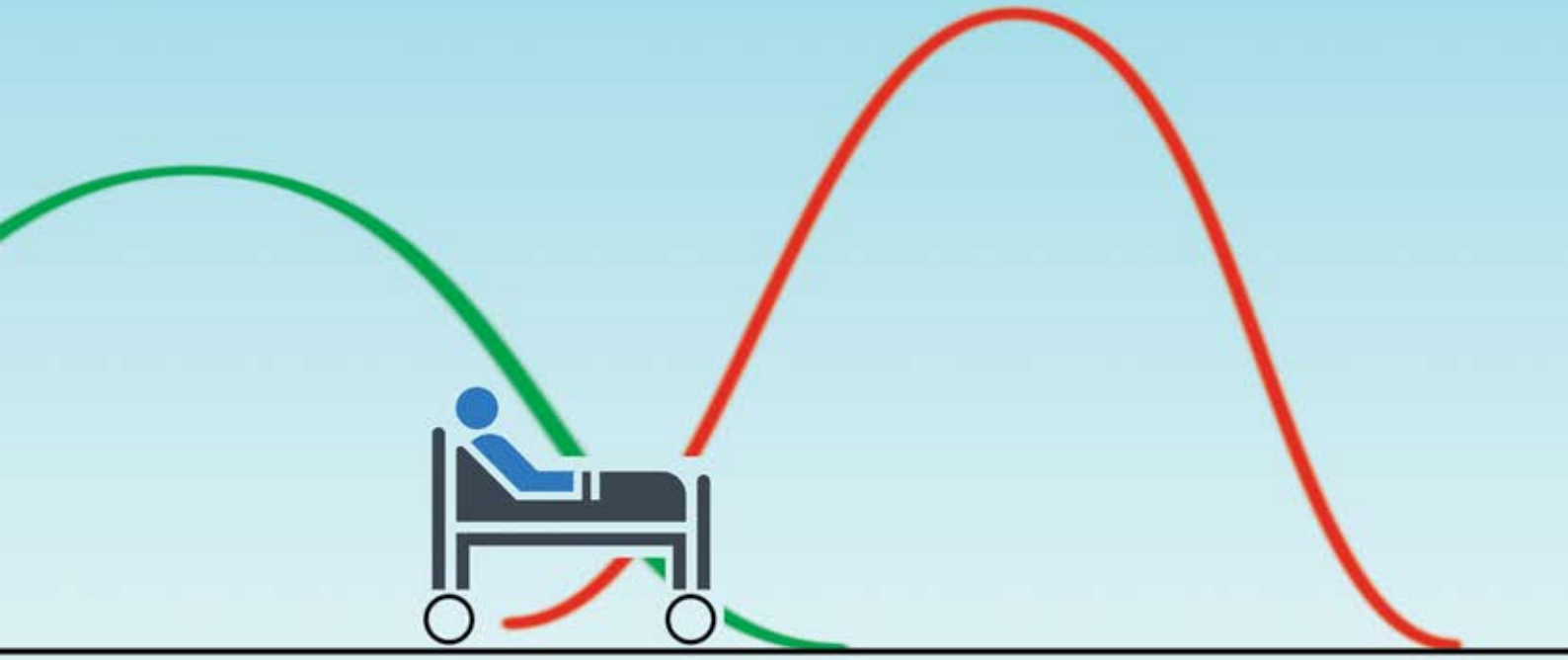
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PhD Thesis

Optimal cutoff points for classification in diagnostic studies: new contributions and software development



Department of Statistics and Operations Research
Santiago de Compostela

Mónica López Ratón

September 2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Universidade de Santiago de Compostela

Ph.D. Thesis

Optimal cutoff points for classification in diagnostic studies: new contributions and software development

Author:
Mónica López Ratón

Advisors:
Carmen Cadarso Suárez
Elisa M. Molanes López

Biostatistics Unit
Department of Statistics and Operations Research

Spain, September 2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

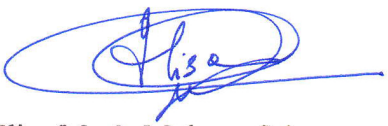
La Dra. Doña Carmen Cadarso Suárez, Catedrática de Universidad del Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela, y la Dra. Doña Elisa María Molanes López, Profesora Ayudante Doctora de la Sección departamental en Medicina del Departamento de Estadística e Investigación Operativa de la Universidad Complutense de Madrid, hacen constar que el trabajo titulado

Optimal cutoff points for classification in diagnostic studies: new contributions and software development

que para la obtención del grado de Doctor en Ciencias Matemáticas presenta Mónica López Ratón, fue realizado bajo su dirección, que lo consideran concluido y solicitan que sea admitido a trámite para su lectura y defensa pública.

En Santiago de Compostela, a 28 de Septiembre de 2015

Las directoras



Carmen Cadarso Suárez y Elisa María Molanes López

La doctoranda

Mónica López Ratón

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A todos los que me acompañan en el camino

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

En primer lugar, quiero expresar mi agradecimiento a las profesoras Carmen Cadarso Suárez y Elisa María Molanes López por su labor de dirección de esta tesis, por todo su apoyo y ánimo en estos años y por toda la confianza que han depositado en mí para la realización de este trabajo.

Gracias a los profesores del Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela, y a los profesores del programa de doctorado de Estadística e Investigación Operativa y del Máster de Bioestadística, por toda la formación recibida a lo largo de estos años. En especial al profesor Xosé Luis Otero Cepeda de la Unidad de Bioestadística de la Facultad de Medicina, como coordinador del convenio de colaboración con el Servicio de Epidemiología de la Consellería de Sanidad de Santiago de Compostela, del que formé parte como becaria.

Quiero agradecer también a los investigadores Francisco Gude Sampedro (Pachicho), de la Unidad de Epidemiología del Complejo Hospitalario Universitario de Santiago de Compostela (CHUS), Emilio Letón del Departamento de Inteligencia Artificial de la UNED en Madrid y Pedro Emmanuel Alvarenga Americano do Brasil del Instituto Nacional de Enfermedades Infecciosas Evandro Chagas y de la Fundación Oswaldo Cruz en Brasil. Y a las investigadoras de la Universidad de Vigo y compañeras María Xosé Rodríguez Álvarez y Mar Rodríguez Girondo. Sin todos vosotros, el trabajo presentado en esta tesis no podría realizarse. Muchas gracias por toda vuestra colaboración e inestimable ayuda.

A Manuel Antonio Botana del Servicio de Endocrinología del Hospital Lucus Augusti de Lugo, a los profesores de la Facultad de Odontología de la Universidad de Santiago de Compostela Mercedes Gallas Torreira y Urbano Santana, y a la investigadora Ana Inocencio Faria, gracias por haberme permitido participar en vuestras investigaciones y retos, y darme la oportunidad de poder aplicar en algunos trabajos la metodología desarrollada en esta tesis.

A todos mis compañeros de la Unidad de Bioestadística por estar siempre ahí y animarme en este camino, dispuestos siempre a ayudar y a echar una mano cuando se os necesita: Laura, Roberto, Ipek, Ana, Isabel, Vicente, Teresa, ...Gracias.

Gracias a mi familia, en especial a mi hermano y a mis padres por todo su apoyo, ayuda y ánimo en cada momento. Y por haberme inculcado la capacidad de trabajo y el espíritu de esfuerzo para poder lograr nuestros objetivos a pesar de las dificultades que puedan surgir en el camino. A Daniel, por su inquebrantable amor por mí, paciencia y perseverancia, por haberme apoyado incondicionalmente en esta tarea a lo largo de todos estos años y por haberme enseñado casi todo el inglés que sé.

A mis amigos y a mis compañeros de secundaria: Raquel, María, Merce, Karina,

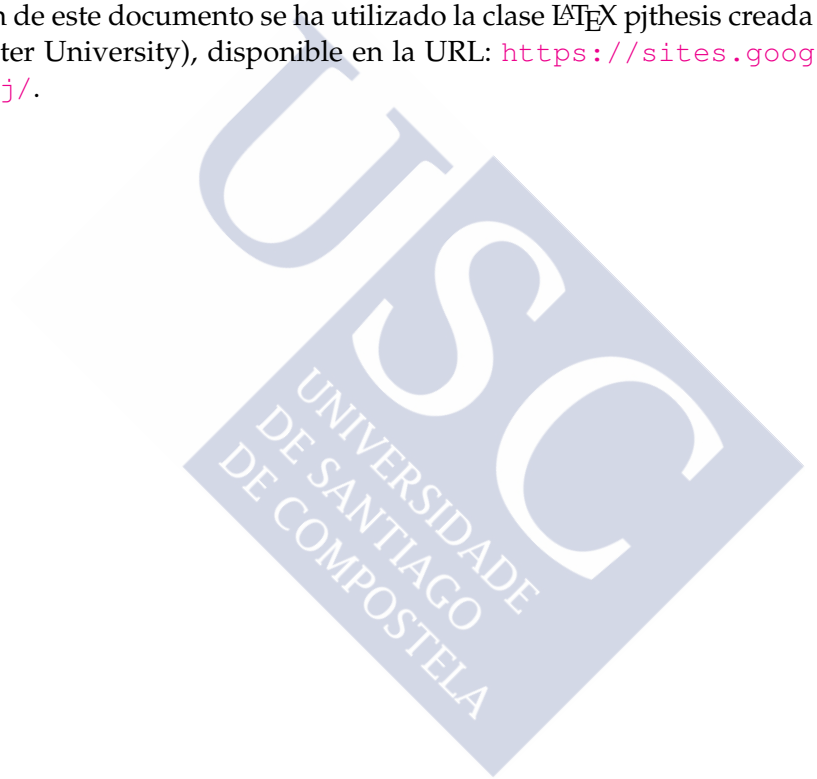
Concha, Chus, ...con los que he compartido y espero seguir compartiendo tan buenos momentos.

Gracias a todos de corazón y perdonad si no he sabido expresar suficientemente bien el agradecimiento que siento por todos los que me han ayudado y apoyado en mi empeño. Gracias!!

Santiago de Compostela, Septiembre 2015
Mónica López Ratón

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación bajo los proyectos MTM2008-01603, MTM2010-09213-E, MTM2011-28285-C02-00, MTM2011-15849-E, y MTM2014-52975-C2-1-R.

Para la creación de este documento se ha utilizado la clase LaTeX pjthesis creada por Peter Jacko (Lancanster University), disponible en la URL: https://sites.google.com/site/latexpj/.



Abstract

Continuous diagnostic tests (biomarkers or risk markers) are often used to discriminate between healthy and diseased populations. For the clinical application of such tests, the key aspect is how to select an appropriate cutpoint or discrimination value c that defines positive and negative test results. In general, individuals with a diagnostic test value smaller than c are classified as healthy and otherwise as diseased. In the literature, several methods have been proposed to select the threshold value c in terms of different specific criteria of optimality. Among others, one of the methods most used in clinical practice is the Symmetry point that maximizes simultaneously both types of correct classifications. From a graphical viewpoint, the Symmetry point is associated to the operating point on the Receiver Operating Characteristic (ROC) curve that intersects the diagonal line passing through the points (0,1) and (1,0). However, this cutpoint is actually valid only when the error of misclassifying a diseased patient has the same severity than the error of misclassifying a healthy patient. Since this may not be the case in practice, an important issue in order to assess the clinical effectiveness of a biomarker is to take into account the costs associated with the decisions taken when selecting the threshold value. Moreover, to facilitate the task of selecting the optimal cut-off point in clinical practice, it is essential to have software that implements the existing optimal criteria in an user-friendly environment. Another interesting issue appears when the marker shows an irregular distribution, with a dominance of diseased subjects in noncontiguous regions. Using a single cutpoint, as common practice in traditional ROC analysis, would not be appropriate for these scenarios because it would lead to erroneous conclusions, not taking full advantage of the intrinsic classificatory capacity of the marker.

In this work, we firstly review the basic concepts on ROC analysis and give a comprehensive overview of the existing methods in the literature to select optimal cutpoints in the setting of a continuous diagnostic test. Secondly, we consider an interesting cost-based generalization of the Symmetry point that incorporates the misclassification costs and we propose how to construct confidence intervals for this optimal cut-off point and its associated accuracy indexes (sensitivity and specificity) using two approaches: a parametric approach based on the Generalized Pivotal Quantity (GPQ) under the assumption of normality and a nonparametric approach based on the Empirical Likelihood (EL) methodology that does not require any parametric assumption. In addition, we describe the main functions of the two `R` packages that we have developed in the framework of this thesis, `OptimalCutpoints` and `GsymPoint`, to facilitate clinicians their usual task of selecting optimal cutpoints in their daily practice. Finally, a new classification rule that provides an improved discriminatory capacity is proposed by transforming the original scale of the biomarker by means of the logistic generalized additive regression

model (GAM), by considering the predicted probabilities obtained from that fit as the new transformed biomarker, and by using the traditional optimal criteria on the transformed scale to finally establish optimal cut-offs or intervals in the original scale on which to base the classification. All this methodology is illustrated along the thesis through real applications based on four real datasets taken from the medical field.

From the results obtained from this research work, we can conclude first that the EL approach is competitive with the GPQ approach when the data follow the Box-Cox model and outperforms the GPQ approach when the data do not follow such a model and the healthy and diseased populations follow different parametric models. Therefore, although the implementation of the EL method is more time consuming than the GPQ method, we recommend the use of the EL method when the distributions of healthy and diseased populations are unknown. In addition, when the relationship between the risk marker and the disease risk (or outcome) is not monotone, we can conclude that using the new transformed biomarker entails an improvement in terms of higher discriminatory capacity. Under these situations, an optimal interval seems more reasonable than a single cutpoint to define the two possible test results (positive or negative result). So, statistical tools designed for flexible modeling can optimize the classificatory capacity of a potential marker using the traditional ROC analysis on the transformed scale. It is therefore important to question linearity in the marker-outcome relationship, in order to take full advantage of the discriminatory capacity of a potential risk marker and avoid erroneous conclusions in clinical practice.

Keywords: Biomarker; Diagnostic test; Empirical Likelihood; Generalized Pivotal Quantity; Logistic GAM regression model; Misclassification costs; Optimal threshold; ROC curve; `R` package; Symmetry point

Resumen

Los tests de diagnóstico (biomarcadores o marcadores de riesgo) continuos se utilizan a menudo para discriminar entre las poblaciones de sanos y enfermos. Para la aplicación clínica de estos tests, el aspecto clave está en cómo seleccionar un punto de corte o valor de discriminación c que defina los resultados positivos y negativos del test. En general, los individuos con un valor del test de diagnóstico menor que c son clasificados como sanos y en caso contrario como enfermos. En la literatura se han propuesto varios métodos para seleccionar el punto de corte c en términos de diferentes criterios específicos de optimalidad. Entre otros, uno de los métodos más utilizados en la práctica clínica es el punto de Simetría que maximiza simultáneamente ambos tipos de clasificaciones correctas. Desde un punto de vista geométrico, el punto de Simetría se corresponde con el punto sobre la curva ROC ("Receiver Operating Characteristic") que interseca con la línea diagonal que une los puntos (0,1) y (1,0). Sin embargo, este punto en realidad sólo es válido cuando el error de clasificar incorrectamente a un paciente enfermo tiene la misma gravedad que el error de no clasificar correctamente a un paciente sano. Dado que este supuesto puede no ser cierto en la práctica clínica, con el objetivo de evaluar la efectividad clínica de un biomarcador, un aspecto importante a la hora de seleccionar el punto de corte es tener en cuenta los costes asociados a las decisiones tomadas en base a dicho punto de corte. Además, para facilitar la tarea de seleccionar el punto de corte óptimo en la práctica clínica es esencial disponer de software que implemente los criterios óptimos existentes en un entorno amigable para el usuario final. Otro aspecto interesante aparece cuando el marcador muestra una distribución irregular, con un dominio de sujetos enfermos en regiones no contigüas. La utilización de un único punto de corte, siguiendo la práctica habitual en el análisis ROC tradicional, no sería apropiado para estos escenarios ya que daría lugar a conclusiones erróneas, no aprovechando al máximo la capacidad de clasificación intrínseca del marcador.

En este trabajo, se llevó a cabo en primer lugar una revisión de los conceptos básicos del análisis ROC y una visión global de los métodos existentes en la literatura para seleccionar puntos de corte óptimos en el contexto de un test de diagnóstico continuo. En segundo lugar, se consideró una interesante generalización del punto de Simetría basada en costes que incorpora los costes derivados de las clasificaciones incorrectas, y se propuso la construcción de intervalos de confianza para este punto de corte óptimo y sus medidas de efectividad ("accuracy") asociadas (sensibilidad y especificidad) utilizando dos planteamientos: un enfoque paramétrico basado en la Cantidad Pivotal Generalizada (o "Generalized Pivotal Quantity", GPQ) bajo la hipótesis de normalidad y un planteamiento no paramétrico basado en la metodología de la Verosimilitud Empírica (o "Empirical Likelihood", EL). Además, se describieron las funciones principales de los

dos paquetes que se desarrollaron en `R` en el marco de esta tesis, `OptimalCutpoints` y `GsymPoint`, con el objetivo de facilitar a los clínicos su tarea habitual de seleccionar puntos de corte óptimos en su práctica diaria. Finalmente, se propuso una nueva regla de clasificación que proporciona una mejora en la capacidad de discriminación, transformando la escala original del biomarcador mediante el modelo de regresión aditivo generalizado ("Generalized Additive Model", GAM) logístico, considerando las probabilidades predichas obtenidas a partir de ese ajuste como el nuevo biomarcador transformado, y utilizando los criterios óptimos tradicionales sobre la escala transformada para finalmente establecer puntos de corte óptimos o intervalos en la escala original sobre los cuales basar la clasificación. Toda esta metodología se ilustra a lo largo de la tesis a través de aplicaciones reales basadas en cuatro conjuntos de datos reales tomados del ámbito médico.

A raíz de los resultados obtenidos en este trabajo de investigación, se puede concluir en primer lugar que la aproximación EL es competitiva con la aproximación GPQ cuando los datos siguen el modelo Box-Cox y es superior a la aproximación GPQ cuando los datos no siguen dicho modelo y las poblaciones de sanos y enfermos siguen diferentes modelos paramétricos. Por lo tanto, aunque la implementación del método EL sea más costosa desde un punto de vista computacional, se recomienda utilizar el método EL cuando las distribuciones en las poblaciones de sanos y enfermos sean desconocidas. Además, cuando la relación entre el riesgo de enfermedad y el marcador de riesgo no sea monótona, se puede concluir que el nuevo biomarcador transformado proporciona una mejoría en términos de mayor capacidad diagnóstica. Bajo estas situaciones, un intervalo óptimo parece más razonable que un solo punto de corte para definir los dos resultados posibles del test (resultado positivo o negativo). De modo que, herramientas estadísticas diseñadas para una modelización flexible pueden optimizar la capacidad clasificatoria de un marcador potencial utilizando el análisis ROC tradicional sobre la escala transformada. Por tanto, es importante cuestionar la relación de linearidad entre el marcador y la respuesta con el objetivo de aprovechar al máximo la capacidad de discriminación intrínseca de un biomarcador potencial y evitar conclusiones erróneas en la práctica clínica.

Palabras clave: Biomarcador; Cantidad Pivotal Generalizada; Costes de clasificación errónea; Curva ROC; Modelo GAM de regresión logística; Paquete de `R`; Punto de corte óptimo; Punto de simetría; Test de diagnóstico; Verosimilitud Empírica.

Resumo

Os tests de diagnóstico (biomarcadores ou marcadores de risco) continuos empréganse a menudo para discriminar entre as poboacións de sans e enfermos. Para a aplicación clínica destes tests, o aspecto clave está en cómo seleccionar un punto de corte ou valor de discriminación axeitado c que defina os resultados positivos e negativos do test. En xeral, os individuos cun valor do test de diagnóstico menor que c son clasificados como sans e noutro caso como enfermos. Na literatura propuxéronse varios métodos para seleccionar o punto de corte c en termos de diferentes criterios específicos de optimalidade. Entre outros, un dos métodos máis empregados na práctica clínica é o punto de Simetría que maximiza simultáneamente ambos tipos de clasificacións correctas. Desde un punto de vista xeométrico, o punto de Simetría correspóndese co punto sobre a curva ROC ("Receiver Operating Characteristic") que interseca coa liña diagonal que une os puntos (0,1) e (1,0). Sen embargo, este punto de corte en realidade só é válido cando o erro de clasificar incorrectamente a un paciente enfermo ten a mesma gravidade que o erro de non clasificar correctamente a un paciente san. Dado que este suposto pode non ser certo na práctica, co obxectivo de evaluar a efectividade clínica dun biomarcador, un aspecto importante á hora de seleccionar o punto de corte é ter en conta os costes asociados ás decisións tomadas en base a ese punto de corte. Ademais, para facilitar a tarefa de elexir os puntos de corte óptimos na práctica clínica é esencial dispoñer de software que implemente os criterios óptimos existentes nun entorno amigable para o usuario final. Outro aspecto interesante aparece cando o marcador mostra unha distribución irregular, cun dominio de suxeitos enfermos en rexións non contigüas. Empregar un só punto de corte, de acordo coa práctica habitual na análise ROC tradicional, non sería apropiado para estes escenarios, xa que daría lugar a conclusións incorrectas, non aproveitando ó máximo a capacidade de clasificación intrínseca do marcador.

Neste traballo, levouse a cabo en primeiro lugar unha revisión dos conceptos básicos da análise ROC e unha visión global dos métodos existentes na literatura para elexir puntos de corte óptimos no contexto dun test de diagnóstico continuo. En segundo lugar, considerouse unha interesante xeralización do punto de Simetría basada en costes que incorpora os costes derivados das clasificacions incorrectas, e propúxose a construcción de intervalos de confianza para este punto de corte óptimo e as súas medidas de efectividade ("accuracy") asociadas (sensibilidade e especificidade) empregando dous planteamentos: un enfoque paramétrico basado na Cantidade Pivotal Xeralizada ("Generalized Pivotal Quantity", GPQ) baixo a hipótese de normalidade e un planteamiento non paramétrico basado na metodoloxía da Verosimilitude Empírica ("Empirical Likelihood", EL). Ademais, describíronse as funcións principais dos dous paquetes que se desenvolveron en `R` no marco desta tese, `OptimalCutpoints` e `GsymPoint`, co obxectivo

de facilitar ós clínicos a súa tarefa habitual de seleccionar puntos de corte óptimos na súa práctica diaria. Para rematar, propúxose unha nova regra de clasificación que proporciona unha mellora na capacidade de discriminación, transformando a escala orixinal do biomarcador mediante o modelo de regresión aditivo xeralizado ("Generalized Additive Model", GAM) loxístico, considerando as probabilidades preditas obtidas a partir dese axuste como o novo biomarcador transformado, e empregando os criterios óptimos tradicionais sobre a escala transformada para finalmente estabrecer puntos de corte óptimos ou intervalos na escala orixinal sobre os cales basear a clasificación. Toda esta metodoloxía é ilustrada ó longo da tese a través de aplicacións reais basadas en catro conxuntos de datos reais tomados do ámbito médico.

Á raíz dos resultados obtidos neste traballo de investigación, pódese concluir en primeiro lugar que a aproximación EL é competitiva coa aproximación GPQ cando os datos seguen o modelo Box-Cox e é superior á aproximación GPQ cando os datos non seguen ese modelo e as poboacións de sans e enfermos seguen diferentes modelos paramétricos. Polo tanto, aínda que a implementación do método EL sexa máis costosa desde un punto de vista computacional, recoméndase empregar o método EL cando as distribucións nas poboacións de sans e enfermos sexan descoñecidas. Ademais, cando a relación entre o risco de enfermidade e o marcador de risco non sexa monótona, pódese concluir que o novo biomarcador transformado proporciona unha mellora en termos da maior capacidade diagnóstica. Porén, baixo estas situacións, un intervalo óptimo parece máis razoable que un só punto de corte para definir os dous resultados posibles do test (resultado positivo ou negativo). De modo que, ferramentas estatísticas diseñadas para unha modelización flexible poden optimizar a capacidade clasificatoria dun marcador potencial empregando a análise ROC tradicional sobre a escala transformada. Polo tanto, é importante cuestionar a relación de linearidade entre o marcador e a resposta co obxectivo de aproveitar ó máximo a capacidade de discriminación intrínseca dun biomarcador potencial e evitar conclusións erradas na práctica clínica.

Palabras clave: Biomarcador; Cantidade Pivotal Xeralizada; Costes de clasificación incorrecta; Curva ROC; Modelo GAM de regresión loxística; Paquete de `R`; Punto de corte óptimo; Punto de simetría; Test de diagnóstico; Verosimilitude empírica.

Contents

List of Figures

List of Tables

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

It is good to have an end to journey toward;
but it is the journey that matters, in the end.

- Ernest Hemingway

Chapter 1

Introduction

1.1 Motivation and objectives

The problem of classification in diagnostic studies is a crucial first step in clinical practice because it will largely determine the subsequent treatment (surgery, drug administration, ...) and posterior care of the patient. There are many situations where the true disease status of the patient is unknown and therefore it is important to have a good diagnostic tool to classify correctly as many individuals as possible, specially in situations of serious or even mortal diseases such as cancer. Continuous **diagnostic tests**, **risk markers** or **biomarkers** are often used for discriminating between alternative states of health. In this work, we will focus on the two state case, the most recurrent and frequent case considered in practice, since normally the first question to answer is if the patient has (or not) the event or outcome of interest, the target condition or disease under study. So, in modern medical practice, it is very important to know how to use a continuous diagnostic test, Y, to discriminate between non-diseased and diseased individuals. For this reason, for the clinical application of such tests it is useful to select an appropriate **cutpoint** or **discrimination threshold** c that defines negative and positive test results. In general, under the common assumption that higher values of the biomarker are associated with the disease, this discrimination task is usually based on a cut-off value, c, such that depending on whether $Y < c$ or $Y \geq c$, the individual is classified as healthy (a negative test result) or diseased (a positive test result), respectively. It should be noted, however, that this classification is not error-free and the diagnostic test can incorrectly classify a patient in the wrong group. Accordingly, it is recommendable that before routine application of a diagnostic test in practice, the misclassification errors are quantified together with the selection of the appropriate cutpoint to avoid erroneous conclusions.

In this context, the **Receiver Operating Characteristic (ROC) curve** is a commonly used tool for evaluating the discriminatory capacity of Y (Metz, 1978) in distinguishing between alternative states of health. The ROC curve is obtained by plotting the pairs $(1 - Sp(c), Se(c))$, with $-\infty < c < \infty$, and where $Sp(c)$ denotes the **specificity** or probability of correctly classifying a healthy patient (the true negative rate, TNR) and $Se(c)$ denotes the **sensitivity** or probability of correctly classifying a diseased patient (the true positive rate, TPR). The ROC curve can be also reparameterized in the interval $(0, 1)$ as $ROC(x) = 1 - F_1(F_0^{-1}(1 - x))$, where $x \in (0, 1)$ is the false positive rate (FPR), that is, $x = 1 - Sp(c) = 1 - F_0(c) = FPR(c)$, F_i denotes the cumulative distribution function

(cdf) of the biomarker in the i-th population, Y_i, and the subindex $i = 0, 1$ is used, respectively, for the healthy and diseased populations. The area under the ROC curve (AUC), although it is a summary index of the global performance or accuracy of a biomarker, it can not be used to select an optimal cut-off value, a key point in this context to yield a binary classification rule that can be used in practice.

Therefore, the main objective here is to select the **optimal cutpoint** of the diagnostic marker that **best discriminates** between patients with and without the disease. However, the optimal cutpoint depends primarily on the situation where it is to be used, and so, in general, one cannot talk in absolute terms of a "best choice" of cutpoint c. For this task, several methods or strategies for identifying an optimal cutpoint in continuous diagnostic tests in the sense of a specific optimality criterion have been proposed in the literature depending mainly on the underlying reason or goal for this choice (see Youden, 1950; Feinstein, SH, 1975; Metz, 1978; Albert and Harris, 1987; England, 1988; Schäfer, 1989; Vermont et al., 1991; Greiner, 1995; Riddle and Stratford, 1999; Pepe, 2004; Rota and Antolini, 2014, among others). The majority of such methods are based on the ROC methodology. In fact, ROC curves can be used for computing the optimal cutpoints taking into account the prevalence of the disease in the study population (the probability of having the disease) and the ratio of relative costs (risk-benefit ratio) of the possible diagnostic decisions (either correct or incorrect classifications) derived from the diagnostic test result. It is very important to take into account the value of such costs when selecting the optimal cutpoint because not all situations can be equally assessed and, depending on the specific circumstances of each situation (characteristics of the disease, available treatment or not, invasive treatment, ...), a study of the selection of the optimal cutpoint incorporating costs (or, equivalently, utilities) should be carried out. Besides, despite the difficulties that may involve allocating costs in practice, the values of these costs must be selected according to those specific circumstances. Thus, for example, in the case of a serious or fatal disease that can not be overlooked, such as cancer, it is necessary to have the highest possible sensitivity, that is, many true positives because it is primordial to diagnose the disease, and consequently, few false negatives. From this perspective, the cost of a false negative is more serious than the cost of a false positive and this should be taken into account when estimating the appropriate optimal cutpoint by assigning a higher value to the false negative than to the false positive cost.

A fundamental aspect for the real application of all these methods that facilitates the task of selecting optimal values in clinical practice is the availability and development of software that implements the different optimal-cutpoint selection criteria in a user-friendly environment and with outputs easy to interpret by professionals of applied research, especially in the biomedical field.

On the basis of all the aspects previously discussed so far, the main objectives that we intend to address in this doctoral thesis are summarized below:

1. To review the basic concepts on the traditional ROC analysis.

2. To provide a comprehensive overview and study of the methods that have been proposed in the literature to select optimal cutpoints for diagnostic tests, discussing their relative advantages and disadvantages, and referring to the conditions of application in clinical practice.

3. To identify which selection methods are equivalent, similar or special cases of others from a theoretical perspective in an intent to reduce the many options currently available.

4. To develop efficient estimation and inference techniques for the selection of optimal cutpoints that take into account the misclassification costs.

5. To develop software in a friendly environment that can be used easily, in a suitable and comfortable way for all those professionals who are interested in the practical use of this methodology.

6. To detect theoretical scenarios where the traditional ROC analysis may fail to take full advantage of the discriminatory capacity of a continuous diagnostic marker, to describe a new methodology that gives an improved solution under those situations, and to study its impact on the selection of the optimal cutpoints.

7. To apply this methodology to real data sets in the field of clinical practice.

8. To identify new interesting lines of future research.

1.2 Structure of the thesis

Considering this chapter devoted to introduce the motivation, objectives and structure, this thesis is structured in seven chapters. In a nutshell, it is intended to provide advances in the study of optimal cutoff points for classification purposes in diagnostic studies and development of user-friendly software that can be used by those professionals interested in this field. The contents of this work are structured as follows.

We first begin with a review of general and basic concepts on diagnostic tests in Chapter 2, describing different types of diagnostic tests, its accuracy measures and the most important or relevant aspects of the traditional ROC methodology. The main objective of this first chapter is to introduce some essential contents that will be used in the following chapters, according to our first objective. A spanish version of some of the contents of this chapter is part of the book entitled "Técnicas de estimación e inferencia de las curvas ROC", published in 2012 by *Editorial Académica Española* (López-Ratón et al., 2012a). Additionally, based on traditional ROC analysis, a real data application has been carried out to compare the validity of two different methods (Olive-Basford's method and Olmos' method) as predictive methods of mandibular third molar impaction. This real application has given rise to the publication entitled "Comparison between two radiographic methods used for the prediction of mandibular third molar impaction" published in 2014 in *Revista Portuguesa de Estomatologia, Medicina Dentária e Cirurgia Maxilofacial* (Gallas-Torreira et al., 2014).

According to our second and third objectives, Chapter 3 is devoted to give a detailed and comprehensive review of the dichotomization methods that have been proposed in the literature to select optimal cutpoints or thresholds in continuous diagnostic tests when there are two groups of interest (healthy and diseased), discussing their relative advantages and disadvantages, and identifying which methods are equivalent, similar

or special cases of others in an intent to reduce the many options currently available. It is important to mention here that in some of the existing methods the user or investigator must select beforehand some specific values, for instance, the probability of disease or the misclassification costs. Having in mind this aspect, the exposition of the currently available methods is classified into two groups: **dichotomization data-driven methods** and **dichotomization methods with user requirements**. The contents of this chapter are partially based on the study carried out in collaboration with researchers from the Evandro Chagas Infectious Diseases National Institute and the Oswaldo Cruz Foundation in Brasil (Pedro Emmanuel Alvarenga Americano do Brasil and Marcel de Souza Borges), from the German Peadiatric Pain Center and Chair for Children's Pain Therapy and Pediatric Palliative Care (Gerrit Hirschfeld), and from the University of Vigo (María Xosé Rodríguez Álvarez). A version of this chapter has given rise to the publication entitled "Methods to estimate decision thresholds for diagnostic tests in medicine: Comparing similarities of 55 methods" which is under review for publication (Alvarenga Americano do Brasil et al., 2015). Additionally, some of the most common criteria for selecting the optimal cutpoints have been applied to a real dataset on diabetes mellitus in order to study the discriminatory capacity of the glycated hemoglobin levels. This real application has yielded the publication entitled "Relationship between glycated hemoglobin and glucose concentrations in the adult Galician population: selection of optimal glycated hemoglobin cut-off points as a diagnostic tool of diabetes mellitus", published in 2012 in *Endocrinología y Nutrición* (Botana-López et al., 2012).

After the revision presented in Chapter 3, we realized that some of the existing criteria for selecting the optimal threshold have been most used in clinical practice and consequently most studied in the literature. One of them is the **Youden index** (Youden, 1950) defined as $J = \max_c \{Se(c) + Sp(c) - 1\}$. However, it should be noted that such an index is a particular case of the Bayes rule for the minimum misclassification error probability (in the sense of considering only binary classification rules based on a single cutoff point), where the disease prevalence is assumed equal to 50% which, from a practical viewpoint, may be far from reality. In addition, there are clinical situations where it could be more adequate to maximize simultaneously both types of correct diagnostic decisions rather than the sum of them (Riddle and Stratford, 1999). So, in Chapter 4 we met our fourth objective. More specifically, we focus on an specific optimality criterion, the **Symmetry point**, defined as the point c_S where $Se(c_S) = Sp(c_S)$, that maximizes simultaneously the two types of correct classifications, and we introduce a cost-based generalization of it, the **Generalized Symmetry point**, taking into account the misclassifications costs. After introducing these two optimal cutpoint criteria from a Bayes decision theory perspective, we construct confidence intervals for the optimal cutpoint obtained by the Generalized Symmetry point, and its associated specificity and sensitivity indexes using two approaches: one based on the **Generalized Pivotal Quantity** (Weerahandi, 1993, 1995) and the other based on **Empirical Likelihood** (Thomas and Grunkemeier, 1975). In addition, we perform an extensive simulation study to check the practical behaviour of both methods and we illustrate their use by means of three real biomedical data sets on melanoma, prostate cancer and coronary artery disease which will be explained in detail in the next section. The study performed for the Symmetry point and its cost-based generalization is the result of the communication entitled "Inference of the Symmetry point

with different costs for the specificity and sensitivity" presented in the 27^{th} International Workshop on Statistical Modelling (López-Ratón et al., 2012b), and the work entitled "Confidence intervals for the Symmetry point: an optimal cutpoint in continuous diagnostic tests", which is under second revision in *Pharmaceutical Statistics* (López-Ratón et al., 2015a).

In the framework of the methodology described and introduced in this thesis and according to our fifth objective, we have developed two `R` (R Core Team, 2015) packages which are freely available from the Comprehensive `R` Archive Network (CRAN). Albeit later on we give a detailed description of each package in Chapter 5, now we present a brief summary of them:
1) The `OptimalCutpoints` package (López-Ratón and Rodríguez-Álvarez, 2014), available at `http://CRAN.R-project.org/package=OptimalCutpoints`, includes almost all methods reviewed in Chapter 3 for selecting the optimal cutpoint in continuous diagnostic tests. This package is described in Section 5.1 of Chapter 5. That description is based on the paper "OptimalCutpoints: An `R` Package for selecting optimal cutpoints in diagnostic tests" published in 2014 in *Journal of Statistical Software* (López-Ratón et al., 2014).
2) The `GsymPoint` package, available at `http://CRAN.R-project.org/package=GsymPoint`, see López-Ratón et al. (2013, 2015b), implements the two inference methods introduced and studied in Chapter 4 for the Generalized Symmetry point. This package is described in Section 5.2 of Chapter 5. A version of the contents of that section will be submitted soon for publication in *Journal of Statistical Software* (López-Ratón et al., 2015c).
In the detailed description of both `R` packages given in Chapter 5, we follow the same structure for each, we first explain the use of the main functions, and we then give an illustration of their practical application using different real data sets.

Traditionally, in classical ROC analysis, marker levels above and below a given cut-off value result in individuals being labeled as diseased or nondiseased, respectively. However, in cases where the marker shows an irregular distribution, with a dominance of diseased subjects in non-contiguous regions, classification by reference to a cut-off value is neither feasible nor logical. Indeed, use of such an analysis would lead to erroneous conclusions, and a modification of the classification rule is therefore necessary (Lustres-Pérez et al., 2010). According to our sixth objective and, in an intent to give a practical solution to this issue, Chapter 6 is devoted to describe a procedure for improving the discriminatory capacity of a continuous biomarker in the above-mentioned situations, by using a flexible regression technique based on generalized additive models (GAMs) (Hastie and Tibshirani, 1990) for binary data. Specifically, a new classification rule is obtained by transforming the original scale of the biomarker by means of the logistic GAM regression model, and by considering the predicted probabilities obtained from that fit as the new transformed biomarker. Based on the scale of this new transformed biomarker, we propose to establish optimal cut-offs or intervals in the original scale of the biomarker and use them for classification purposes. In addition, we validate this methodology under different theoretical scenarios, and we present a real application using data from a prospective study of patients undergoing surgery at a University Teaching Hospital, where the first objective was to examine the plasma glucose levels as a potential biomarker for postoperative infection (Figueiras and Cadarso-Suárez, 2001). This

chapter is partially based on the contents of the manuscript entitled "Application of generalized additive models to the evaluation of continuous markers for classification purposes", recently published in 2015 in *International Journal of Statistics in Medical Research* (López-Ratón et al., 2015d). Besides, this GAM based methodology has been applied to determine the diagnostic accuracy of the use of surface electromyography (sEMG) as a potential diagnostic tool to detect chronic temporomandibular disorders (TMD). This real application has given rise to the publication entitled "Surface raw electromyography has a moderate discriminatory capacity for differentiating between healthy individuals and those with TMD: A diagnostic study" published in 2014 in *Journal of Electromyography and Kinesiology* (Santana-Mora et al., 2014).

Finally, this work ends with a brief discussion in Chapter 7, that provides not only some conclusions of the studies carried out in Chapters 3–6 but also new interesting extensions of them for future research lines.

1.3 Data sets

To exemplify and illustrate all the methodology developed along the thesis, we will use four real biomedical applications on melanoma, prostate cancer, coronary artery disease and postoperative infection, covering thus pathologies with different characteristics. In this section we explain in detail such data sets.

1.3.1 Melanoma dataset

In order to determine if a suspicious pigmented lesion on the skin is a melanoma, dermatologists use a clinical scoring scheme without dermoscope (CSS) or a dermoscopic scoring scheme (DSS) to clinically evaluate the lesion on the basis of several visible features such as asymmetry, border irregularity, colouration and size. Venkatraman and Begg (1996) have analyzed a dataset on 72 patients with suspicious lesions of being a melanoma. All these patients underwent a biopsy and it turned out that 21 out of them were diagnosed as suffering from melanoma. In cutaneous melanoma, the true diagnosis is produced when removing the lesion and examining the anatomo-pathological characteristics of the piece biopsied. In order to avoid this invasive technique, it is interesting to analyze if there exist other less invasive alternatives on which to base the diagnose of melanoma, such as the CSS or the DSS markers. We will analyze here and in some of the following chapters, the CSS marker, as a potential continuous diagnostic marker to detect the presence of melanoma. Here we simply collect in Table 1.1 the main descriptive statistics for each group. It is observed that higher values of CSS are associated with the presence of melanoma.

Table 1.1: Descriptive statistics of melanoma dataset.

	n	Mean	Median	Variance	sd	Min	Max
Melanoma	21 (29%)	0.557	0.406	2.321	1.523	-2.763	3.032
Non melanoma	51 (71%)	-2.426	-2.293	3.099	1.760	-5.881	1.751
Total	72	-1.556	-1.696	4.701	2.168	-5.881	3.032

1.3.2 Prostate cancer dataset

In order to design an appropriate treatment strategy for a patient with prostate cancer, it is important to know first whether cancer has spread or not to the neighboring lymph nodes. Although the true status of the patient could be confirmed through a laparoscopic surgery, it is interesting to find a non-invasive diagnostic method to predict whether nodal involvement is present. Some possible predictors include X-ray, palpation, biopsy and "acid phosphatase level in blood serum", among others (Le, 2006). We analyze here the continuous measurement of the level of acid phosphatase in blood serum (APBS) $\times 100$ as a potential continuous diagnostic test for predicting nodal involvement, using a dataset on prostate cancer taken from Miller et al. (1980) and that has been previously studied in detail in Le (2006). This dataset has 20 patients with nodal involvement and 33 without (Le, 2006). For the practical use of this marker in clinical practice, it is necessary to dichotomize the marker so that we can classify the individuals into the two groups of interest (patients with nodal involvement and patients without nodal involvement). For the case of APBS, clinicians have long observed that an elevated level tend to be associated with a nodal involvement, that is, higher marker values are associated with disease. In Table 1.2 we show the basic descriptive statistics (mean, median, variance and standard deviation in parentheses, minimum and maximum value) of the APBS $\times 100$ for each group.

Table 1.2: Descriptive statistics of prostate cancer dataset.

	n	Mean	Median	Variance	sd	Min	Max
With nodal involvement	20 (38%)	77.500	74	515.105	22.696	48	136
Without nodal involvement	33 (62%)	64.515	55	744.133	27.279	40	187
Total	53	69.42	65	686.517	26.201	40	187

1.3.3 Coronary artery disease dataset

The main disadvantage of the existing non-invasive methods to detect coronary artery disease (CAD), such as electrocardiographies (with or without effort) and echocardiographies (with or without stress), is their limited success in the detection of CAD in its early stages. Therefore, in order to provide an early diagnosis of CAD, it is of main interest to analyze the diagnostic capability of biochemical factors, such as the peripheral blood leukocyte elastase concentration, a proteolytic enzyme contained in the azurophilic granules of neutrophils, that has been implicated in the pathophysiology of ischaemic heart disease. We analyze here this biochemical factor as a potential diagnostic indicator of CAD using the data from a study conducted on 141 consecutive patients admitted to the Cardiology Department of a Teaching Hospital in Galicia (northwest of Spain) for evaluation of chest pain or cardiovascular disease (Amaro et al., 1995). All patients underwent coronary angiography during investigation: 96 had coronary lesions and 45 had non-stenotic coronaries. In Table 1.3 we present some basic descriptive statistics for each group. From these results, it is observed that higher elastase values are associated with coronary lesions and that there is more variability in this group.

Table 1.3: Descriptive statistics of coronary artery disease dataset.

	n	Mean	Median	Variance	sd	Min	Max
Coronary lesions	96 (68%)	49.729	43	752.621	27.434	13	163
Non-stenotic coronaries	45 (32%)	29.522	31	213.704	14.619	5	56
Total	141	43.28	39	667.230	25.831	5	163

1.3.4 Postoperative infection dataset

Infections after surgical procedures can cause several problems such as pain, need for further treatment including antibiotics, longer hospital stays, increased health care costs, and even severe complications as failure of the operation, sepsis, organ failure or death (Torpy et al., 2010). So, it is important to avoid the presence of postoperative infection (POI) in order to prevent all these possible problems and complications. We considered here data drawn from a prospective study of patients who underwent surgical interventions at the Hospital Clínico Universitario de Santiago (Santiago de Compostela, Spain) in the period from January 1996 to March 1997 (Figueiras and Cadarso-Suárez, 2001). The aim of this study was to ascertain factors associated with the appearance of postoperative infection (POI). Some known risk factors for postoperative infection are diabetes, obesity, older age, emergency operations and obvious contamination (with pus, stool, or other substances) of the injury or the surgical area (Torpy et al., 2010). Of the 2353 individuals contained in the database, postoperative infection was detected in 460 during follow-up, and infection was strongly correlated with the type of surgery performed. The classification of the surgery type was based on the Altemeier classification (Altemeier, 1979), which categorizes surgery into the following 4 types: clean; clean-contaminated; contaminated; and dirty to a lesser or greater degree of bacterial contamination. This study evaluated whether glucose levels could predict appearance of infection in the immediate postoperative period. In order to prevent the presence of possible confounding variables, only the group of clean surgical interventions on nondiabetic individuals was considered for analysis purposes in the following chapters of the thesis. The final sample consisted of 836 patients, 45 of whom had POI. In Table 1.4 we show some descriptive statistics of this reduced dataset that, from here on, we will refer to it as the postoperative infection dataset. From these results, it is observed that subjects with POI present higher mean glucose values, although the maximum value is reached in the non-POI group and the minimum value in the group with POI. This dataset will be specially interesting for the illustrative application of the GAM-based methodology presented in Chapter 6.

Table 1.4: Descriptive statistics of coronary artery disease dataset.

	n	Mean	Median	Variance	sd	Min	Max
POI	45 (5%)	116.600	106.000	1393.568	37.331	56.000	240.000
Non POI	791 (95%)	103.400	98.000	556.458	23.589	68.000	265.000
Total	836	104.100	99.000	608.802	24.674	56.000	265.000

All human knowledge begins with intuitions,
proceeds from thence to concepts,
and ends with ideas.

- Immanuel Kant

Chapter 2

Basic concepts on diagnostic tests

As a first step to address the study of the selection of optimal cutoff points for classification in diagnostic studies, it is necessary to know some basic concepts on diagnostic tests. In this chapter we review the basic concepts primarily related to the different accuracy measures of diagnostic tests, specifying their advantages, disadvantages and indications for further application in clinical practice. This review aims to be didactic and the cornerstone to understand in an easy and logical way the need of selecting the optimal cutoff point in continuous diagnostic tests, and the rationale of the different criteria of optimality proposed in the literature (see Chapter 3 for more details).

The **diagnosis** can be considered as one of the most important results of medical practice, the key that leads to the treatment and prognosis of the patient. The different phases of the diagnostic process generally involve several sources of information, such as medical history, physical examination, epidemiological information and complementary tests (Sandler, 1980; Sackett et al., 1994). A **diagnostic test** could be defined as any kind of medical test performed to aid in the diagnosis or detection of disease and it is an essential tool in the practice of modern medicine (Barrat et al., 1999). Other possible definitions are: a diagnostic test is any process, more or less complex, that seeks to determine in the patient the presence of some supposedly pathological condition that it is likely to not be directly observed (with any of the five basic senses). For instance, a diagnostic test is useful to diagnose diseases, to confirm that a person is free from disease and/or to measure the progress or recovery from disease. So, diagnostic tests are often used for discriminating between healthy and diseased populations and it is important to know how to do this classification. However, from a statistical point of view we may consider medical diagnose in a much broader context, since statistical methodologies pertinent to diagnosing disease also apply to other important classification problems in medicine, for example, the prognosis or disease screening test.

The interpretation of a diagnostic test in turn depends on several factors:

(i) The intrinsic ability of the diagnostic test to distinguish between healthy and diseased patients (ability or discrimination accuracy).

(ii) The individual characteristics of each patient.

(iii) The environment where the diagnostic test is applied.

Some authors (Begg and Greenes, 1983; Silva, 1987) stated that the use of diagnostic tests for the detection and evaluation of various diseases in clinical practice has increased significantly and tends to keep increasing exponentially. On one hand, the incorporation into clinical practice of more sophisticated and novel diagnostic methods (due largely to great technological development in recent years) have made this a reality. On the other hand, however, these technological advances have led to a somewhat indiscriminate use by some clinicians that do not always act in the most appropriate way.

Assessment of the reliability, accuracy and impact of the available tests is essential to guide on the selection of the optimal test and to interpret appropriately tests results. It is clear that a good diagnostic test is that which provides positive results in diseased individuals and negative results otherwise. Therefore, according to the ideas presented in Morrison (1992), the required conditions of a diagnostic test are the following:

- **Validity**: the degree to which the test measures what it is supposed to measure (Khan and Chien, 2001; Seffinger et al., 2004). This involves comparing the results of the diagnostic test to the results of a reference test or gold standard that even when it does provide more reliable results on the state of the patient regarding the condition under study, it has the disadvantage of being too expensive and/or impractical for routine use in clinical practice; that is, the validity answer the question of "how often the test result is confirmed by more complex and rigorous diagnostic procedures?"

- **Reproducibility**: the ability of the test to give the same results when applied in similar circumstances or conditions (Khan and Chien, 2001; Seffinger et al., 2004; Bruno et al., 2011). A test to be of potential use in practice has to give consistent results. Several sources of variability such as biological variability of the observed fact, variability introduced by the observer himself (or herself) and intrinsic variability of the test itself, determine its reproducibility.

- **Safety**: the degree to which the test will predict the presence or absence of disease. For example, what is the probability that a positive diagnostic test result indicates the presence of disease?

In turn, it is desirable that a diagnostic test is simple to apply or implement, accepted by the patients or the population in general, to have the minimum adverse effects and be economically acceptable.

Depending on the outcome of the test, these are classified into:

- **Binary tests**: The test has only two possible outcomes, positive or negative for indicating the presence of disease, that is, the test is a binary variable. For instance, developing a certain disease or not.

- **Discrete tests**: The test can take different values (usually integer values) and the set of possible outcomes is finite. For instance, positive, negative or uncertain results.

- **Ordinal tests**: The test can take different values (usually integer values), the set of possible outcomes is finite and, unlike the discrete diagnostic tests, an order can be established on the different test results. Tests that involve an element of

subjective assessment are often ordinal in nature (Pepe, 2004). For example, from the radiologists reading of an image the possible results to detect a disease could be "definitely", "probably", "possible" or "definitely not". Clinical symptoms are also used as diagnostic tools and they are often classified as "severe", "moderate", "mild" or "not present".

- **Continuous tests (or biomarkers)**: The test can take infinite (real) values. For example, glycated hemoglobin (HbA1c) as a marker for diabetes mellitus or leukocyte elastase to diagnose coronary artery disease (CAD).

In all these types of tests the objective is the same: to classify an individual as healthy or diseased depending on the absence or presence (respectively) of a particular symptom or sign, which is presumed to be indicative of the absence or presence of the condition under study. For binary tests, the outcome directly provides the classification rule. Nevertheless, in ordinal or continuous tests, the classification rule is a bit more complex.

The most common diagnostic tests used in clinical practice are either binary tests because of its simplicity and ease of handling, or continuous tests given the continuous nature of many analytical measurements. So, we will focus here on these two types of tests.

2.1 Binary tests

The simplest diagnostic test is one in which the test result is dichotomous, that is, it is a test that has only two possible outcomes, positive or negative for the disease. In this type of tests the classification is performed straightforward from the test result, each patient is classified as healthy or diseased depending on whether the test result is negative or positive, respectively, as in general a positive result is associated with the presence of disease and a negative result with the absence of disease.

Defining D as the random variable that models the true disease status or gold standard result, $D = 1$ indicates the patient has the disease, and $D = 0$ the patient does not have such a disease. The disease prevalence is the probability that a subject has the disease and it is normally denoted by p or π, that is, $p = \pi = \Pr(D = 1)$. Let Y be the random variable that models the test results in the population under study. The notation $Y+$ indicates a positive test result (tentatively indicating the presence of disease), and $Y-$ indicates a negative test result (tentatively indicating the absence of disease).

When a sample of patients is studied, the data obtained from the binary test allow us to classify the subjects into four groups according to a 2×2 table known as **confusion matrix** or **classification table**. For an example, see Table 2.1. In the confusion matrix, the diagnostic test result (in rows) is faced with the true disease status of the patient determined by the reference test or the gold standard (in columns).

The classification of the true status (healthy/diseased) of the patient based on the result of a diagnostic test is not error-free. So, it is necessary to measure the errors to check its validity on diagnosing, that is, to evaluate the **diagnostic accuracy** or ability to discriminate or differentiate between healthy and diseased populations. The test may

err or fail in its task of detecting the disease in two different ways, namely, by incorrectly classifying a healthy patient (a **false positive**, denoted by FP or $F+$) or, alternatively, by declaring a patient to be healthy when he or she is in fact diseased (a **false negative**, denoted by FN or $F-$). Conversely, the test may correctly classify a healthy patient (a **true negative**, denoted by TN or $T-$) or a diseased patient (a **true positive**, denoted by TP or $T+$).

Table 2.1: Classification of test results by disease status.

	True Disease Status (D)		
Test result (Y)	Diseased ($D = 1$)	Healthy ($D = 0$)	Total
Positive ($Y+$)	True Positive (TP)	False Positive (FP)	TP+FP
Negative ($Y-$)	False Negative (FN)	True Negative (TN)	FN+TN
Total	TP+FN	FP+TN	N= (TP+FP+FN+TN)

The accuracy of a diagnostic test to detect an individual with the disease or exclude an individual without the disease can be often measured in terms of different **accuracy measures**. Such measures must be individually considered by the clinicians in order to appropriately interpret the results that they obtain when performing the test on a patient (Bland, 1987; Murphy et al., 2006), but many clinicians are frequently unclear about the practical application of these terms (Steurer et al., 2002). In the following, the most important accuracy measures of a binary diagnostic test are presented, defined from the probabilities of the four possible events (collected in the confusion matrix, see Table 2.1) and the disease-specific classification probabilities (that is, the disease prevalence, $p = \pi = \Pr(D = 1)$, and its complementary, $1 - p = 1 - \pi = 1 - \Pr(D = 0) = \Pr(H)$, where H is an alternative notation to refer to the event $D = 1$).

2.1.1 Sensitivity and specificity

The validity of a diagnostic test in clinical practice (Feinstein, AR, 1975; Swets, 1979; Swets et al., 1979; Altman and Bland, 1994a) is usually measured by the conditional probabilities of correctly classifying a patient, the **sensitivity** and **specificity**, the most well-known accuracy measures. Yerushalmy (1947) introduced these two terms as statistical indicators for evaluating the degree of inherent efficacy of a diagnostic test in relation to a reference criterion that is considered to yield "the truth" (Feinstein, AR, 1975; Griner et al., 1981; Robertson et al., 1983; Fescina et al., 1985; Sox, 1986; Taube, 1986; Silva, 1987; Knottnerus and Leffers, 1992).

- **Sensitivity** (Se) is the probability of correctly classifying a diseased patient, that is, the probability that a diseased patient gets a positive test result (Bland, 1987; Altman, 1991; Hastie et al., 2000; Chien and Khan, 2001; Davidson, 2002). Sensitivity is, therefore, the ability of the test to detect the disease when it is present, that is, how "sensitive" is the test to the presence of the disease (Feinstein, AR, 1975; Griner et al., 1981; Robertson et al., 1983; Feinstein, 1985; Fescina et al., 1985; Sox, 1986; Taube, 1986; Hlatky et al., 1987; Silva, 1987; Sackett et al., 1989; Riegelman and Hirsch, 1991; Knottnerus and Leffers, 1992), and it is also known as true positive fraction (TPF) or true positive rate (TPR). Keep in mind that the sensitivity can only be calculated from those individuals who have the disease (Mayer, 2004).

This means that the sensitivity of a test tells us nothing about whether or not some people without the disease would test also positive and, if so, in what proportion (Akobeng, 2007). From the confusion matrix given in Table 2.1, it is easy to estimate the sensitivity as the proportion of diseased individuals who tested positive, that is:

$$Se = \Pr(Y+|D=1) = \Pr(TP) = TP/(TP+FN).$$

- **Specificity** (Sp) or true negative fraction (TNF) is the probability of correctly classifying a healthy patient, that is, the probability that for a healthy subject a negative test result is obtained (Bland, 1987; Altman, 1991; Hastie et al., 2000; Chien and Khan, 2001; Davidson, 2002). In other words, specificity can be defined as the ability of the test to exclude the disease in patients who are disease-free (Feinstein, AR, 1975; Griner et al., 1981; Robertson et al., 1983; Feinstein, 1985; Fescina et al., 1985; Sox, 1986; Taube, 1986; Hlatky et al., 1987; Silva, 1987; Sackett et al., 1989; Riegelman and Hirsch, 1991; Knottnerus and Leffers, 1992). Similarly to sensitivity, specificity can only be calculated from those people who do not have the disease, and therefore, specificity tells us nothing about whether or not some people with the disease would also have a negative result and, if so, in what proportion (Akobeng, 2007). From the confusion matrix given in Table 2.1, the specificity is estimated as:

$$Sp = \Pr(Y-|D=0) = \Pr(TN) = TN/(TN+FP).$$

The sensitivity and specificity of a binary diagnostic test can be estimated either from a single sample (cross sampling study) as from two independent samples of diseased and non-diseased individuals (case-control studies). The only difference is that in case-control sampling, the number of diseased individuals and healthy subjects, that is, $TP+FN$ and $TN+FP$ are set by the researcher.

Both sensitivity and specificity are probabilities of success in the diagnosis of disease. From these correct decisions, the probabilities of the corresponding classification errors or misclassifications are defined as follows:

- **1-sensitivity** ($1-Se$), also called false negative fraction (FNF), is the probability of incorrectly classifying a diseased patient:

$$1-Se = \Pr(Y-|D=1) = \Pr(FN) = FN/(TP+FN).$$

- **1-specificity** ($1-Sp$), also called false positive fraction (FPF), is the probability of incorrectly classifying a healthy patient:

$$1-Sp = \Pr(Y+|D=0) = \Pr(FP) = FP/(TN+FP).$$

The sum of the probability of a true positive (sensitivity) and the probability of a false negative is equal to one:

$$\Pr(Y+|D=1) + \Pr(Y-|D=1) = 1.$$

Similarly, the sum of the probability of a true negative (specificity) and the probability of a false positive is equal to one:

$$\Pr(Y-|D=0) + \Pr(Y+|D=0) = 1.$$

In the context of the statistical hypothesis testing of the null hypothesis (D = 0) versus the alternative hypothesis (D = 1), the sensitivity (TPF) is the statistical power β and 1-specificity (FPF) is the significance level α.

The sensitivity and specificity represent the efficacy of the diagnostic test and they depend only on the intrinsic ability of the test for distinguishing between healthy and diseased individuals. Therefore, both parameters depend on the physical, chemical, biological, ... basis of the test to discriminate between the two types of patients. It should be noted that a test with high sensitivity is useful to exclude the disease (that is, for "ruling out" the disease if an individual tests negative (Sackett et al., 2000; Davidson, 2002), since the probability of a false negative is low), and a test with high specificity is useful to confirm the disease (for "ruling in" the disease if an individual tests positive (Sackett et al., 2000; Davidson, 2002) since the probability of a false positive is low). The mnemonic SnNout (high Sensitivity, Negative test = rule out) and SpPin (high Specificity, Positive test, rule in) are useful ways of remembering these principles (Sackett et al., 2000). An ideal test has FPF = 0 and TPF = 1 and a useless test has $TPF = FPF$.

In general, screening tests should have a high sensitivity to detect all diseased subjects. A highly sensitive test will be specially useful in situations where not to diagnose the disease can be fatal for the diseased individuals, as in the case of serious but treatable diseases, such as tuberculosis or lymphoma, or diseases in which a false positive does not cause serious psychological or economic problems for the patient (for example, performing a mammography in breast cancer). Furthermore, confirmatory diagnostic testing must have a high specificity to avoid false positives. Tests with high specificity are needed in severe diseases for which there is no treatment available, when there is great interest in the absence of disease, or when diagnosing a patient who does not really have the disease can have severe consequences, whether psychological, physical or economic, for example, in the case of AIDS (Pita-Fernández and Pértega-Díaz, 2003). A test with high sensitivity is more useful to the clinicians when the test result is negative, while a high specificity test is more useful to the clinician when the test result is positive.

To illustrate the usefulness of such indexes or measures, we consider an example corresponding to real data of a follow-up study which includes a total of 2641 patients with suspected prostate cancer who attended an Urology consultation for a specified interval time. During their exploration, the result of the digital rectal examination (DRE) (the diagnostic test) performed to each patient was collected, according to normal or abnormal, and this result was compared with the subsequent diagnosis obtained from a prostate biopsy performed to each patient (the reference test or gold standard). Data from this study are presented in Table 2.2. By prostate biopsy a total of 1121 cases with prostate cancer were detected (the 42.45% of the patients enrolled). However, only for 903 patients the DRE was abnormal (the 34.19% of the patients enrolled). The DRE marker for the diagnosis of prostate cancer had a sensitivity $= 634/(634 + 487) = 634/1121 = 0.566$, so that the DRE was abnormal in 56.6% of cases with prostate cancer, that is, 56.6% of patients with prostate cancer were correctly classified by the DRE marker. Therefore, 43.44% (= 100% - 56.56%) of patients who actually had cancer, had a normal rectal examination, and so they were incorrectly classified as healthy patients. This clearly indicates the need for other more sensitive markers such as PSA or their derivatives to make the di-

agnosis more precise (Pita-Fernández and Pértega-Díaz, 2003). In addition, a specificity $= 1251/(1251 + 269) = 1251/1520 = 0.823$ was obtained, which means that DRE was normal in 82.3% of cases who did not finally developed prostate cancer, that is, 82.3% of patients without cancer were correctly classified.

Table 2.2: Results of exploration and prostate biopsy of a sample of patients with suspicious of prostate cancer.

	Prostate biopsy		
Rectal exam	Diseased	Healthy	Total
Positive or abnormal	634	269	903
Negative or normal	487	1251	1738
Total	1121	1520	2641

The concepts of sensitivity and specificity allow therefore, assess the validity of a diagnostic test. Despite the fact that such measures are considered the fundamental operating characteristics of a diagnostic test, in some situations they may lack utility in clinical practice because they can not be used to estimate the probability of disease for each patient. Both the sensitivity and specificity provide information about the probability of a particular result (positive or negative) based on the true disease status of the patient. However, in general, when a diagnostic test is performed to a patient, he (or she) presents a series of symptoms rather than a diagnosis, the clinician (or observer of a diagnostic test) does not have a priori information about his (or her) true diagnosis, and sometimes, it is more interesting to answer the questions in reverse, that is, what is the probability that an individual who has tested positive is actually diseased, and vice-versa, that is, what is the probability that an individual who has tested negative is disease-free. Sensitivity and specificity cannot be used to answer these two questions, but those two probabilities can be computed from the sensitivity and specificity applying the Bayes theorem, whenever the "a priori" probability that the subject is diseased is known beforehand, which is also known as the pre-test probability. If we have no additional information on the subject, the pre-test probability will be considered equal to the disease prevalence, which is only applicable in the case of screening programs designed for the general population. In clinical practice, when a diagnostic test is performed to a patient is because there is the suspicions from previous symptoms and, therefore, the probability of disease under suspicion will be higher than the disease prevalence. The **predictive values** are the measures that allow us to answer these questions.

2.1.2 Predictive values

The accuracy of a diagnostic test can be quantified by how well the test result predicts true disease status, that is, it makes a diagnosis (Altman and Bland, 1994b). The **predictive values** allow us to measure the safety of the diagnostic test, that is, the probability that the known specific test result, positive or negative, ensures the presence or absence of disease, respectively. Based on this, the positive and negative predictive values can be defined as follows (Feinstein, AR, 1975; Griner et al., 1981; Robertson et al., 1983; Fescina et al., 1985; Sox, 1986; Taube, 1986; Silva, 1987; Knottnerus and Leffers, 1992).

- The **positive predictive value** (PPV) or the predictive value of a positive test is the

probability of developing the disease if a positive test result is obtained (Altman, 1991; Hastie et al., 2000; Chien and Khan, 2001; Davidson, 2002). PPV is, sometimes, also referred to as the "post-test probability of disease given a positive test". Therefore, the positive predictive value can be estimated based on the proportion of patients with a positive test result that ultimately proved to be diseased. From Bayes theorem, the PPV is related to the sensitivity and specificity measures, and the prevalence of the disease under study (Last, 2001):

$$PPV = \Pr(D = 1|Y+) = \frac{pSe}{pSe + (1-p)(1-Sp)} = TP/(TP + FP).$$

- The **negative predictive value** (NPV) or the predictive value of a negative test is the probability that a subject who tested negative is actually healthy (Altman, 1991; Hastie et al., 2000; Chien and Khan, 2001; Davidson, 2002). It is estimated as the ratio of the number of true negatives to the total number of the patients who tested negative, and similarly to PPV, it can be estimated based on sensitivity and specificity measures, and disease prevalence by Bayes theorem (Last, 2001):

$$NPV = \Pr(D = 0|Y-) = \frac{(1-p)Sp}{(1-p)Sp + p(1-Se)} = TN/(TN + FN).$$

It can be observed that the roles of true disease status D and diagnostic test Y are reversed in the predictive values relative to their roles in the classification probabilities.

A perfect test (that will predict disease perfectly) has PPV = 1 and NPV = 1. On the other hand, a useless test has no information about the true disease status of the patients, so that $\Pr(D = 1|Y+) = \Pr(D = 1)$ and $\Pr(D = 0|Y-) = \Pr(D = 0)$, that is, $PPV = p = \pi$ and $NPV = 1 - p = 1 - \pi$.

From the positive and negative predictive values that measure the probabilities of a correct prediction of the true disease status of the patient, the probabilities of performing an incorrect prediction can be also defined:

- **1-positive predictive value** ($1-PPV$), also called the false discovery rate (FDR), is the post-test probability of not having the disease if a positive test result is obtained. Similarly, it is related to measures of sensitivity and specificity, and the prevalence of the disease. According to Bayes theorem, it follows that:

$$1 - PPV = \Pr(D = 0|Y+) = \frac{p(1-Se)}{(1-p)(1-Sp) + p(1-Se)} = FP/(TP + FP).$$

- **1-negative predictive value** ($1 - NPV$), also called the false omission rate (FOR), is the post-test probability of disease if a negative test result is obtained. According to Bayes theorem, it also relates to sensitivity and specificity measures, and disease prevalence:

$$1 - NPV = \Pr(D = 1|Y-) = \frac{(1-p)(1-Sp-Se)}{pSe + (1-p)(1-Sp)} = FN/(TN + FN).$$

The predictive values can be estimated without any problems in a cross sectional study, because the sensitivity, specificity and prevalence can be estimated in such studies. The prevalence can be estimated from the sample prevalence as the ratio between the number of diseased cases in the sample and the total number of cases. However, in a case-control study, we need to have an estimate of the disease prevalence taken from other studies, since in a case-control study the total number of healthy and diseased individuals is set by the researcher and, consequently, the sample prevalence would not be a good estimate of the true prevalence.

If we consider the previous example of prostate cancer, the negative predictive value was NPV = 1251/1738 = 0.7198. So, 72% of the patients with no abnormality test result (normal DRE) were actually healthy, not suffering from prostate cancer. The positive predictive value was PPV = 634/903 = 0.7021. This means that 70% of patients with normal DRE were indeed diseased individuals suffering from prostate cancer.

As we have seen, the values of sensitivity and specificity allow define the validity of a diagnostic test, to measure how well the test reflects the true disease status, but have the disadvantage of not providing relevant information when making a clinical decision. The patient and the caregiver are most interested in how likely is that the individual is diseased given the test result. However, Se and Sp have the additional advantage that they are intrinsic properties to the diagnostic test and they define the diagnostic test validity regardless of the disease prevalence in the study population.

By contrast, predictive values PPV and NPV are extremely useful when making clinical decisions and transmit information to patients about their diagnosis, quantifying the clinical value of the test. Therefore, it seems natural to use them as comparison indices when evaluating two different diagnostic methods. However, as we have previously discussed, they depend on the disease prevalence in the study population, and so, it will be wrong for clinicians to directly apply published predictive values of a test to their own populations if the disease prevalence differs from that.

The higher the disease prevalence, the higher the PPV and the lower the NPV, that is, the more likely a positive result is able to predict the presence of disease. When the prevalence of disease is low, a negative result rules out the disease more safely, with greater negative predictive value; but a positive result does not allow to confirm the diagnosis, resulting in a low positive predictive value, even when using a test with high sensitivity and specificity (Altman, 1991). So, a low PPV may simply be a result of a low prevalence or it may be due to a test that does not reflect the true disease status of the subject very well. Hence, we can say that either a diagnostic test is useful to confirm the disease if the positive predictive value is high, or it is useful to rule out the disease if the negative predictive value is high, or it is useful for both reasons if the two predictive values are high.

Zweig and Campbell (1993) did an in depth discussion of the concepts of accuracy versus usefulness. In many biomedical research studies, both the disease-specific classification probabilities and the predictive values are reported; what this means in practice is that the clinical usefulness of the diagnostic test for a patient depends on the prevalence of the disease in the study population. The diagnostic value of a test is greatly improved if based in the clinical assessment, that is, if the use of the test is limited to those patients

who are likely to have the disease. So, a positive or negative test result is more likely to be meaningful in that setting than when the test is applied indiscriminately to patients. A diagnostic test should be used as a supplement rather than as a substitute for clinical judgment (Akobeng, 2007).

As prevalence is a determinant factor in the predictive values of a test, they can not be used as indices when comparing two different procedures nor for extrapolating the results of other studies to our own data or vice versa. Thus, it may be useful and necessary to determine indices of valuation (responding to the real needs for patient classification, that is, being clinically useful), that are not dependent on the proportion of diseased in the sample. Besides, the use of sensitivity, specificity and predictive values alone can occasionally lead clinicians to make misleading inferences regarding the value of a clinical test and the results derived from it when used in clinical practice (Chien and Khan, 2001; Bruno et al., 2011). So, in addition to the concepts of sensitivity, specificity and predictive values, we often talk in terms of **likelihood ratios** (Sackett et al., 1985; Boyko, 1994; Dujardin et al., 1994; Boyd, 1997).

2.1.3 Likelihood ratios

Other measures of the performance of a binary diagnostic test are the **likelihood ratios** (LRs). Feinstein (1985) describes the LR as a "recent and popular" indicator of the effectiveness of a diagnostic test. Some authors (Sackett et al., 1985; Boyko, 1994; Boyd, 1997) have discussed its usefulness in the evaluation of diagnostic tests versus other measures and its application in clinical research (Giard and Hermans, 1993; Kerlikowske et al., 1996). These measures are likelihood ratios in the true statistical sense, although they are also called diagnostic likelihood ratios ($DLRs$) to distinguish the context. The diagnostic likelihood ratios measure how much more likely is an specific result (positive or negative) according to the presence or absence of disease, and they can also be used to compute the probability of disease for single patients (Deeks and Altman, 2004). So, they are literally, the ratios of the likelihood of the observed test result in the diseased versus non-diseased populations (Halkin et al., 1998; Pepe, 2004) and it is a similar concept to the relative risk as used in modern epidemiology. So, for a test result i, the likelihood ratio is a ratio of two probabilities defined as:

$$DLR = \frac{\Pr(Y = i|D = 1)}{\Pr(Y = i|D = 0)}.$$

There are two likelihood ratios, two dimensions of accuracy, one for a positive test and one for a negative test. If the test result is positive, the likelihood ratio is called **positive likelihood ratio**, while if the test result is negative, the **negative likelihood ratio** is obtained.

- The **positive diagnostic likelihood ratio** ($DLR+$) is computed by dividing the probability of a positive test result in diseased individuals between the probability of a positive result among healthy. Note that it can be obtained in terms of sensitivity and specificity measures as the ratio between the true positive fraction (sensitivity) and the false positive fraction (1-specificity). So, the positive diagnostic likelihood ratio is defined as:

$$DLR+ = \Pr(Y+|D = 1)/\Pr(Y+|D = 0) = \frac{Se}{1 - Sp}.$$

The $DLR+$ responds therefore to the question of how many times more likely the test is positive in diseased patients than in non-diseased.

- The **negative diagnostic likelihood ratio** ($DLR-$) is computed by dividing the probability of a negative test result in the presence of the disease between the probability of a negative result in the absence of it. It is estimated, therefore, as the ratio between the false negative fraction (1-sensitivity) and true negative fraction (specificity):

$$DLR- = \Pr(Y-|D=1)/\Pr(Y-|D=0) = \frac{1-Se}{Sp}.$$

The $DLR-$ responds therefore to the question of how many times more likely the test is negative in diseased patients than in non-diseased.

It is important to consider both $DLRs$ in order to have a complete idea of the effectiveness of the test. Therefore, although the sensitivity and specificity can not be used to estimate the probability of disease of an individual, they can be combined into a single measure, the likelihood ratio, which can be clinically more useful than the sensitivity and specificity indexes.

Likelihood ratios take values between zero and infinite and do not dependent on the disease prevalence. When the diagnostic test and the gold standard are independent, then both likelihood ratios are equal to one (the test is useless, because it does not alter the probability), and if the diagnostic test correctly classifies all patients (having or not the disease), then $DLR+ = \infty$ (it is not possible to specify an upper limit for the $DLR+$) and $DLR-$ = 0. If the likelihood ratios are greater than one, it indicates an increase in the disease probability. Higher values of $DLR+$ indicate improved ability to diagnose the presence of the disease. Thus, a value $DLR+ > 1$ indicates that a positive test result is more likely in a diseased subject than in a non-diseased subject, and a $DLR- \geq 1$ means that a negative test result is more likely in diseased people than in non-diseased people. A DLR less than 1 indicates a decreased likelihood of disease; thus, a $DLR- < 1$ indicates that a negative test result is more likely in a non-diseased patient than in a diseased patient, and a $DLR+ < 1$ indicates that a positive test is less likely in a diseased patient than in a non-diseased patient. In general, one can say that for patients with a positive test result, values of $DLR+ \geq 10$ significantly increase the likelihood of disease ("rule in" disease) while values of $DLR+$ very low ($DLR+ \leq 0.1$) virtually "rule out" the chance that an individual has the disease (Jaescheke et al., 2002). In a similar way, for patients who had a negative test result, $DLR-s \geq 10$ significantly increase the probability of disease ("rule in" disease) while values of $DLR-$ very low ($DLR- \leq 0.1$) virtually "rule out" the chance that an individual has the disease (Jaescheke et al., 2002).

If we consider again the above example of prostate cancer, we obtain that the digital rectal exam for prostate cancer diagnosis has a positive diagnostic likelihood ratio equal to $DLR+$ = 0.566/(1-0.823) = 3.20 and a negative diagnostic likelihood ratio equal to $DLR-$ = (1-0.566)/ 0.823 = 0.53. This means that any patient with a positive test result would be three times more likely to have than not to have prostate cancer, and that a negative result is almost twice (1/0.53 = 1.89) more likely in a patient without prostate

cancer than in a patient with the disease.

One of the main advantages of likelihood ratios is that they can be used to help the clinician when adjusting the values of sensitivity and specificity of the test to individual patients. The likelihood ratios quantify the increase in knowledge about presence of disease that is gained through the application of the diagnostic test, as discussed below.

Before performing the diagnostic test (that is, in the absence of the test result) usually the clinician has a rough estimate of the patient's chance, probability or odds of having the disease. The estimated probability of disease before the test result is known as the pre-test probability, which is usually estimated on the basis of the local prevalence, the clinician's experience and published reports (Espallardo, 2003):

$$\text{pre-test odds} = \frac{\Pr(D=1)}{\Pr(D=0)} = \frac{p}{1-p}.$$

The most important reason why the clinician performs the test is to obtain further information which may modify the pre-test probability of disease. So, a negative test may reduce the pre-test probability and a positive test may increase such probability. The probability that the patient has the disease after the test result is known as the post-test probability. This is the probability that clinicians and patients are more interested in, and it can serve as a guide or aid in deciding whether to confirm a diagnosis, rule out a diagnosis or even perform further and additional tests. Therefore, after the test is performed, that is, with knowledge of the test result, the odds of disease can be defined as:

$$\text{post-test odds} = \frac{\Pr(D=1|Y=i)}{\Pr(D=0|Y=i)}.$$

Also, if we know or we can make an estimate of the pre-test probability of a subject suffering from the disease, once the test result is known, we can "correct" the value of that probability using likelihood ratios according to the result obtained (in such a way that it increases or decreases depending on whether the result is positive or negative), using the following formula:

$$p - post = \frac{pDLR}{1 + p(DLR - 1)},$$

where p is the pre-test probability, DLR is the likelihood ratio (positive if we want to compute the probability that the patient has the disease and negative otherwise) and $p - post$ is the post-test probability.

In the previous example of prostate cancer, if we assume that the pre-test probability of having the disease is 0.7, and the result of DRE is positive for the diagnosis of prostate cancer, it follows that the probability of having the disease has changed to 0.88. If the test result was, however negative, this probability would decrease to 0.55.

Therefore, due to everything mentioned above, the likelihood ratios relate the pre-test to post-test odds and quantify the change in the odds of disease obtained by taking into account the test result. They are also called Bayes factors, because if we consider

$\Pr(D = 1)$ as the prior probability of disease and $\Pr(D = 1|Y)$ as the posterior probability in a Bayesian sense, the DLR is the Bayesian multiplication factor relating the prior and posterior distributions. The positive likelihood ratio is the multiplicative change in the post-test odds of being sick after having a positive test result and the negative likelihood ratio is the multiplicative change in the post-test odds of not being diseased after a negative test result:

$$\text{post-test odds } (Y+) = DLR+ \text{(pre-test odds)},$$

$$\text{post-test odds } (Y-) = DLR- \text{(pre-test odds)}.$$

Likelihood ratios offer the advantage to relate the sensitivity and specificity in one index, and are available according to several levels of a new measure. Moreover, they are motivated by the concept of predicting disease status from the test result, as the predictive values. Indeed, the post-test odds also can be defined in terms of the predictive values:

$$\frac{PPV}{1 - PPV} = \text{post-test odds } (Y+),$$

$$\frac{1 - NPV}{NPV} = \text{post-test odds } (Y-).$$

In the sense that they relate to prediction but do not depend on the population prevalence, they can be considered as a compromise between classification probabilities and predictive values. However, the advantage of the $DLR+$ and $DLR-$ compared to positive and negative predictive values is that, unlike them, do not depend on the proportion of patients in the sample, but only on the classification probabilities of sensitivity and specificity, hence likelihood ratios are useful to compare different tests for the same diagnosis. So, if we compare two diagnostic tests A and B, we can make this comparison based on the corresponding likelihood ratio. For instance, if we have $DLR+_A > DLR+_B$, then the test A is better than B to confirm the presence of the disease. In addition, if we have that $DLR-_A < DLR-_B$, then the test A is better than B to confirm the absence of disease.

2.1.4 Youden index

The **Youden index** (Youden, 1950) (denoted in the literature by J or YI) is another measure of the effectiveness of a diagnostic test defined as

$$J = Se + Sp - 1 = Se - (1 - Sp),$$

in order to maximize the effectiveness of the biomarker. J takes values between 0 and 1 and is the maximum difference between the probability of correctly classifying diseased subjects or true positive rate (sensitivity) and the probability of misclassifying healthy individuals or false positive rate (one minus the specificity).

This index has the following properties:

- When the test is independent of the disease, the specificity and sensitivity are complementary, that is, if the test and gold standard are independent: $Se + Sp = 1$. Therefore in this case a value of $J = 0$ is obtained, that is, it corresponds to a completely ineffective test to diagnose the disease.

- Youden index measures the discrepancy of the diagnostic test with the independence of the gold standard, so that the maximum discrepancy is obtained when $Se = Sp = 1$. In this situation we talk of total dependence, providing a value of $J = 1$ that indicates a perfect diagnostic effectiveness.

- The situation where both probabilities are zero results in a situation that makes no sense in practice.

- When the diagnostic test is related to the disease (a desirable situation for a diagnostic test), then you get that $Se + Sp > 1$. This result is very important because otherwise, that is, in case that $Se + Sp < 1$, one of the two terms has inevitably be less than 0.5, which implies that the test fails more than a random allocation rule.

The Youden index is a measure of the accuracy of a diagnostic test that does not depend on the disease prevalence and is preferred to the combination of the simple values of sensitivity and specificity (Feinstein, SH, 1975; Feinstein, 1985), but it has the disadvantage of a more difficult interpretation since the idea of whether the diagnostic test is good in sensitivity or specificity is lost. Feinstein, SH (1975) based this statement on the following example: if the Youden index is equal to 0.55, it may happen that the sensitivity is 0,95 and that the specificity is 0.60, or vice versa.

2.2 Continuous tests

So far we have studied the simplest test with a dichotomous outcome (positive or negative), but often the confirmation of a diagnosis must be made from a quantitative diagnostic test, especially when it is a laboratory test or an analytical determination, such as measuring blood glucose levels in an individual. A continuous diagnostic test takes values in a continuous range or scale. In this type of tests, the following question arises: how could we classify an individual as healthy or disease from the values of Y? When the result of the diagnostic test Y is ordinal or continuous, the classification rule is more complex than in the case of a binary test. In continuous diagnostic tests it is necessary to select a **cutpoint** or **discrimination value** c, a certain value of the test, that defines the boundary between healthy and diseased, that is, to allow us to define the positive and negative test results. In general, without loss of generality, we will assume that higher values of the diagnostic test are associated with disease. Under this assumption, you have the following classification rule (see Figure 2.1): if the diagnostic test value Y is equal to or higher than the cutpoint c, the patient is classified as diseased (positive test result $Y+$), whereas if the test value is lower than c, the patient is classified as healthy (negative test result $Y-$). If this assumption is not valid, it would suffice to change the sign of the diagnostic test.

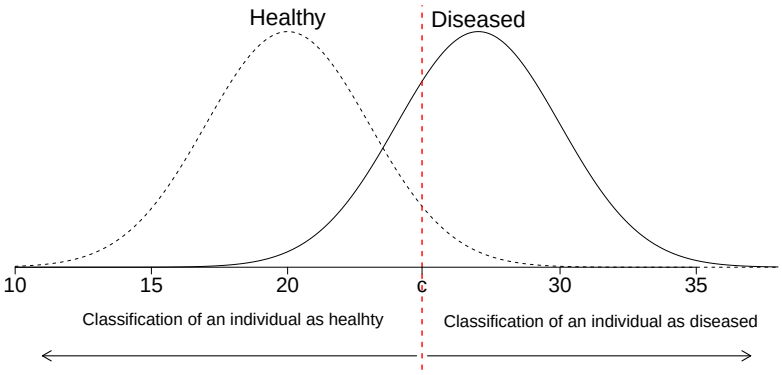


Figure 2.1: Classification of an individual as healthy or diseased as function of a fixed cutoff c.

Analogously to a binary diagnostic test, this classification is not error-free and the continuous diagnostic test may fail in its task of detecting disease (see Table 2.3). Therefore, before routine application of a continuous diagnostic test in practice, the misclassification errors must be quantified.

Table 2.3: Classification of continuous test results by disease status.

	True Disease Status (D)	
Test result (Y)	Diseased ($D = 1$)	Healthy ($D = 0$)
Positive ($Y \geq c$)	True Positive (TP(c))	False Positive (FP(c))
Negative ($Y < c$)	False Negative (FN(c))	True Negative (TN(c))
Total	TP(c)+FN(c)	FP(c)+TN(c)

Obviously, each choice of c gives rise to a binary test, and then you can use all the considerations made so far for a binary test. For instance, with respect to any given cutpoint, different measures of accuracy can be considered such as the the sensitivity and specificity (see Figure 2.2). The essential difference now is that you do not have a single pair of **sensitivity** and **specificity** values that define the accuracy of the test, but rather a set of pairs of values corresponding each at a different cutpoint c:

$$Se(c) = TPF(c) = \Pr(Y + | D = 1) = \Pr(Y \geq c | D = 1),$$

$$Sp(c) = TNF(c) = \Pr(Y - | D = 0) = \Pr(Y < c | D = 0).$$

From the above measures, you can define the corresponding probabilities of incorrect classifications of diagnosis (see Figure 2.3):

$$1 - Se(c) = FNF(c) = \Pr(Y < c | D = 1),$$

$$1 - Sp(c) = FPF(c) = \Pr(Y \geq c | D = 0).$$

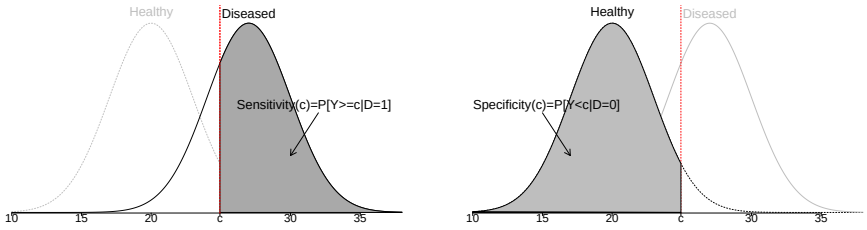


Figure 2.2: Sensitivity and specificity measures of the diagnostic test Y for the cut-off value c.

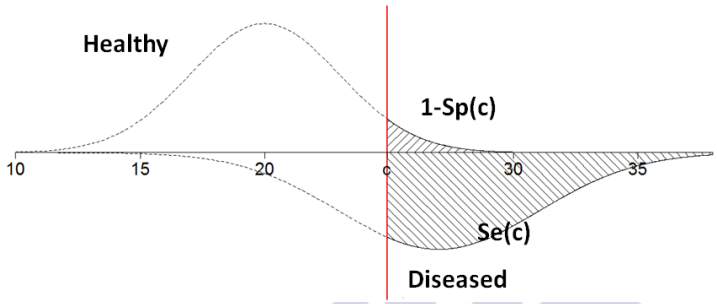


Figure 2.3: Sensitivity and 1-specificity (FPF) of the diagnostic test Y for the cut-off value c.

In continuous diagnostic tests, the selection of an optimal cutoff point is the most important task. However, this choice is not an easy task. As is illustrated in Figure 2.4, different choices of c will differ for all kinds of correct and incorrect diagnosis decisions frequencies. In Chapter 3 we present a review of many of the existing criteria in the literature for selecting the optimal threshold value in the sense of different optimality criteria.

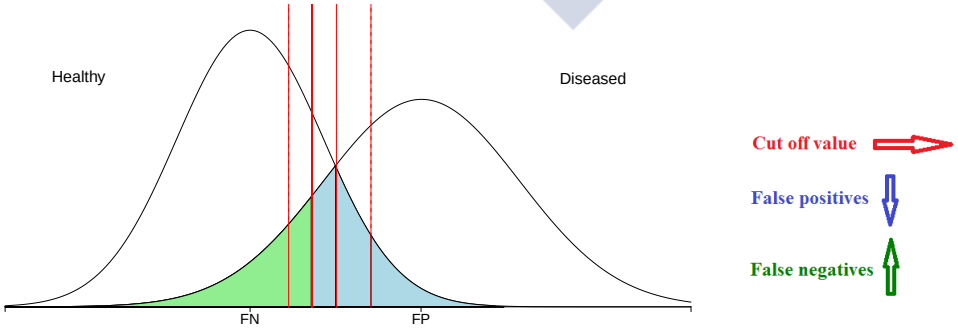


Figure 2.4: Healthy and diseased distributions and variation of the cutoff point depending on the different diagnostic decisions.

In general, there is a range of potential test results for which the distributions of healthy subjects and diseased overlap. If you want to increase the probability of detecting diseased patients, that is, if the cutpoint is moved to the left, then the number of false positives also increases. On the contrary, if the cutoff is moved to the right, false positives will be reduced, but at the cost of increasing false negatives. Therefore, as the sensitivity decreases the specificity increases and vice versa. Hence, when selecting the "best" cutoff c, a balance between sensitivity and specificity measures will be sought.

Similarly to a binary test, on the basis of the specificity and sensitivity measures, other accuracy measures can be also defined for each cutpoint c, such as positive and negative predictive values (PPV and NPV) and diagnostic likelihood ratios ($DLR+$ and $DLR-$):

- **Positive Predictive Value** (PPV):

$$PPV(c) = \Pr(D = 1|Y \geq c) = \frac{pSe(c)}{pSe(c) + (1-p)(1-Sp(c))}.$$

- **Negative Predictive Value** (NPV):

$$NPV(c) = \Pr(D = 0|Y < c) = \frac{(1-p)Sp(c)}{(1-p)Sp(c) + p(1-Se(c))}.$$

- **Diagnostic positive likelihood ratio** ($DLR+$):

$$DLR+ = \frac{\Pr(Y \geq c|D = 1)}{\Pr(Y \geq c|D = 0)} = \frac{Se(c)}{1-Sp(c)}.$$

- **Diagnostic Negative Likelihood Ratio** ($DLR-$):

$$DLR- = \frac{\Pr(Y < c|D = 1)}{\Pr(Y < c|D = 0)} = \frac{1-Se(c)}{Sp(c)}.$$

We remind here that predictive values measure the safety of the diagnostic test and they depend on the disease prevalence. Besides, from positive and negative predictive values that measure the probabilities of a correct prediction of the disease status, you can also define the probabilities of making a wrong prediction:

$$1 - PPV(c) = \Pr(D = 1|Y < c) = \frac{(1-p)(1-Sp(c)-Se(c))}{pSe(c) + (1-p)(1-Sp(c))},$$

$$1 - NPV(c) = \Pr(D = 0|Y \geq c) = \frac{p(1-Se(c))}{(1-p)(1-Sp(c)) + p(1-Se(c))}.$$

Note that all these definitions of the measures of accuracy of a diagnostic test are valid under the conventional assumption that larger values of Y are more indicative of disease. In probabilistic terms, this means that the random variable Y in the diseased group is stochastically greater than the random variable Y in the healthy group. It should be noted however that, when high marker values are linked to health, a positive test result would be considered when $Y \leq c$ and the above definitions should be changed accordingly. Another possibility for not to change the previous definitions is to consider $-Y$

instead of Y. All these measures constitute the classic indexes for assessing the diagnostic accuracy of a test. However, these indices have the disadvantage of depending on the cutoff c considered to define the binary rule of classification (Turner, 1978).

Table 2.4: Classification of elastase results by CAD status for cutpoint $c = 22\ \mu gl^{-1}$.

	CAD		
elastase	Diseased	Healthy	Total
Positive (elastase $\geq 22\ \mu gl^{-1}$)	92	28	120
Negative (elastase $< 22\ \mu gl^{-1}$)	4	17	21
Total	96	45	141

In order to illustrate these measures, we consider here the coronary artery disease dataset, previously introduced in Section 1.3 of Chapter 1. For instance, considering the use of leukocyte elastase determination as a screening test prior to performing a coronary angiography for detecting CAD, one would be interested in high sensitivity in order to identify as many diseased patients as possible. In Table 2.4 we show the confusion matrix for this cutpoint $c = 22\ \mu gl^{-1}$, which provides a sensitivity of 0.96, and consequently we could conclude that a coronary angiography would be performed on any patient having a leukocyte elastase level greater or equal to 22 μgl^{-1}. Using this cutpoint, 96% of CAD patients would be correctly classified (sensitivity), whereas only 38% (specificity) of patients without CAD would be correctly identified (28 false positive classifications). Moreover, 81% of the patients (the negative predictive value) with elastase below 22 μgl^{-1} are correctly classified as healthy (4 false negative results); and 77% of the patients (the positive predictive value) with an elastase value greater than or equal to 22 μgl^{-1} had in fact CAD. In addition, any patient having an elastase value $\geq 22\ \mu gl^{-1}$ would be 1.5 times more likely to have CAD than to not ($DLR+$ = 1.54), and a negative result of the test, that is, an elastase value $< 22\ \mu gl^{-1}$ is more likely in a patient without CAD than in a diseased patient, in particular the probability of a negative result is nine times higher in healthy patients than in patients with CAD ($1/DLR-$=1/0.11 = 9.09).

Table 2.5: Classification of elastase results by CAD status for cutpoint $c = 38\ \mu gl^{-1}$.

	CAD		
elastase	Diseased	Healthy	Total
Positive (elastase $\geq 38\ \mu gl^{-1}$)	65	15	80
Negative (elastase $< 38\ \mu gl^{-1}$)	31	30	61
Total	96	45	141

If we consider, however, a higher cutpoint, for example, an elastase value of 38 μgl^{-1}, that provides an equilibrium between both measures of sensitivity and specificity, 68% of patients with CAD and 67% of patients without CAD would be correctly detected. Moreover, about half of the patients with a negative test result (elastase $< 38\ \mu gl^{-1}$) are really free of CAD. However, 81% of patients with a positive test result (elastase ≥ 38 μgl^{-1} are really diseased. In addition, the probability of negative test result is twice as high in patients without CAD than in patients with CAD, and the patients with a positive test result would be two times more likely to have CAD than not. Thus, we

confirm that when increasing the cutoff point, the specificity, the false negative rate, the positive predictive value and the likelihood ratios increase, at the cost of decreasing the sensitivity, the false positive rate, and the negative predictive value. In Table 2.5 the corresponding confusion matrix is shown.

2.2.1 Receiver Operating Characteristic curve

An useful tool for assessing the diagnostic capacity of a quantitative test for all possible cutoff values of a continuous diagnostic test is called the **Receiver Operating Characteristic (ROC) curve** (Swets, 1979; Swets and Pickett, 1982). The ROC curve is constructing by plotting the sensitivities in the y-axis and the corresponding false positive fractions (1-sensitivity) in the x-axis:

$$ROC(.) = (1 - Sp(c), Se(c)), c \in (-\infty, \infty).$$

In Figure 2.5, we show the typical shape of the ROC curve of a continuous diagnostic test whose graph is in the unit square.

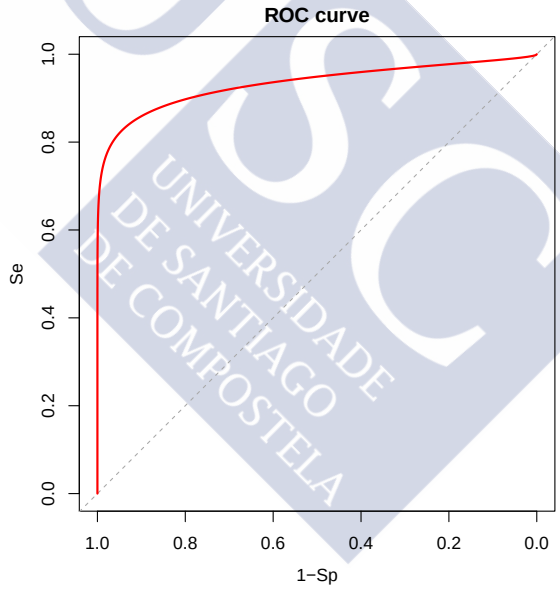


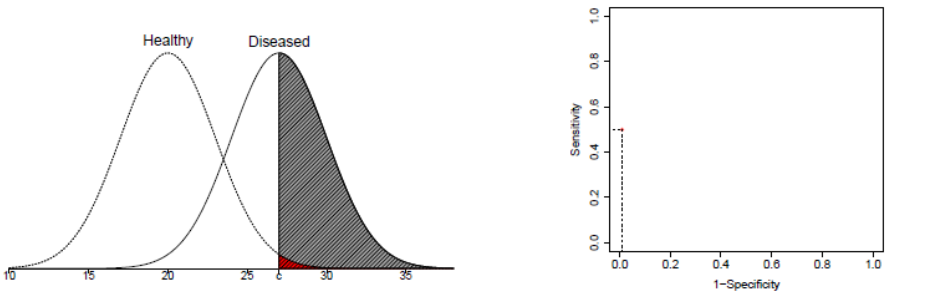
Figure 2.5: ROC curve of a continuous diagnostic test.

The construction process of the ROC curve is then made from different cutoff values c, in such a way that if the cutoff c moves from left to right, the operating points on the ROC curve are generated from right to left (see Figure 2.6).

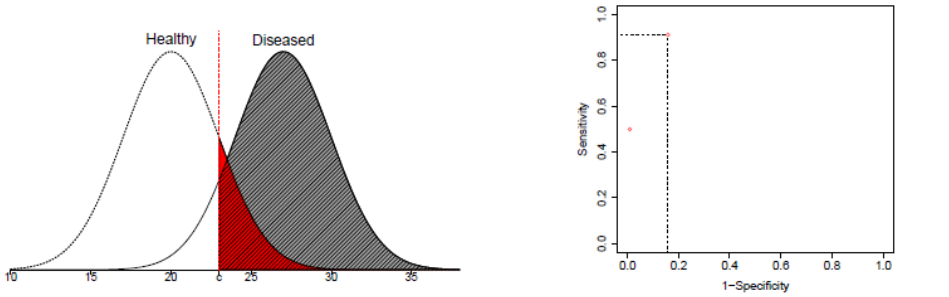
If the ROC curve lies below the diagonal, the diagnostic test distribution has the wrong orientation and a reversal is needed (either allocating to the diseased group when $Y \leq c$ rather than when $Y \geq c$ or considering $-Y$ rather than Y without having to interchange the allocating rule between groups).

The ROC curve was originated in the theory of signal detection in the years 1950-1960 (Green and Swets, 1966; Egan, 1975) to discriminate between signal and noise, and

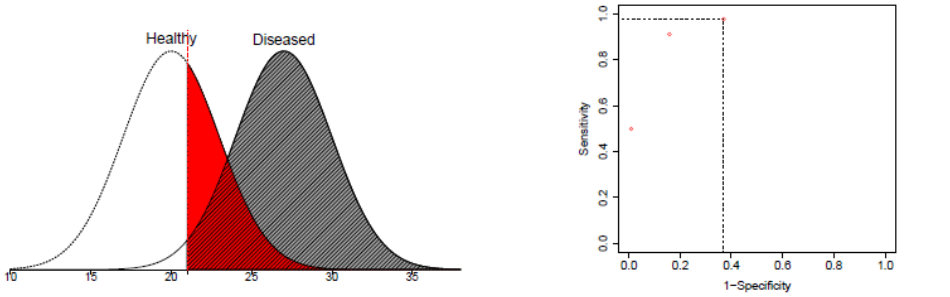
its potential to be used in diagnostic tests was recognized in 1960 (Lusted, 1960). In fact, after the publication of Swets and Pickett (1982), they became very popular in the field of biomedicine, especially in the field of radiology (Metz, 1989), but they have been used in many other scientific areas such as psychiatry (Hsiao et al., 1989), epidemiology (Aoki et al., 1997) or biomedical informatics (Lasko et al., 2005), among others. It has recently increased its use in biomedical problems to assess the effectiveness of continuous diagnostic markers to differentiate between groups of healthy and diseased individuals (Shapiro, 1999; Greiner et al., 2000; Zhou et al., 2002; Pepe, 2004).



Sensitivity and *1-Specificity* for $c = 27$ | *1-Specificity* vs. *Sensitivity* for $c = 27$



Sensitivity and *1-Specificity* for $c = 23$ | *1-Specificity* vs. *Sensitivity* for $c = 23$



Sensitivity and *1-Specificity* for $c = 21$ | *1-Specificity* vs. *Sensitivity* for $c = 21$

Figure 2.6: Generating process of ROC curve.

The ROC curve can be represented in terms of the cdfs of the diagnostic test in both populations (healthy and diseased) as follows:

$$ROC(t) = 1 - F_D(F_{\overline{D}}^{-1}(1-t)), t \in (0,1),$$

where F_D and $F_{\overline{D}}$ are the distribution functions of the diagnostic test in the diseased (D) and non-diseased ($\overline{D}$) populations, that is,

$$F_D(y) = \Pr(Y \leq y | D = 1),$$

$$F_{\overline{D}}(y) = \Pr(Y \leq y | D = 0).$$

A graphical example can be seen in Figure 2.7.

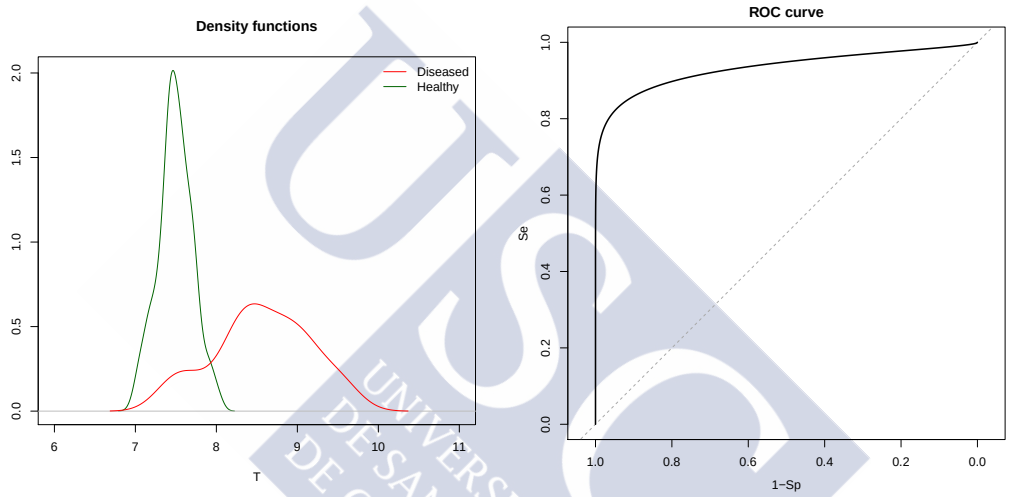


Figure 2.7: Diagnostic test distributions in healthy and diseased populations and the corresponding ROC curve.

Equivalently, the ROC curve can be represented in terms of survival functions (defined as one minus the distribution functions) as follows:

$$ROC(t) = S_D(S_{\overline{D}}^{-1}(t)), t \in (0,1),$$

where S_D and $S_{\overline{D}}$ are the survival functions of the diagnostic test in the diseased (D) and healthy ($\overline{D}$) populations, that is,

$$S_D(y) = \Pr(Y \geq y | D = 1),$$

$$S_{\overline{D}}(y) = \Pr(Y \geq y | D = 0).$$

This last representation is very important in a ROC regression context.

The ROC curve has also a number of interesting properties:

- The ROC curve is a monotone increasing function defined in the unit square $[0,1]x[0,1]$.
- The ROC curve joins the points (0,0) and (1,1).

- The ROC curve is invariant to strictly increasing transformations of the marker Y.

From a practical point of view, the ROC curve has a very simple and interesting interpretation:

- It is a global measure of the discriminatory ability (accuracy) of a continuous (or ordinal) diagnostic test, independent of the cutoff point and the disease prevalence, that fulfills the two conditions set by Swets (1988) as indispensable characteristics of any indicator of the diagnostic accuracy.

- It represents the discriminative ability of a diagnostic test along all possible cutoffs in the scale or range of the diagnostic marker.

- It serves as a guide when selecting the optimal cutoff point in diagnostic tests.

- It provides a method for comparing different diagnostic tests.

- It reflects the degree of overlapping of test results in both healthy and diseased populations. When there exists total overlapping between healthy and diseased distributions, the diagnostic test is useless (that is, it is not valid to differentiate between the two populations) and the corresponding ROC curve coincides with the positive diagonal line of the unit square since for any cutpoint c, it is satisfied that $Se(c) = 1 - Sp(c)$. However, when the overlap between the two distributions is null, the test is perfect, that is, it discriminates perfectly between the two groups of healthy and diseased individuals. In this ideal situation, the corresponding ROC curve is a line that would pass through the left and top sides of the unit square because it happens that for any point either $Se = 1$ or $1 - Sp = 0$, having a cutoff point where both situations are met ($Se = 1$ and $1 - Sp = 0$). In general, the overlap of test results between healthy and diseased will be partial in practice, generating ROC curves between the two extreme situations previously described (Weinstein and Fineberg, 1980).

Numerical indices for ROC curves are often used to summarize the accuracy of a diagnostic test. We explain in detail such indices in the next subsection.

2.2.2 Summary measures

As previously stated, the closer to the upper left corner it is placed the ROC curve, the better the discriminatory power of the corresponding diagnostic test. This raises the idea of identifying measures to reduce or summarize a ROC curve into one quantitative index indicating the efficiency of diagnosis, similar to the use of average and variance as summary measures to describe a probability distribution. In addition, these measures will be very convenient for comparison of various diagnostic tests from their corresponding ROC curves. For example, if the ROC curve of a test, say A, is always above the ROC curve of another, say B, then at any given value of the FPF, the corresponding sensitivity of the test A is higher than the sensitivity of the test B. Similarly, if we choose thresholds c_A and c_B for which $Se_A(c_A) = Se_B(c_B)$, the corresponding FPF are ordered in favour of test A, so that $FPF_A(c_A) < FPF_B(c_B)$.

So, it is common to summarize the information provided for the ROC curve into a single global value or index. Several such indexes or measures are defined in the literature and they have been used in several applications (Shapiro, 1999; Greiner et al., 2000). Some of these summary measures which have been defined in the literature are the area under the ROC curve (AUC) (Swets, 1979), single points of the ROC curve, the partial area under the ROC curve (pAUC) (McClish, 1989; Jiang et al., 1996) and the Youden index (J) (Youden, 1950).

Area under the ROC curve

Just a way to assess comprehensively the ability of a diagnostic test to discriminate is to calculate the polygon area that remains under the ROC curve (see Figure 2.8), which is called the area under the curve (AUC) (Swets, 1979). The AUC is independent of the disease prevalence and then it serves as a reference index for comparing different diagnostic tests (Zweig and Campbell, 1993; Altman and Bland, 1994c; Burgueño et al., 1995; López de Ullibarri Galparsoro and Pita Fernández, 1998). In fact, the AUC is the most popular ROC based summary index of the overall diagnostic performance of a test and it is given by:

$$AUC = \int_0^1 ROC(t)dt.$$

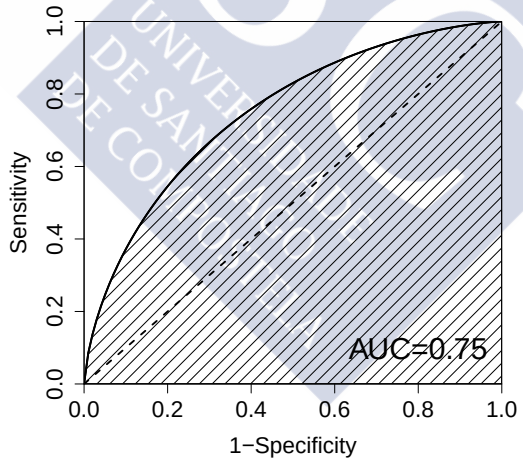


Figure 2.8: Area under the ROC curve (AUC)

By simply looking at the construction of the AUC and the ROC curve, we see that the AUC takes values between 0.5 (uninformative test, the same as a random prediction and corresponds to the diagonal ROC curve that passes through the points (0,0) and (1,1)) and 1 (perfect test, all cases are correctly classified), and represents the diagnostic capacity of the test Y, so that the larger the area the better the diagnostic ability of Y. Thus, values closer to 1 indicate that Y optimally discriminate between healthy and diseased individuals, while values near 0.5 indicate that the test is not informative.

Bamber (1975) showed that the AUC can be interpreted as the probability that a randomly chosen diseased individual will have a test value greater than that of a randomly chosen healthy individual:

$$AUC = \Pr(Y_1 > Y_0),$$

where Y_1 and Y_0 represent values of the diagnostic test from the diseased and healthy groups, respectively.

If two diagnostic tests A and B are ordered in such a way that A is uniformly better that B in the sense that $ROC_A(t) \geq ROC_B(t)$ for all $0 < t < 1$, then it is also satisfied that $AUC_A \geq AUC_B$. However, the converse is not necessarily true.

Partial area under the ROC curve

When the AUC values are globally the same but the two ROC curves intersect and are significantly different in a specific range of interest from a clinical point of view, it is necessary to carry out a partial analysis of the ROC curve. The "global" index used here is the partial area under the ROC curve (pAUC) (McClish, 1989; Jiang et al., 1996) in the interval of clinical interest (t_0, t_1):

$$pAUC = \int_{t_0}^{t_1} ROC(t)dt.$$

The pAUC region can be defined as a portion of specificity (McClish, 1989) or as a portion of sensitivity (Jiang et al., 1996). The estimation and inference of pAUC is an area in continuous development (Walter, 2005; Cai and Dodd, 2008; Komori and Eguchi, 2010, among others).

Youden index

Although the area under the ROC curve (AUC) is the overall index of diagnostic accuracy most commonly used, the Youden index (J) (Youden, 1950) is also another summary measure commonly used in clinical practice (Aoki et al., 1997; Grmec and Gasparovic, 2001). As it will be clear soon, unlike the AUC, the Youden index is related to a specific cutpoint c in the scale of the diagnostic test.

The Youden index, previously defined for binary tests, also makes sense for continuous diagnostic tests which can be defined as follows:

$$J = max_c\{Se(c) + Sp(c) - 1\}.$$

Equivalently, it can be rewritten in several ways:

$$J = max_c\{TPF(c) - FPF(c)\} = max_c\{F_0(c) - F_1(c)\} = max_t\{ROC(t) - t\}.$$

From the first expression above, we observe clearly that the Youden index is the maximum difference between the true positive fraction (TPF) and the false positive fraction (FPF). From the second expression, we interpret the Youden index as the maximum separation between both healthy and diseased populations and it is somehow related to the well-known Kolmogorov-Smirnov statistic.

Therefore, J represents the diagnostic capacity of the diagnostic marker Y, and takes values between 0 and 1, so that values close to 1 indicate that Y optimally discriminate

between healthy and diseased populations, while values close to 0 indicate that the test Y is uninformative. In fact, a complete separation of the two distributions yields a value of $J = 1$, while a full overlap between the two distributions gives $J = 0$. From the third expression of J, we obtain an interesting geometrical interpretation of the Youden index: it corresponds to the maximum vertical distance between the ROC curve and the positive diagonal line in the unit square.

It also presents a very interesting property that does not have the AUC, the cutoff c defining the index, the cutoff c_J for which $Se(c) + Sp(c) - 1$ is maximized, can often be used to establish a reasonable optimal cut-off for the diagnostic marker in clinical practice. In fact, there are a number of recent studies in the literature on the estimation of the Youden index and its corresponding optimal cutpoint (Fluss et al., 2005; Lai et al., 2012; Molanes-López and Letón, 2011, among others). This optimal criterion to select the optimal cutoff point will be treated with a more rigorous detail in Chapter 3.

2.2.3 Predictive Receiver Operating Characteristic curve

There has been a tendency in recent years to consider only the ROC curves to graphically determine the cutoffs and sometimes a low sensitivity is intuitively associated with a low positive predictive value. However, this is not the case. For high values of the cutoff, sensitivity tends to be very small while the PPV is close to 1, that is, points in the left corner of the ROC curve correspond to cutpoints where sensitivity is small but where the PPV is close to 1.

If positive and negative predictive values are considered instead of sensitivity and specificity in a similar way as for the ROC curve, the **Predictive ROC curve** (PROC) is obtained (Vermont et al., 1991; Gallop et al., 2003), that is, the PROC curve is obtained by plotting the coordinate points $(1 - NPV(c), PPV(c))$ (see Figure 2.9).

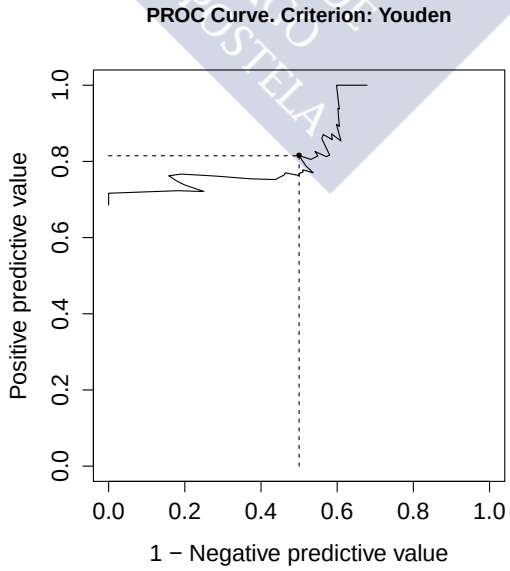


Figure 2.9: PROC curve of a continuous diagnostic test

The PROC curve is generally not as convenient as the ROC curve, specially because it depends on the disease prevalence. However, it is still helpful in some aspects. Under the generally accepted restriction that only the threshold values c satisfying that $PPV(c) \geq p$ and $NPV(c) \geq (1-p)$ provide relevant information on the diagnostic performance of the test, PROC curves of interest are those located in the rectangle defined by the vertical and horizontal lines passing through the point with ordinate and abscissa equal to the disease prevalence p and by the vertical and horizontal lines passing through the point (1,1).

2.3 Estimation and inference for ROC curve and accuracy measures

There are several approaches for estimating the ROC curve and consequently its associated accuracy and summary measures. All these approaches differ in the way the distribution functions of both populations are estimated based on sample values. In general, we can distinguish between parametric and nonparametric methods, depending on if some known parametric distribution is assumed or not for the diagnostic marker in both populations, respectively. In López-Ratón et al. (2012a), a review of all these methods to estimate ROC curve and its accuracy measures is performed.

Consider two independent samples of i.i.d. (independently and identically distributed) observations, $\{Y_{0k_0}\}_{k_0=1}^{n_0}$ and $\{Y_{1k_1}\}_{k_1=1}^{n_1}$, taken from the healthy and diseased populations, Y_0 and Y_1, respectively, with sample sizes n_0 and n_1.

2.3.1 Parametric methods

Parametric estimation methods consider a known parametric distribution of the diagnostic test in both healthy and diseased populations. The normal or lognormal distributions are the most popular alternatives (Tosteson and Begg, 1988; Hajian-Tilaki et al., 1997; Walsh, 1997), taking into account that the more appropriate the choice of such distribution is (Hajian-Tilaki et al., 1997; Walsh, 1997), the better the results. In practice it is not easy to select an appropriate distribution and even sometimes is something subjective (Sorribas et al., 2000), which is a big disadvantage of this approach and, specially, when there is no much information on the true distribution of the marker.

Binormal model

The first parametric method is based on the assumption that the distribution of the marker (or any monotonic transformation of it, for example of Box-Cox type) in both healthy and diseased populations follows a normal distribution. This approach is known as the **binormal method** (Metz, 1986; Hanley, 1996; Hajian-Tilaki et al., 1997; Walsh, 1997; Metz et al., 1998).

So, assuming that Y_i follows a normal distribution with mean μ_i and standard deviation σ_i, for $i = 0, 1$, (using if necessary a monotone transformation of Box-Cox type to

achieve normality), for each cutpoint c, it follows that

$$FPF(c) = 1 - Sp(c) = P(Y_0 \geq c) = P\left(Z \geq \frac{c - \mu_0}{\sigma_0}\right),$$

$$Se(c) = P(Y_1 \geq c) = P\left(Z \geq \frac{c - \mu_1}{\sigma_1}\right),$$

where Z denotes a random variable that follows the standard normal distribution.

Under this condition, there is a linear relationship between the quantiles of the standard normal distribution defined by the values of sensitivity and 1-specificity, that is, it follows that:

$$\Phi^{-1}(Se(c)) = a + b\Phi^{-1}(1 - Sp(c)),$$

where $a = \frac{\mu_1 - \mu_0}{\sigma_1}$ is the intercept, $b = \frac{\sigma_0}{\sigma_1}$ is the slope, and Φ denotes the standard normal cdf. Consequently, it follows that the ROC curve can be expressed as:

$$ROC(t) = \Phi(a + b\Phi^{-1}(t)),$$

where $t = 1 - Sp(c)$ is the FPF.

Under the binormal model, the AUC can be rewritten also in terms of a, b and Φ as follows:

$$AUC = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right).$$

In the literature, several inference methods have been proposed for AUC, mainly based on the binormality assumption and the bootstrap technique.

For the sake of illustration, if we consider the real melanoma dataset, introduced in Section 1.3 of Chapter 1, we can assume normality for the CSS marker in both populations, according to the p-values > 0.5 obtained from the Shapiro-Wilk normality test (p-value = 0.4719 for the healthy group and p-value = 0.9084 for the diseased group), and then we can estimate the ROC curve by assuming the binormal model. Specifically, in this example, it follows that $Y_0 \approx N(\mu_0 = -2.42649, \sigma_0 = 1.760389)$ and $Y_1 \approx N(\mu_1 = 0.556619, \sigma_1 = 1.523482)$. Therefore, for a cutpoint value of c = -2.763, it follows that

$$FPF = 1 - Sp = \Pr(Y_0 \geq -2.763) = \Pr(Z \geq (-2.763 - (-2.42649))/1.760389) = 0.5757985,$$

$$Se = \Pr(Y_1 \geq -2.763) = \Pr(Z \geq (-2.763 - 0.556619)/1.523482) = 0.985333.$$

Analogously, for the cutpoint c = 1.751, it turns out that

$$FPF = 1 - Sp = \Pr(Y_0 \geq 1.751) = \Pr(Z \geq (1.751 - (-2.42649))/1.760389) = 0.00882,$$

$$Se = \Pr(Y_1 \geq 1.751) = \Pr(Z \geq (1.751 - 0.556619)/1.523482) = 0.2165.$$

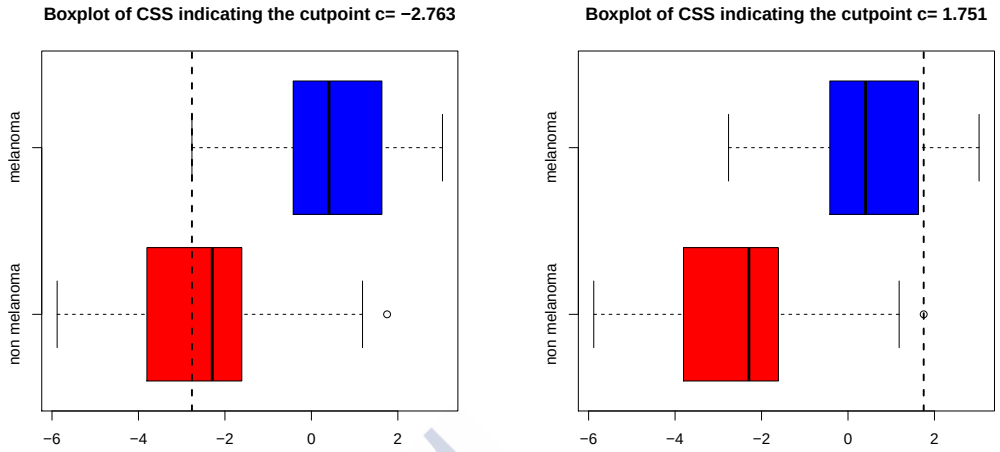


Figure 2.10: Boxplot of CSS as a marker of melanoma, with the cutpoints c = -2.763 (*left panel*) and c = 1.751 (*right panel*) indicated through *dotted vertical lines*.

In Figure 2.10 we plot the boxplots in both populations indicating by a vertical line the two cutpoints c = -2.763 and c = 1.751, previously considered. These graphs allow us to visualize the separation existing between melanoma and non-melanoma populations as a function of such threshold values. Computing in a similar way the pairs (FPF,Se) for all possible cutpoints c in the marker scale, we obtain the binormal ROC curve of the CSS marker for diagnosing melanoma (see Figure 2.11).

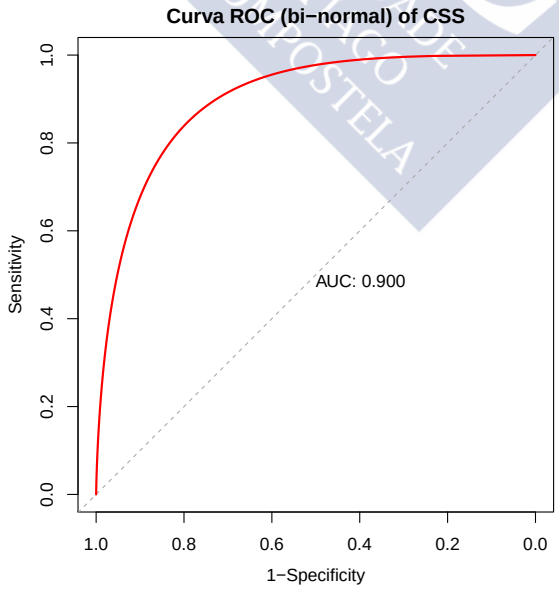


Figure 2.11: ROC curve estimate of CSS marker for diagnosing melanoma based on the binormal model.

The AUC value estimated by the binormal method is equal to 0.90, indicating a very good classification performance of the CSS marker for diagnosing melanoma.

2.3.2 Nonparametric methods

To solve the problem of specifying an appropriate distribution, different nonparametric methods have been proposed in the literature. The nonparametric methods are characterized by not assuming any known distribution for the diagnostic tests in both healthy and diseased populations.

Empirical method

The empirical method is one of the most common and simple nonparametric methods proposed in the literature (Zweig and Campbell, 1993; Campbell, 1994). It estimates the distribution function of the diagnostic test by using the empirical distribution function of the sample; that is, the empirical estimator of the ROC curve simply applies the definition of the ROC curve to the observed data. So, for each possible cut-off value, c, the empirical $TPF(Se)$ and the empirical FPF are estimated by $\widehat{Se}(c)$ = proportion of diseased individuals in the sample with $Y \geq c$, and $\widehat{FPF}(c)$ = proportion of healthy individuals with $Y \geq c$, that is,

$$\widehat{Se}(c) = \widehat{TPF}(c) = \frac{1}{n_1}\sum_{i=1}^{n_1} I[Y_{1i} \geq c] = \frac{\#[Y_{1i} \geq c]}{n_1},$$

$$1 - \widehat{Sp}(c) = \widehat{FPF}(c) = \frac{1}{n_0}\sum_{j=1}^{n_0} I[Y_{0j} \geq c] = \frac{\#[Y_{0j} \geq c]}{n_0}.$$

Therefore, the empirical ROC curve is a plot of the pairs $(\widehat{FPF}(c), \widehat{TPF}(c))$ for all $-\infty < c < \infty$.

Equivalently, we can write the empirical ROC curve as follows:

$$\widehat{ROC_{emp}}(t) = 1 - \widehat{F_1}(\widehat{F_0}^{-1}(1-t)),$$

where $\widehat{F_0}$ and $\widehat{F_1}$ are the empirical distribution functions obtained from the two samples of healthy and diseased populations, respectively:

$$\widehat{F_0}(c) = \frac{1}{n_0}\sum_{i=1}^{n_0} I[Y_{0i} \leq c],$$

$$\widehat{F_1}(c) = \frac{1}{n_1}\sum_{j=1}^{n_1} I[Y_{1j} \leq c].$$

The empirical method presents the following properties:

- It is an increasing step function that approaches the theoretical ROC curve, as it is based on empirical distribution functions.

- If there are no ties in the data, horizontal jumps are sized $1/n_0$ and vertical jumps are sized $1/n_1$.
- If, on the contrary, there are some ties in the data, they provide larger jumps, where the size of the jumps depend on the number of ties.
- It depends only on the ranks of the combined set of scores, so it is invariant under strictly increasing transformations of the data.

The typical shape of an empirical ROC curve can be seen in Figure 2.12, where we have plotted the empirical ROC curve obtained with the melanoma dataset (previously introduced in Section 1.3 of Chapter 1).

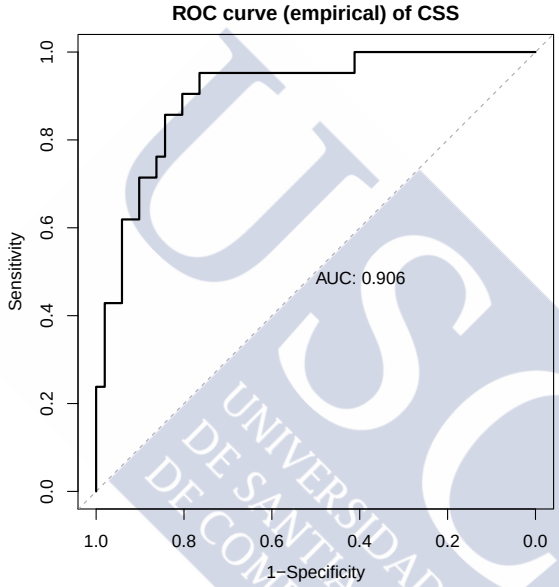


Figure 2.12: Empirical ROC curve of CSS as marker of melanoma

From the empirical ROC curve, you can obtain empirical estimators of summary measures such as the AUC and Youden index, by simply applying their definition to the empirical ROC curve.

For instance, the empirical estimator of the AUC is given by:

$$\widehat{AUC} = \int_0^1 \widehat{ROC}_{emp}(t)dt.$$

If there are no ties in the data, the empirical estimate of the AUC is equivalent to the well-known Mann-Whitney statistic:

$$\widehat{AUC} = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I[Y_{1j} \geq Y_{0i}].$$

In order make inference on the AUC, it is necessary to study the asymptotic behavior of the empirical estimate of the AUC. On one hand, the construction of confidence intervals can be done using the nonparametric methods proposed by Hanley and McNeil

(1982); Delong et al. (1988). On the other hand, it is of great interest to test whether a specific diagnostic marker Y can discriminate between healthy and diseased populations, and this can be done by solving a hypothesis test based on AUC, where the null hypothesis $H_0 : AUC = 0.5$ is tested versus the alternative hypothesis $H_1 : AUC > 0.5$, which is equivalent to testing if Y_1 is stochastically greater than Y_0 or, equivalently, if there is some degree of separation between the two distributions. For large sample sizes, the test will reject the null hypothesis if:

$$\frac{\widehat{AUC} - 0.5}{\sqrt{\frac{n_0 + n_1 + 1}{12 n_0 n_1}}} > z_{1-\alpha}.$$

Taking into account the equivalence established between the empirical estimator of the AUC and the Mann-Whitney statistic, this hypothesis test can also be solved for small sample sizes using the exact distribution of the Mann-Whitney statistic.

Similarly to the previous test, only useful for detecting whether a marker or diagnostic procedure is better than random allocation, it may also be of interest to determine if it has a satisfactory AUC. In this case, denoting by θ_0 the minimum acceptable value for AUC, the corresponding contrast assumption would be:

$$H_0 : AUC = \theta_0 \text{ versus } H_1 : AUC > \theta_0.$$

Thus, given a level of signification, α, the null hypothesis H_0 will be rejected if:

$$\frac{\widehat{AUC} - \theta_0}{\widehat{var}(\widehat{AUC})} > z_{1-\alpha},$$

where $\widehat{AUC}$ is the empirical estimator of the AUC, and $\widehat{var}(\widehat{AUC})$ is the estimate of $var(\widehat{AUC})$.

The empirical estimator of the Youden index is given by:

$$\widehat{J} = \max_t \left\{ \widehat{ROC}(t) - t \right\},$$

and it is equivalent to the well-known Kolmogorov-Smirnov statistic for comparison of distribution functions:

$$\widehat{J} = \sup_c \left\{ \widehat{TPF}(c) - \widehat{FPF}(c) \right\} = \sup_c \left\{ (1 - \widehat{F}_1(c)) - (1 - \widehat{F}_0(c)) \right\} = \sup_c \left\{ \widehat{F}_0(c) - \widehat{F}_1(c) \right\}.$$

Therefore, the following hypothesis test:

$$H_0 : J = 0 \quad \text{versus} \quad H_1 : J > 0$$

can be solved based on the classical Kolmogorov-Smirnov test for equality of distributions.

For illustrative purposes, we consider the same melanoma dataset. Instead of assuming normality, in Tables 2.6-2.7, we show the confusion matrix for the cutpoints c = -2.763

and c = 1.751, respectively, based on empirical estimates. From these tables, it follows that $Se = 1$ and FPF= 0.588 for the cutpoint c = -2.763, and Se = 0.238 and FPF = 0.020 for the cutpoint c = 1.751.

Table 2.6: Classification of CSS results by melanoma status for c = -2.763.

CSS	melanoma	non melanoma	Total
Positive (CSS ≥ -2.763)	21	30	51
Negative (CSS < -2.763)	0	21	21
Total	21	51	72

Table 2.7: Classification of CSS results by melanoma status for c = 1.751.

CSS	melanoma	non melanoma	Total
Positive (CSS ≥ 1.751)	5	1	6
Negative (CSS < 1.751)	16	50	66
Total	21	51	72

The empirical method is well-known and applied due to its simplicity, but the main drawback of the empirical method is that the empirical estimator of the ROC curve is a step function rather than a continuous function (built from empirical cdfs and empirical quantile distributions, both non-continuous). Since we are dealing with continuous diagnostic tests, the ROC curve should be a continuous function and therefore it seems reasonable to construct estimators that are also continuous. The continuity of the ROC curve estimator can be obtained parametrically, assuming a parametric model for the ROC curve (as we saw in the previous section), but also non parametrically using **smoothing techniques**. In the following, we briefly discuss two smoothing techniques.

Kernel smoothing

Kernel smoothing is a smooth approach based on kernel estimation (Zou et al., 1997; Lloyd, 1998) of the distribution functions F_1 and F_0, thus providing a smooth estimate of these functions. Specifically F_1 and F_0 are estimated by, respectively,

$$\widehat{F}_{1,h_1}(y) = \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{K}\left(\frac{y - Y_{1i}}{h_1}\right),$$

$$\widehat{F}_{0,h_0}(y) = \frac{1}{n_0}\sum_{j=1}^{n_0}\mathbb{K}\left(\frac{y - Y_{0j}}{h_0}\right),$$

where $\mathbb{K}(u) = \int_{-\infty}^{u} K(v)\mathrm{d}v$, K denote a kernel function (a symmetric density) and $h_i, i = 0, 1$ is the smoothing parameter for each population.

The kernel smooth estimate of the ROC curve is obtained by simply taking into account the previous estimates, that is, by plotting the corresponding pairs of coordinate points, that is,

$$\widehat{ROC}(t) = 1 - \widehat{F}_{1,h_1}(\widehat{F}^{-1}_{0,h_0}(1-t)), t \in (0,1)$$

In kernel smoothing it is necessary to choose the kernel function K, for instance, the Gaussian kernel or the Epanechnikov Kernel (Epanechnikov, 1969) and, most importantly, the optimal smoothing bandwidth h (Zou et al., 1997; Lloyd and Yong, 1999; Hall and Hyndman, 2003).

The kernel method has some limitations. For instance, it often suffers from boundary effects, that is, it has trouble to estimate at the borders, and therefore, the kernel estimator may not be valid at the ends of the ROC curve. Besides, the kernel estimator of the ROC curve is not invariant under a monotone transformation of the data.

Considering the melanoma dataset, we plot in Figure 2.13 the smooth ROC curve of CSS for diagnosing melanoma using three different bandwidths proposed in the literature. We observe that the three estimations are very similar.

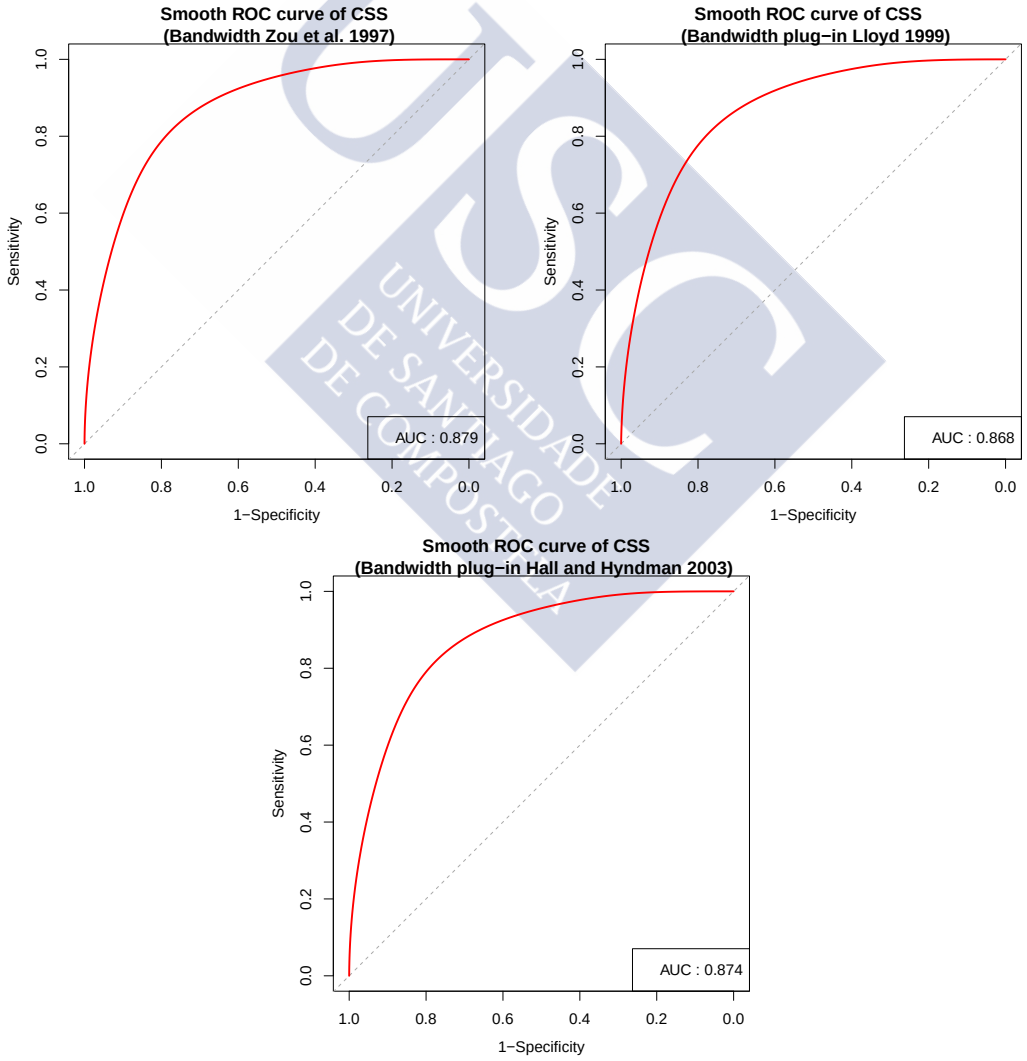


Figure 2.13: Smooth ROC curve of CSS as marker of melanoma

P-spline smoothing

P-spline smoothing is an alternative nonparametric approach (Eilers and Marx, 1992, 1996). Before discussing this smoothing method in some more detail, it is necessary to introduce first some basic concepts.

An **spline** is a numeric function piecewise-defined by polynomial functions. A **B-spline basis (of order** $m+1$**)** is constructed such that each basis function is non-zero over the interval between *m+3* adjacent nodes, with *m+1* the order of the basis (for instance, *m*=2 for a cubic spline). In order to define a B-spline basis of k parameters it is therefore necessary to consider *k+m+1* nodes, $x_1 \leq x_2 \leq ... \leq x_{k+m+1}$, where B-splines basis functions are defined recursively (de Boor, 1978) as follows:

$$B_i^m(x) = \frac{x - x_i}{x_{i+m+1} - x_i} B_i^{m-1}(x) + \frac{x_{i+m+2} - x}{x_{i+m+2} - x_{i+1}} B_{i+1}^{m-1}(x), i = 1, ..., k$$

$$B_i^{-1}(x) = \begin{cases} 1, & x_i \leq x \leq x_{i+1}, \\ 0, & \text{otherwise.} \end{cases}$$

Based on this B-spline basis, the estimation is given by

$$\sum_{i=1}^{k} B_i^m(x)\beta_i.$$

P-spline smoothing uses a B-spline basis, usually defined in equidistant nodes, with a penalty directly on the parameters β_i:

$$P = \sum_{i=1}^{k-1} (\beta_{i+1} - \beta_i)^2 = \beta_1^2 - 2\beta_1\beta_2 + 2\beta_2^2 - 2\beta_2\beta_3 + ... + \beta_k^2.$$

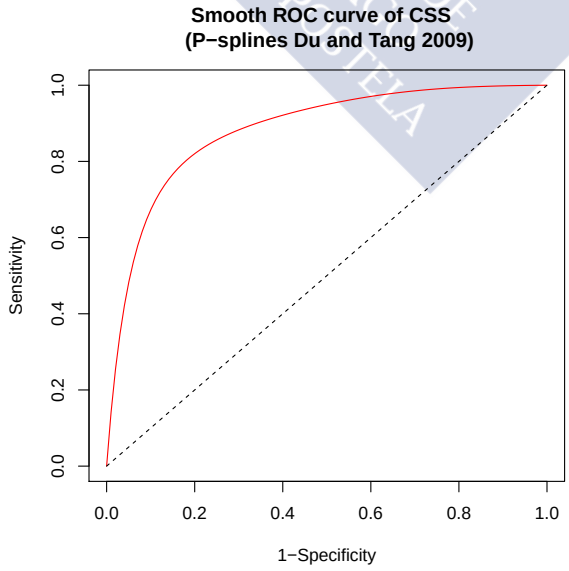


Figure 2.14: Smooth P-spline ROC curve of CSS as marker of melanoma

Du and Tang (2009) proposed the estimation of the ROC curve by P-spline smoothing based on plugging the P-spline estimation of the distribution functions of both populations into the theoretical expression of the ROC curve. In Figure 2.14, we show the P-spline smooth ROC curve of CSS estimated using source code in R kindly provided by Du and Tang (2009).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

He had the vague sense of standing on a threshold, the crossing of which would change everything.

- Kate Morton

Chapter 3

Criteria to select the optimal cutoff point in diagnostic tests

The selection of the appropriate cutpoint is crucial to avoid erroneous conclusions being drawn in clinical practice. In medical decision theory and epidemiologic research, determining a cutpoint for a quantitative variable is a common problem, and has indeed been an active area of research (Miller and Siegmund, 1982; Altman et al., 1994; Lausen and Schumacher, 1996; Mazumdar and Glassman, 2000). The objective is to select the optimal cutpoint c of a continuous diagnostic marker that best discriminates between patients with and without the disease, under the assumption that high values of the test are associated with disease, and in such a way that individuals with a diagnostic test value Y equal to or higher than c are classified as diseased (positive test), whereas patients with a lower value are classified as healthy (negative test). It must be borne in mind that the optimal cutpoint depends on the situation in which it is to be used, and so generally one cannot talk in absolute terms of a "best choice" of cutpoint c. The optimal selection of a cutoff point will depend on the analysis of the use proposed for that point and such analysis together with the study of the corresponding empirical curves and/or theoretical formulas will influence the determination of a cutpoint and will allow its justification. This is the reason why several strategies for selecting optimal cutpoints in continuous diagnostic tests have been proposed in the literature (see Youden, 1950; Feinstein, SH, 1975; Metz, 1978; Albert and Harris, 1987; England, 1988; Schäfer, 1989; Vermont et al., 1991; Greiner, 1995; Riddle and Stratford, 1999, among others). While the choice of the number and values of the cutpoints may be made in accordance with criteria already established by earlier studies or, for theoretical reasons, be based on clinical, biological or physiological information (this is the desirable approach), at other times, however, this information is not available and it is the researcher him/herself who has to decide on the cutpoints that are to be set on the basis of certain criteria. Therefore, in the absence of any "a priori" clinical information showing the prognostic relationship between the continuous marker and the presence of certain disease, the choice of a cutpoint must be determined based on graphs and numerical results (Greiner, 1995; Hilsenbeck and Clark, 1996; Lausen and Schumacher, 1996; Mazumdar and Glassman, 2000; Abdolell et al., 2002).

In this chapter of the thesis we show a comprehensive overview of the dichotomization methods that have been proposed in the literature to select cutpoints in the setting of a continuous diagnostic test, discussing their relative advantages and disadvantages,

giving an overview of their use in many medicine fields, and finally, identifying which methods are similar or special cases of others in an intent to reduce the many options currently available.

The methods for selecting the optimal cutoff value in continuous diagnostic tests may initially be divided into two groups: bench level methods (or analytical methods) and clinical or epidemiological methods. Bench level methods are based on the analytical characteristics of the test. Often researchers use methods to select the optimal cutoff point based on arbitrary values that may have no relationship with patient outcomes. Thus, a procedure that has been widely used for selecting the cutoffs is based on choosing the values of the quartiles or specific percentiles (median, $25th$ percentile, $75th$ percentile) of the sample distribution, the median or some other value that has been admitted so far as clinically relevant. But these methods are in general arbitrary, may not be useful in many practical situations, and do not determine the true prognostic value of a diagnostic marker (Courdi et al., 1988; Altman, 1991; Altman et al., 1994; Hilsenbeck and Clark, 1996; Mazumdar and Glassman, 2000). At other times, the optimal cutoff points considered in clinical practice are often determined as the mean plus several (usually two or three) standard deviations of the observed results in the non-diseased sample (Richardson et al., 1983). For example, some authors have determined the optimal cutoff points to be the mean of the control group $+1.00SD$, $+1.64SD$ and $+2.00SD$. Under the assumption of normality, these cutoff points correspond to the $84th$, $95th$, and $97.7th$ percentiles, respectively, that is, they automatically achieve an specificity value equal to 84%, 95%, and 97.7%, respectively (Barajas-Rojas et al., 1993). Recently, some authors have indicated that such a method for computing the optimal cutoff point without taking into account the value of the sensitivity cannot reflect the best cutoff point for discriminating between diseased and healthy populations (Greiner and Böhning, 1994).

Unlike bench level methods, the clinical (or epidemiological) methods are concerned with the subject classification and they assume, whatever characteristic of interest (an analytic measure, a biomarker, an image, signs or symptoms,...) that the measurement represents the status of a patient at a particular time. Clinical methods are generally based on a balance between the ability of the test to correctly classify patients with the disease (sensitivity) and the ability to correctly classify patients without the disease (specificity). In general, this balance is usually represented by the ROC curve, and therefore the ROC-based optimal-cutpoint selection methods are included here. A popular method of this group is to maximize the Youden index (Youden, 1950). These clinical methods, in turn, can be further divided into several groups. Some authors (Gönen and Sima, 2013) talk in general of two main statistical approaches to the problem of selecting an optimal cutpoint, an approach based on the ROC curve, and another approach that seeks to maximize an appropriately chosen statistical test (Mazumdar and Glassman, 2000). The ROC analysis furnishes several optimal cutpoint selection criteria based on sensitivity and specificity measures (Green and Swets, 1966; Zweig and Campbell, 1993; Coffin and Sukhatme, 1997; Pepe, 2004), by imposing certain specifications on such measures such as to assume certain values or to consider a linear combination or any other function of both measures. Furthermore, ROC-curve criteria allow for the choice of optimal cutpoints based on the disease prevalence and the relative cost ratio (risks and benefits) of the possible (correct and incorrect) medical decisions derived from the diagnostic test result. The other widely used approach is the maximization of an appropriate statistical

test, often one that is based on two samples and compares the groups that result from the dichotomization. This method first appeared in the context of a binary diagnostic test using the Pearson Chi-squared test (Miller and Siegmund, 1982), and it is frequently known as the maximum Chi-squared or minimum p-value method (Mazumdar and Glassman, 2000).

Despite this initial classification, given the large number of dichotomization criteria existing in the literature, we consider appropriate to split them in the following two groups:

(i) Completely data driven methods

(ii) Methods with researcher (or user) requirements

In Section 3.1, we present an exploratory graphical analysis (Williams et al., 2006) that allows the graphical visualization of possible cutpoints and that it is recommended to carry out before a more formal selection. Then, Sections 3.2-3.3 are devoted to explain in detail each of these two groups of methods or strategies for selecting optimal cutpoints in continuous diagnostic tests.

3.1 Exploratory analysis

Once it has been determined that a continuous variable is significantly associated with the outcome variable or presence of the event of interest (here, the presence of disease), the first thing you should do is to examine this relationship graphically. The exploratory graphs can reveal obvious thresholds suggesting possible cutpoints or a range of values which should be taking into account a priori (Williams et al., 2006). In the literature, a series of graphs have been suggested: scatter-plots, grouped data plots, lowess smoothed plots, ...

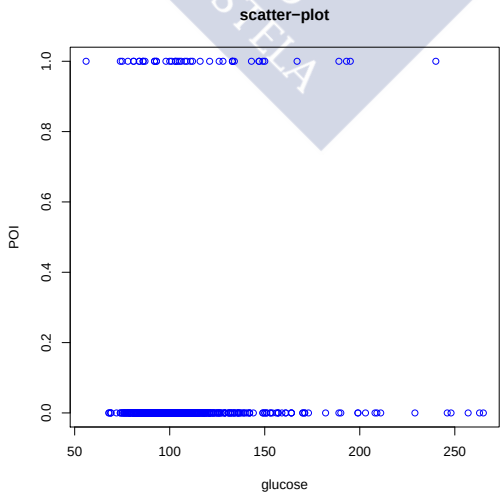


Figure 3.1: Scatter-plot of plasma glucose levels (a continuous variable) as potential marker for predicting postoperative infection.

For a binary response variable (for instance, presence/absence), a scatter-plot graph on the observed outcome of the prognostic variable or marker Y ideally shows the degree to which the factor separates patients into risk groups. This plot will be generally uninformative unless there is a fairly clear separation, in which case you will see a step function that displays the existence of at least one cutoff point. However, this graph is often difficult to interpret, since all points are distributed in only two ordinate values due to the binary nature of the event of interest (see Figure 3.1 where we have plotted the scatter-plot corresponding to real data from a prospective study for examining plasma glucose as a potential marker for predicting postoperative infection (Figueiras and Cadarso-Suárez, 2001), that has been previously introduced in Section 1.3 of Chapter 1). A more illustrative graph is obtained by dividing first the continuous variable into equal intervals (for example, deciles or some other grouping into quantiles) and representing then the proportion of occurrences in each interval against the midpoint of each interval. These graphs are called grouped data plots and they are an useful diagnostic graphic when the outcome is binary (Mazumdar and Glassman, 2000). However, the disadvantage of this type of representation is that it can be sensitive to the interval width used, so it is a good idea to consider different amplitudes, and especially to check the frequency of individuals in each interval, because if it is too small, the information will not be accurate. In Figure 3.2, we see an example of a grouped data plot with deciles for the same data considered in Figure 3.1.

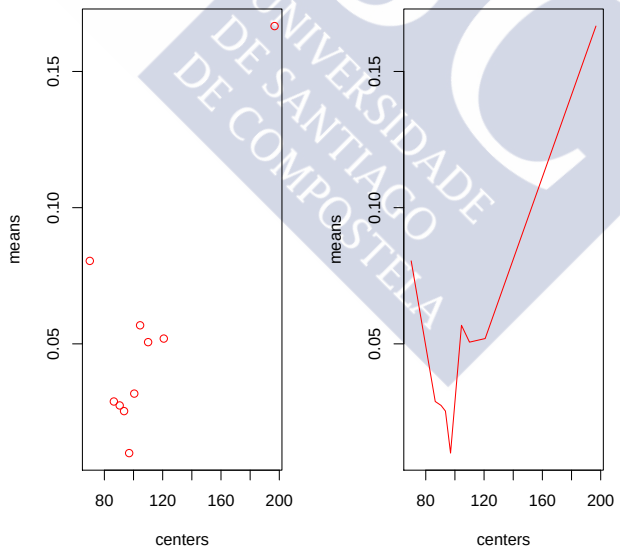


Figure 3.2: Grouped data plot of plasma glucose levels (a continuous variable) as potential marker for predicting postoperative infection.

There exist various techniques that allow flexible modeling of a continuous variable, including smoothing splines, piecewise polynomial regression and nonparametric regression (Greenland, 1995; Harrell, 2001). Of these techniques, the smoothing splines are those who have received more attention and, in fact, they are incorporated in many

standard statistical packages. The results of the regression models using splines allow the representation of the continuous variable as a function of the response or any transformed version of the response, for instance, the odds ratio (OR) of a logistic regression model. In Figure 3.3 a lowess smoothed plot of plasma glucose levels as a potential marker for predicting postoperative infection is shown.

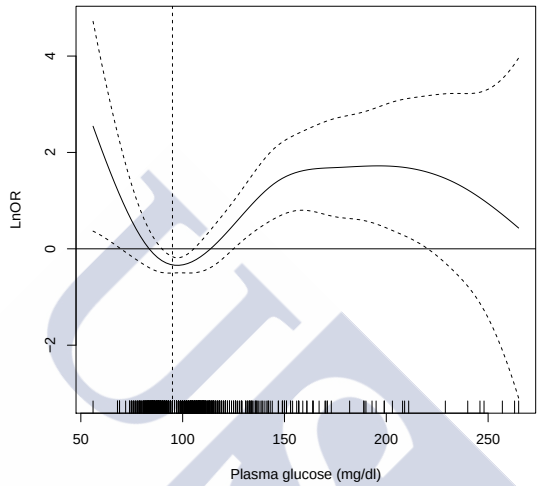


Figure 3.3: Lowess smoothed plot of plasma glucose levels (a continuous variable) as a potential marker for predicting postoperative infection.

3.2 Completely data-driven methods

We use the term "completely data-driven methods" to refer to those methods that use only the observed test results to compute the threshold. This group of methods is mostly based on ROC analysis, that is, criteria based mainly on sensitivity and/or specificity measures.

First of all, we list below the criteria included in this group of completely data-driven methods. As you can see, some of them have been additionally grouped in different subcategories: I. Criteria based on sensitivity and specificity measures; II. Criteria based on diagnostic likelihood ratios; III. Criteria based on predictive values; IV. Prevalence-based methods; V. Methods not based on ROC analysis; VI. Other methods.

I. Criteria based on sensitivity and specificity measures

- 1. Maximization of sensitivity
- 2. Maximization of specificity
- 3. Sensitivity-specificity equality approach (or simultaneous maximization of sensitivity and specificity)

- 4. Maximization of the Youden index (maximization of the sum of sensitivity and specificity, maximization of concordance, or maximization of net gain)
- 5. Minimum ROC distance (point on ROC curve closest to $(0, 1)$, ROC plot-based approach, K index, Euclidean distance or geometric approach)
- 6. Maximization of concordance probability (accuracy related area or percent correct answers area)

II. Criteria based on diagnostic likelihood ratios

- 7. Maximum diagnostic positive likelihood ratio
- 8. Minimum diagnostic negative likelihood ratio

III. Criteria based on predictive values

- 9. Maximization of positive predictive value
- 10. Minimization of negative predictive value
- 11. Positive Predictive value-Negative Predictive value equality approach
- 12. Simultaneous maximization of positive and negative predictive values
- 13. Maximum Predictive Summary index (or maximization of the sum of positive and negative predictive values)
- 14. Criterion of the point on PROC curve closest to (0,1) point (or minimum P-ROC distance)
- 15. Maximization of the product of positive and negative predictive values
- 16. P-R (precision-recall) break-even point approach ($PPV = Se$)
- 17. P-R (precision-recall) plot-based approach

IV. Prevalence-based methods

- 18. Observed prevalence
- 19. Mean predicted probability
- 20. Prevalence approach (or prevalence = predicted prevalence)

V. Methods not based on ROC analysis

- 21. Split test values at median
- 22. Maximum Chi-squared (or minimum p-value)

VI. Other methods

- 23. Maximization of diagnostic odds ratio
- 24. Maximization of Kappa index

I. Criteria based on sensitivity and specificity measures

1. Maximization of sensitivity
2. Maximization of specificity

In some diagnostic situations, it is desirable to have a higher probability of detecting a true negative or a true positive result, and in such a case the optimal cutpoint should therefore be chosen with this aim in mind. Therefore, one criterion for selecting the optimal value is based on the maximization of one of the two accuracy measures, i.e., the *maximization of sensitivity* (Filella et al., 1995; Hoffman et al., 2000; Álvarez García et al., 2003) or the *maximization of specificity* (Bortheiry et al., 1994; Hoffman et al., 2000). So, in these methods the cutpoint c that maximizes either $Se(c)$ or $Sp(c)$, respectively, is chosen. If there are different values verifying such a condition, the value that has the higher value of the non maximized measure (either specificity or sensitivity) is usually selected. However, these procedures prove somewhat extreme, since the choice of an optimal cutpoint should generally imply an equilibrium between both Se and Sp. Consequently, the equality or simultaneous maximization of these two quantities (sensitivity and specificity) (Riddle and Stratford, 1999; Peng and So, 2002; Gallop et al., 2003) or the maximization/minimization of a given combination of such quantities tend in general to be more appropriate criteria.

3. Sensitivity-specificity equality approach

Another strategy for selecting the optimal cutoff point is to choose the optimal value as the cutoff point c_0 for which both sensitivity and specificity measures are similar or virtually identical: $Se(c_0) \cong Sp(c_0)$ (Amaro et al., 1995; Greiner et al., 1995; Cantor et al., 1999; Hosmer and Lemeshow, 2000; Álvarez García et al., 2003; Liu et al., 2005; Chen et al., 2006; Freeman and Moisen, 2008; Brandao et al., 2009; Caraguel et al., 2011). This is equivalent to select the cutpoint c_0 minimizing the absolute value of the difference of such measures, that is, $c_0 = \arg\min_c |Sp(c) - Se(c)|$. This point is known as *the Symmetry point*, also known in the literature as *the point of equivalence* (see Greiner et al., 1995; Defreitas et al., 2004; Adlhoch et al., 2011) and it corresponds to the operating point where the ROC curve and the line $y = 1 - x$ (the perpendicular line to the positive diagonal passing through the $(0, 1)$ point) intersect.

The point on the ROC curve with sensitivity equal to specificity can also be seen as the point that minimizes the total error, or equivalently, as the point that maximizes simultaneously both types of correct classifications (Riddle and Stratford, 1999; Gallop et al., 2003), that is, it balances the two types of correct classifications and therefore it corresponds to the probability of correctly classifying any individual, whether he/she is healthy or diseased (Jiménez-Valverde, 2012, 2014). So, it may also be used as an accuracy measure and to compare different diagnostic tests, since it does not depend on the test scale. However, it tends to overestimate (underestimate) the positive predictive value for low (high) prevalence situations. Moreover, in clinical practice, the false negative and false positive results often do not have the same cost (Rutter and Miglioretti, 2003). The extension of this criterion taking into account different costs for false negative and false positive misclassifications will be treated in detail in Chapter 4.

Sometimes, a level α is set and the optimal cutoff point is the value for which the absolute value of the difference between the sensitivity and specificity measures is below that level. When there exist multiple cutpoints verifying such a condition, sometimes the optimal value is defined as the average of those points, and since in some practical situations, the sensitivity achieved is not exactly equal to the specificity, if you want to work with a single accuracy value, this value is obtained, for instance, as the mean of the values (practically equal) of sensitivity and specificity measures.

The optimal cutpoint computed from the criterion that sets an equal value for the sensitivity and specificity indexes, can be estimated by parametric methods (for instance, a normal distribution can be assumed for the marker distribution in both populations of healthy and diseased individuals) or non-parametric methods (when a deviation of the normal distribution is observed).

In addition, this criterion is theoretically equivalent to the *simultaneous maximization of sensitivity and specificity*, that is, instead of selecting either a high sensitivity or a high specificity, an alternative for selecting the optimal cutoff point is to determine the operating point on the ROC curve where both measures are maximized simultaneously: Se and Sp (Riddle and Stratford, 1999; Gallop et al., 2003).

This criterion has been applied in various fields, including studies in psychotherapy (Crits-Christoph et al., 2001), cancer (Yang et al., 2004), pediatrics (Schurman et al., 2007) and cardiology (Maneesai and Krittayaphong, 2007).

In the following we will prove the equivalence between *the sensitivity-specificity equality approach* and *the simultaneous maximization of sensitivity and specificity*.

The criterion based on simultaneous maximization of sensitivity and specificity defines the optimal cutpoint by:

$$\max_c \{\min\{Sp(c), Se(c)\}\}.$$

Taking into account that:

$$\{c : Se(c) \leq Sp(c)\} = \{c : ROC(t) \leq 1 - t, \text{where } t = 1 - F_0(c)\},$$

where F_0 denotes the cdf of the marker in the healthy population, it follows that:

$$\begin{aligned} \min\{Sp(c), Se(c)\} &= \begin{cases} Se(c), & \text{if } ROC(t) \leq 1 - t, \\ Sp(c), & \text{if } ROC(t) \geq 1 - t, \end{cases} \\ &= \begin{cases} Se(c), & \text{if } Se(c) \leq Sp(c), \\ Sp(c), & \text{if } Se(c) \geq Sp(c). \end{cases} \end{aligned}$$

Therefore, for t such that $ROC(t) \leq 1 - t$, we have to compute the maximum of $ROC(t)$, and given that the theoretical ROC curve is always monotone increasing, the maximum of $ROC(t)$ where $ROC(t) \leq 1 - t$ will be attained at the largest t, that is, at the point t_0, where $ROC(t_0) = 1 - t_0$, with $t_0 = 1 - F_0(c_0)$, and, consequently, $Se(c_0) = Sp(c_0)$. Analogously, for t such that $ROC(t) \geq 1-t$, we need to maximize $1-t$ and this maximum is reached at the smallest t value, that is, at the same point where $Se(c_0) = Sp(c_0)$. See Figure 3.4 for a graphical example that illustrates this equivalence.

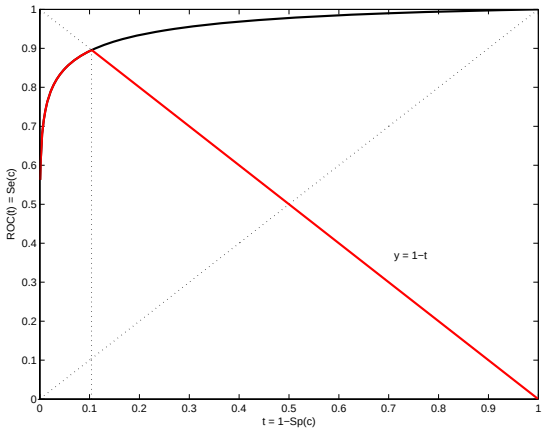


Figure 3.4: Objective function to maximize in the simultaneous maximization of sensitivity and specificity criterion.

4. Maximization of the Youden index

One of the most frequently studied methods of this group of completely data-driven methods is *the maximization of the Youden index* or simply called *the Youden index*. We remind that the Youden index is a summary measure of the ROC curve frequently used in clinical practice, first introduced in the context of binary tests (Youden, 1950; Aoki et al., 1997; Shapiro, 1999; Greiner et al., 2000; Grmec and Gasparovic, 2001; Kramar et al., 2001; Fluss et al., 2005; Schisterman et al., 2005; Perkins and Schisterman, 2006, among others), and therefore an overall measure of the effectiveness of a diagnostic marker. So, the Youden index maximization corresponds to the cutpoint c where the sum of the sensitivity and specificity of the test reaches its maximum:

$$YI(c) = \max_{c}\{Se(c) + Sp(c) - 1\}.$$

As we already said in Chapter 2, the Youden index is another commonly used optimal criteria for selecting the optimal cutpoint. The Youden index, also known in the literature as the *Kolmogorov-Smirnov index*, see for example, Pepe (2004), maximizes the sum of the two correct classification probabilities and it can be seen geometrically as the point that maximizes the Euclidean distance from the ROC curve to the positive diagonal or no discrimination line, that is, the maximum of the Youden index is the furthest point from chance (Perkins and Schisterman, 2006; Schisterman and Perkins, 2007). The most intuitive characteristic of the Youden index is that it takes into account the distribution of the test results from those individuals with (sensitivity) and without (specificity) the target condition, and it also adapts to skewed distributions. The Youden index is a very well-known and used criterion in practice to select the classification threshold, although it coincides with the Bayes rule for the minimum misclassification error probability when the prevalence of disease is assumed 0.5 (Webb, 2002; Skaltsa et al., 2010), which may be far from reality. In fact, several published articles (Perkins and Schisterman, 2006; Liu, 2012; Rota and Antolini, 2014) have encouraged its use. In addition, there are several

recent studies in the literature about the estimation of the Youden index and the corresponding optimal cutpoint (Fluss et al., 2005; Molanes-López and Letón, 2011; Lai et al., 2012, among others). The bias of the estimation of the Youden index and its corresponding cutpoint depends on the test accuracy, disease prevalence, underlying population distribution, and assumptions made for sample estimation (Fluss et al., 2005). The estimation of the optimal cutpoint maximizing the Youden index in case-control studies does not differ from the estimation in cross-sectional studies (Fluss et al., 2005).

Taking into account that $Se(c) = 1 - F_1(c)$ and $Sp(c) = F_0(c)$ where F_i, $i = 0, 1$, is the marker distribution in healthy and diseased populations, respectively, the Youden index can be also expressed in terms of both distribution functions as follows:

$$YI = \max_c \{F_0(c) - F_1(c)\}.$$

Therefore, a plug-in type estimate of the Youden index can be performed by estimating the distributions F_0 and F_1, and substituting such estimations in the above equation, that is, $\widehat{YI} = \max_c\{\hat{F}_0(c) - \hat{F}_1(c)\}$. Different approximations for estimating F_0 and F_1 provide different estimates of the Youden index and its associated optimal cutpoint.

It should be noted that *the Youden index method* is identical (from an optimization point of view) to the method that *maximizes the sum of sensitivity and specificity* (Albert and Harris, 1987; Zweig and Campbell, 1993), to the criterion that *maximizes concordance*, which is a function of the AUC defined as $Se + Sp - 0.5$ (Begg et al., 2000; Mogulkoc et al., 2001; Silman and Macfarlane, 2002; Schwarz et al., 2009; Gönen and Sima, 2013), and to the method based on *the maximum net gain*: $(\sqrt{2}/2)(Se(c) + Sp(c) - 1)$ (Koepsell and Connell, 2002). All these criteria correspond to a point that minimizes the probability of misclassification of any type (false positives and false negatives).

5. Minimum ROC distance

Whereas the ideal situation is when sensitivity and specificity are equal to 1, a possible method for selecting the optimal cutoff point is to find a cutoff such that the pair of values (Se, Sp) is as close as possible to the pair of perfect classification given by (1,1). This is equivalent to the point on the ROC curve closest to the (0,1) point or K *index* (Metz, 1978; Vermont et al., 1991; Cantor et al., 1999; Liu et al., 2005; Perkins and Schisterman, 2006; Freeman and Moisen, 2008; Irwin and Irwin, 2011; Liu, 2012; Zou et al., 2013; Rota and Antolini, 2014). It is also called *the North-West corner* or *the closest-to-(0,1) criterion*. It is also one of the completely data-driven methods, most frequently studied.

This optimal cutpoint minimizes the Euclidean distance between the pair (Se, Sp) (or equivalently the point $(1-Sp, Se)$ on the ROC plane) and the point $(1, 1)$ (or equivalently the point $(0, 1)$ on the ROC plane), that is, the point on the ROC curve which has the shortest distance to the top-left corner $(0, 1)$:

$$c = \arg\min_c \left\{ \sqrt{(1 - Se(c))^2 + (1 - Sp(c))^2} \right\},$$

or, equivalently, in terms of the survival distributions in both populations (Perkins and Schisterman, 2006):

$$c = \arg\min_c \left\{ \sqrt{(1 - S_D(c))^2 + (S_{\bar{D}}(c))^2} \right\}.$$

Moreover, deriving the above expression with respect to the cutoff c and equating the derivative to zero, it follows that the slope of the ROC curve at this optimal cutpoint is equal to $S = \dfrac{1 - Sp(c)}{1 - Se(c)}$, where c denotes the optimal cutpoint.

This optimal criterion is commonly used in studies of diagnostic accuracy (Costello-Boerrigter et al., 2006; Wheeler et al., 2007; Al-Mawali et al., 2009; Bartlett et al., 2012). Often this strategy provides a point near the intersection of the ROC curve with the equation $y = 1 - x$, which is the negative diagonal of the unit square. Geometrically, this criterion is similar to the Pythagoras theorem, where the smallest hypotenuse should be chosen; or the shortest radius from the $(0, 1)$ point. This is the reason why the closest-to-(0,1) criterion is also known in the literature as *the geometric approach*.

It may not yield the same threshold as the Youden index. Indeed, Perkins and Schisterman (2006) do not recommend the use of the North-West corner because it involves the minimization of a quadratic term that does not own a clinical meaning, and advocate the use of the Youden index instead, which maximizes the sum of the two types of correct classifications. In addition, this method tends to overestimate the positive predictive value.

6. Maximization of concordance probability

Similarly to using the sum of sensitivity and specificity measures as the objective function, the product of both measures has been also considered in the literature to define another optimal criterion, in this case, the cutpoint that provides *the maximum product of sensitivity and specificity measures* (Johnson, 1998; Lewis et al., 2008):

$$c = \arg\max_{c} \{Se(c)Sp(c)\}.$$

According to Liu (2012) and Rota and Antolini (2014), the empirical estimate of this criterion has small bias when compared to the empirical estimate of the Youden index and minimum p-value approaches.

II. Criteria based on diagnostic likelihood ratios

7. Maximization of the diagnostic positive likelihood ratio
8. Minimization of the diagnostic negative likelihood ratio

When the aim of the diagnostic test is predictive, cutpoints based on the diagnostic likelihood ratio may be more useful (Boyko, 1994). In fact, optimal-cutpoint selection criteria based on DLR_+ and DLR_- have been proposed in the literature. Specifically, *the maximum diagnostic positive likelihood ratio*, that is, the cutpoint c maximizing $DLR_+(c) = Se(c)/(1 - Sp(c))$ (Caraguel et al., 2011); and *the minimum diagnostic negative likelihood ratio*, i.e., the value c minimizing $DLR_-(c) = (1 - Sp(c))/Se(c)$ (Caraguel et al., 2011).

III. Criteria based on predictive values

As with Se and Sp, in the case of PPV and NPV, there are similar strategies for selecting an optimal cutpoint (Vermont et al., 1991; Schapire et al., 1998; Gallop et al., 2003; Liu et al., 2005), such as selecting the point at which the predictive values are practically the same (Vermont et al., 1991), the point on the predictive ROC (PROC) curve closest to the point (0,1) (Vermont et al., 1991; Gallop et al., 2003), the average probability/suitability approach (Cramer, 2003; Liu et al., 2005) or the precision and recall-combined approaches: the called precision-recall break-even point and precision-recall plot-based approaches (Schapire et al., 1998; Liu et al., 2005), among others. In the following we explain in more detail these criteria based on the predictive values.

9. Maximization of positive predictive value
10. Minimization of negative predictive value

In some diagnostic situations, it is desirable to have the highest probability of a positive prediction or the lowest probability of a negative prediction and in such cases the optimal cutpoint should therefore be chosen taking into account these objectives. Therefore, one criterion for selecting the optimal value is based on the *maximization of positive predictive value* (Caraguel et al., 2011) or the *minimization of negative predictive value* (Bortheiry et al., 1994; Hoffman et al., 2000; Caraguel et al., 2011). So, in these methods the cutpoint c maximizing $PPV(c) = TP(c)/(TP(c) + FP(c))$ or minimizing $NPV(c) = TN(c)/(TN(c) + FN(c))$ respectively is chosen.

However, these procedures turn out to be somewhat extreme in the sense that the choice of an optimal cutpoint should generally imply equilibrium between both predictive values. Accordingly, the equality or simultaneous maximization of the positive predictive value and negative predictive value (Vermont et al., 1991; Gallop et al., 2003) or the maximization/minimization of a given combination of such quantities tends to be more appropriate in general.

11. Positive predictive value-Negative predictive value equality approach

Similarly to the criterion based on the equality of sensitivity and specificity measures, this strategy selects the threshold value c where both predictive values are similar or practically equal (Vermont et al., 1991), that is, the cutpoint c minimizing $|NPV(c) - PPV(c)|$. Sometimes, a level α is set and the optimal cutoff point is the value for which the absolute value of the difference between the positive predictive value and negative predictive value is below that level. When multiple cutpoints satisfy this condition, the optimal value is usually defined as the average of those points, and since in some practical situations, the positive predictive value achieved is not exactly equal to the negative predictive value, if a single value is desired, this is obtained, for example, by averaging the positive and negative predictive values (practically equal).

12. Simultaneous maximization of positive and negative predictive values

Instead of maximizing only one of the predictive values, this method maximizes both quantities simultaneously, that is, the optimal cutpoint c defined by this criterion maximizes the minimum of $NPV(c)$ and $PPV(c)$.

13. Maximization of Predictive summary index

Analogously to the Youden index that can be seen as a summary measure of the ROC curve, the predictive summary index (PSI) can be seen as a summary index of the PROC curve (Linn and Grunau, 2006; Zetterberg, 2006). For a fixed threshold value c, it is defined as follows:

$$\begin{aligned} PSI(c) &= PPV(c) + NPV(c) - 1 \\ &= \frac{pSe(c)}{pSe(c) + (1-p)(1-Sp(c))} + \frac{(1-p)Sp(c)}{(1-p)Sp(c) + p(1-Se(c))} - 1. \end{aligned}$$

The threshold value c that maximizes $PSI(c)$ is selected as the optimal cutpoint, that is

$$c = \arg\max_c \{PSI(c)\}.$$

The Predictive summary index provides more information than the Youden index and the predictive values, making it more appropriate in clinical settings. The inverse of the Predictive summary index is known as the "number needed to predict" ($NNP = 1/PSI$) and represents how many patients are needed to be examined in the population in order to correctly identify (predict) the positive diagnosis of one individual. It describes how much more likely the patient is to be correctly diagnosed with the disease after a positive test result, and how much more likely the patient is not to be incorrectly diagnosed with the disease after a negative test result.

14. Criterion of the point on PROC curve closest to (0,1) point

Considering that the ideal situation is when both predictive values are equal to one, a possible method for selecting the optimal cutpoint is to select the cutoff for which the pair of values $(PPV(c), NPV(c))$ is as close as possible to the pair (1,1). This is equivalent to the point on the PROC curve closest to the (0,1) point (Vermont et al., 1991; Gallop et al., 2003). So, the optimal cutpoint is the value c that minimizes the Euclidean distance between the pair $(PPV(c), NPV(c))$ and the point (1,1) (or equivalently, the distance between the PROC curve and the point (0,1)), that is, the point on the PROC curve which has the shortest distance to the top-left corner (0,1)):

$$c = \arg\min_c \{\sqrt{(1-PPV(c))^2 + (1-NPV(c))^2}\}$$

or, equivalently: $c = \arg\min_c \{(1-PPV(c))^2 + (1-NPV(c))^2\}$.

This strategy often provides a point near the intersection of the PROC curve with the equation $y = 1 - x$, which is the negative diagonal on the unit square passing through the points (0,1) and (1,0), that is, the corresponding predictive values are almost equal.

15. Maximization of the product of positive and negative predictive values

This criterion maximizes the product of the two predictive values (positive and negative), that is, $c = \arg\max_c \{NPV(c)PPV(c)\}$.

16. P-R (precision-recall) break-even point approach

Precision and recall are indices widely used in the field of information retrieval. Precision is the proportion of the retrieved items that are relevant, that is, the proportion of predicted presences that are real presences; here in diagnostic situations the presence of a target condition or disease, and therefore it corresponds to the positive predictive value. Recall is the proportion of the relevant items that are retrieved, which is equal to sensitivity.

P-R (precision-recall) break-even point approach for selecting the optimal cutpoint is based on selecting the optimal cutpoint as the value c where $PPV(c)$ (precision) = $Se(c)$ (recall), that is, the absolute value of the difference between precision and recall $|PPV(c) - Se(c)|$ is minimized (Schapire et al., 1998; Liu et al., 2005). Often, interpolation of the scores to obtain the break-even point is necessary. Interpolation gives values not achievable by the system. The point where recall equals precision is neither a desirable nor an informative target from a researcher's perspective (Schapire et al., 1998).

17. P-R (precision-recall) plot-based approach

The P-R (Precision-Recall) curve is the graph of sensitivity against positive predictive value. Similarly to the criteria of the point on ROC curve closest to (0,1) point and the point on PROC curve closest to (0,1), the point on the P-R (i.e, precision-recall) curve that is closest to the upper-right corner (1,1) in the P-R plot can also be used to determine the threshold, since the point in this corner represents a perfect classification with 100% precision and recall. So, according to this criterion, the cutoff corresponding to the point on the P-R curve which has the shortest distance to the top-right corner (1,1) is chosen as the optimal cutpoint (Schapire et al., 1998; Liu et al., 2005), that is,

$$c = \arg\min_c \left\{(1 - Se(c))^2 + (1 - PPV(c))^2\right\}.$$

IV. Prevalence-based methods

18. Observed prevalence
19. Mean predicted probability
20. Prevalence approach

Strategies have also been proposed for the selection of optimal prevalence-based cutpoints, designed mainly for situations in which the marker assumes values from 0 to 1 (prevalence values), e.g., the probabilities obtained on the basis of a statistical model (Manel et al., 2001; Kelly et al., 2008).

The observed prevalence criterion simply consists of selecting as optimal the value closest to the observed prevalence, that is,

$$c = \arg\min_c |c - \tilde{p}|,$$

where $\tilde{p}$ denotes the sample prevalence.

The mean predicted probability criterion chooses the closest value to the mean predicted probability.

The prevalence approach selects the cutoff value for which the prevalence predicted on the basis of the statistical model is practically equal to the observed prevalence (Liu et al., 2005), that is,

$$c = \arg\min_c |p(1 - Se(c)) - (1 - p)(1 - Sp(c))|.$$

Criteria 18 and 20 are useful strategies in cases in which preserving prevalence is of crucial interest.

V. Methods not based on ROC analysis

21. Split test values at median

It consists simply on selecting the optimal value as the median of the diagnostic test values (Liu et al., 2005). This criterion is not recommended mainly because it ignores the distribution of those with and without the target condition.

22. Maximum Chi-squared

Another approach for selecting the optimal cutpoint consists of maximizing a statistical test which represents the association between the marker and the binary result obtained based on the cutoff value (Mazumdar and Glassman, 2000). The minimal p-value criterion is based on either the exact Fisher test or the Pearson Chi-squared test applied on 2×2 contingency tables, where each table is a binary split of the test scale at all potential thresholds, that is, the pertinent Chi-squared test is calculated for each of the observed diagnostic marker values (candidates for the optimal cutpoint), except for the most extreme values. The optimal cutpoint is chosen at the split/point for which the maximum Chi-squared or, equivalently, the corresponding minimum p-value is obtained (Miller and Siegmund, 1982; Mazumdar and Glassman, 2000). This criterion has been criticized mainly due to multiplicity (Altman et al., 1994); so, in a attempt to overcome the most serious problems associated with multiple testing, several correction methods have been proposed for adjusting for the increase of the type-I error which is associated with the minimum p-value approach, such as the maximally selected rank statistical method (Schulgen et al., 1994; Lausen and Schumacher, 1996) or the use of a permutation test approach (Hilsenbeck and Clark, 1996). The former is an easily applicable method but has the drawback of being too conservative in cases where there are few cutpoints.

VI. Other methods

23. Maximization of the diagnostic odds ratio

Many quantitative indicators of the performance of a diagnostic test have been introduced in the literature, some already discussed, such as sensitivity, specificity, predictive values, concordance, likelihood ratios, and area under the ROC curve (AUC), among others. Less known is the odds ratio (OR) as a simple indicator of the effectiveness.

The OR is a very well-known statistic used in epidemiology, expressing the strength of association between exposure and disease. It can also be applied to measure the strength of association between the test results and the presence of disease. In a diagnostic context, it is commonly known as the diagnostic odds ratio and denoted by DOR rather than OR.

For each cutpoint c, the DOR of a diagnostic test is the ratio of the odds of a positive outcome in the diseased group relative to the odds of a positive result in the non-diseased individuals (Kraemer, 1992), that is,

$$DOR(c) = \frac{\frac{TP(c)}{FN(c)}}{\frac{FP(c)}{TN(c)}} = \frac{\frac{Se(c)}{(1-Se(c))}}{\frac{(1-Sp(c))}{Sp(c)}} = \frac{Se(c)}{(1-Se(c))}\frac{Sp(c)}{(1-Sp(c))}.$$

Alternatively, the DOR can be understood or interpreted as the ratio of the odds of disease in positive test results relative to the odds of disease in negative test results:

$$DOR(c) = \frac{\frac{TP(c)}{FP(c)}}{\frac{FN(c)}{TN(c)}} = \frac{\frac{PPV(c)}{(1-PPV(c))}}{\frac{(1-NPV(c))}{NPV(c)}} = \frac{PPV(c)}{(1-PPV(c))}\frac{NPV(c)}{(1-NPV(c))}.$$

Moreover, the DOR can be expressed in terms of the likelihood ratios:

$$DOR(c) = \frac{\frac{TP(c)}{FP(c)}}{\frac{FN(c)}{TN(c)}} = \frac{DLR+}{DLR-}.$$

The possible values for DOR are between 0 and ∞, with higher values indicating better discriminatory power of the test. A value of 1 means that the test does not discriminate at all between patients with and patients without disease. Lower values of 1 indicate an incorrect interpretation of the test (more negative tests among diseased individuals, that is, a greater number of false negatives). Note that the inverse of the DOR can be interpreted as the ratio of the odds of negative test results among diseased patients relative to the odds of negative results in the healthy group. The DOR increases abruptly where the sensitivity or specificity are almost perfect (Kraemer, 1992).

As concluded from the above formulas, the DOR does not depend on the disease prevalence. Another point to consider is that it can not be used as a global measure to judge the error rates of the test. Two tests with the same DOR can have a sensitivity and a specificity very different, with different clinical consequences.

If the 2×2 contingency table of the different diagnostic test decisions contains zeros, the DOR is not defined. Adding 0.5 to all the cells in the table is a common method for computing an approximation of the DOR in this case (Haldane, 1955; Littenberg and Moses, 1993).

According to the DOR, how to select the optimal cutpoint? Some authors (Greiner et al., 2000; Böhning et al., 2011; Caraguel et al., 2011) compute the optimal cutpoint c based on the maximization of the DOR, that is,

$$c = \arg\max_c \{DOR(c)\} = \arg\max_c \left\{ \frac{\frac{Se(c)}{(1-Se)(c)}}{\frac{(1-Sp)(c)}{Sp(c)}} \right\} = \arg\max_c \left\{ \frac{DLR_+}{DLR_-} \right\}.$$

Other authors (Magder and Fix, 2003) point out that the optimal cutpoint should be selected as the value that maximizes the precision of the DOR estimate, for instance, the mean squared error (MSE), which is the mean squared distance between the estimate and the true value (Bickel and Doksum, 1977). Moreover, for determining the precision of the DOR estimate it is very common and convenient to work on a logarithmic scale. Given the estimate, the problem is thus reduced to find the cutoff value minimizing the MSE.

The main advantage of this criterion is that the OR (here, the DOR) has several well-known mathematical properties. Besides, it is appealing because it is an accuracy measure that combines both Se and Sp. However, it presents some disadvantages. For instance, it is noisy (unstable) along the range of test results due to its multiplicative nature and its log transformation leads to a convex curve. In addition, it may easily lead to cutoff values on the boundary of the parameter range (Böhning et al., 2011). For this reason, this criterion is not recommended for the selection of the optimal cutpoint in clinical practice.

24. Maximum Kappa index

The variation of an observer in himself, a gold standard or other observers can be measured by means of agreement to examine and classify a number of elements (patients, radiographs, biological samples, ...). In this case we have the situation of a diagnostic test (which plays the role of the observer) and a reference test, which here is the true disease state of the patient. Several indexes of agreement that provide quantitative information that attempts to measure the degree of agreement and/or variation have been proposed. For instance, the so called *Kappa (K) index* originally proposed by Cohen (1960) for the case of two evaluators or two methods. This index relates the agreement that the observer shows with the standard or reference test, beyond the potential agreement by chance.

How to compute the Kappa index?
In essence, the process of constructing the index is as follows: the difference between the proportion of observed agreement and the proportion of agreement expected by chance

is computed; if it is zero, then the degree of agreement that has been observed can be entirely attributed to chance; if the difference is positive, it indicates that the degree of agreement is higher than it was expected by chance; and vice versa, if the difference is negative (really unlikely) then the degree of agreement is less than it was expected by chance.

The Kappa index is the ratio between that amount and the maximum expected without the intervention of chance

$$K = \frac{P_0 - P_e}{1 - P_e},$$

where P_0 is the ratio of observed agreements and P_e is the the ratio of expected agreements by chance, that is, the proportion of agreements on the assumption of independence between observers, that is, agreements by chance.

From the corresponding 2×2 contingency table which includes the true positive (TP), false positive (FP), false negative (FN) and true negative (TN) decisions of the diagnostic test, the observed agreement is computed as follows:

$$P_0 = \frac{TP + TN}{TP + FP + FN + TN}.$$

Once you have the observed values of the four types of decisions taken, the expected values, under the assumption that the classification had been made at random, are constructed from those:

$$TP_e = \frac{(TP + FP)(TP + FN)}{TP + FP + FN + TN},$$
$$TN_e = \frac{(TN + FN)(TN + FP)}{TP + FP + FN + TN},$$

where TP_e and TN_e denote the number of true positive and true negative classifications that would be expected by chance, respectively. Thus, the ratio of expected agreements by chance (potential agreement) is calculated as follows:

$$P_e = \frac{TP_e + TN_e}{TP + FP + FN + TN}.$$

The Kappa index takes values between -1 and 1. The maximum value is reached if there is total agreement between the classification of the diagnostic test and the true state, that is, it occurs only when the observed agreement is 100% (a perfect agreement). This would correspond to an ideal test with a sensitivity and specificity of 100%.

The question about the Kappa index refers to how it can be considered as an indicator of good agreement. There is no an exact answer to this issue, since what is considered appropriate (or not) will depend on the problem under study. Landis and Koch (1977) considered acceptable values for the Kappa index those greater than or equal to 0.40 and excellent those greater than 0.75, and proposed the following scale value interpretation:

- $\leq$ 0,00: without agreement
- 0,00 - 0,20: insignificant
- 0,21 - 0,40: discrete

- 0,41 - 0,60: moderate
- 0,61 - 0,80: substantial
- 0,81 - 1,00: almost perfect

How to compute the optimal cutpoint from the Kappa index?
The optimal cutpoint c derived from the Kappa index is simply that which reaches the maximum value of the Kappa index.

3.3 Methods with researcher (or user) requirements

Unlike the completely data driven methods, which rely on maximizing accuracy without considering usefulness, which has been defined as the practical clinical value of the information provided by the marker (Zweig and Campbell, 1993), some authors have addressed the question of incorporating usefulness into the cutpoint estimation (see, for example, Metz, 1978; DeNeef and Kent, 1993; Zweig and Campbell, 1993; Halpern et al., 1996). Usefulness depends on the prevalence and costs of classification made on the basis of the test results and thus, several kinds of usefulness could be defined with the same accuracy level.

We use the term "methods with researcher or user requirements" to refer to those methods that use the test values and also external information set by the researcher to compute the optimal threshold value. This external information often includes the disease prevalence, and/or the costs of wrong classifications or utilities of correct classifications. The methods included in this group will give the researcher options either to penalize the wrong classifications (loss or cost), or weight the correct classifications (utility or profit) preferentially toward those with or without the target condition. In essence, utilities and costs are weights which will represent the researcher preferences toward either sensitivity or specificity.

It should be noted that the validity of the cutpoints or decision thresholds generated by some of the methods included in this group do not only depend on the sampling strategy, but also on the veracity of the assumptions, that is, the values fixed by the researcher. This can be seen as a disadvantage because it may happen that different users set different values for the utilities and costs under the same setting that lead to different thresholds. The minimum average cost method is the most frequently studied in this group of methods.

We will review some of the optimal cutoff selection criteria that can be included in this group of methods with researcher requirements. As you can see from the list below, some of these methods have been additionally grouped in different subcategories: I. Criteria based on sensitivity and specificity measures; II. Criteria based on predictive values; III. Criteria based on diagnostic likelihood ratios; IV. Methodology based on cost-benefit analysis of the diagnosis.

I. Criteria based on sensitivity and specificity measures

- 1. Required specificity (pre-established minimum value for specificity or maximum sensitivity at fixed specificity)

- 2. Required sensitivity (pre-established minimum value for sensitivity or maximum specificity at fixed sensitivity)
- 3. Required sensitivity and specificity (or pre-established minimum values for sensitivity and specificity)
- 4. Range of pre-established minimum values for sensitivity
- 5. Range of pre-established minimum values for specificity
- 6. Pre-established value for specificity
- 7. Pre-established value for sensitivity

II. Criteria based on predictive values

- 8. Required negative predictive value (pre-established minimum value for negative predictive value or maximum positive predictive value at fixed negative predictive value)
- 9. Required positive predictive value (pre-established minimum value for positive predictive value or maximum negative predictive value at fixed positive predictive value)
- 10. Required negative and positive predictive values (or pre-established minimum values for negative predictive value and positive predictive value)
- 11. Set an arbitrary positive predictive value
- 12. Set an arbitrary negative predictive value
- 13. Posterior probabilities
- 14. Information theory (or mutual information)

III. Criteria based on diagnostic likelihood ratios

- 15. Set an arbitrary diagnostic negative likelihood ratio
- 16. Set an arbitrary diagnostic positive likelihood ratio

IV. Methodology based on cost-benefit analysis of the diagnosis

- 17. Cost-benefit ratio (minimum average cost or slope of isoutility) criterion
- 18. Particular cases of the cost-benefit ratio criterion
- 19. Maximization of expected utility (highest net utility or indifference curve approach)
- 20. Criterion based on the relative utility (RU) curve
- 21. F-measure (or F approach) criterion

I. Criteria based on sensitivity and specificity measures

1. Required specificity

This criterion for selecting the optimal value is adequate when it is desirable to have a higher probability of detecting a true negative result (specificity) than a true positive one (sensitivity). So, in this case, the optimal cutpoint is the test value which increases the accuracy of detecting a true negative result. This can be treated in two different ways:

a) A minimum value is chosen above which the specificity must be, that is, specificity is larger than or equal to the minimum value set and the sensitivity must be as high as possible (Vermont et al., 1991; Gallop et al., 2003; Hadzi-Pavlovic, 2008; Bartlett et al., 2012)

b) A desired specificity Sp is set and the upper limit of a confidence interval for the true unknown Sp percentile of the distribution of the diagnostic test in the healthy population is selected as the optimal cutpoint (Clopper and Pearson, 1934; Greenhouse and Mantel, 1950; David, 1970; Tietz, 1986; Linnet, 1987; Schäfer, 1989).

The advantage of this criterion is that it is a simple method, not influenced by the disease prevalence. However, the main disadvantage is that the error of misclassifying those without the condition is controlled at the expense of completely disregarding the rate of misclassifying those with the condition. Besides, the optimal cutpoint obtained with this method may vary substantially as different preferred values for Sp are set (Freeman and Moisen, 2008).

2. Required sensitivity

Analogously to the previous situation, sometimes the interest is the selection of a cutoff to increase the accuracy of detecting a true positive result (sensitivity). For this, one possibility is that the sensitivity is higher than or equal to a predetermined minimum value x (Tietz, 1986; Schäfer, 1989; Vermont et al., 1991; Gallop et al., 2003; Hadzi-Pavlovic, 2008; Bartlett et al., 2012), that is, the optimal cutpoint c is the value that maximizes $Sp(c)$ while $Se(c) \geq x$.

Similarly to the previous method, this method enjoys the same advantages: it is a simple method and it is not influenced by the disease prevalence. However, its main disadvantage now is that the error of misclassification of those with the condition is controlled at the expense of completely disregarding the rate of misclassifying those without the condition.

In both procedures (required specificity and required sensitivity) the preselected minimum level is usually fairly high. Since the objective is to maximize the probability of a correct decisions (either TN or TP), a possible choice that makes sense is to consider the set of values traditionally used for the power levels in hypothesis testing, such as 80%, 90% or 95%. For example, when the required specificity level is 80%, you seek for a probability of at least 80% for the correct classification of a negative result.

3. Required sensitivity and specificity

Selecting an optimal cutpoint usually involves a balance between sensitivity and specificity. It may be the case that rather than a clear preference to a high sensitivity or a high specificity, minimum values are desired for both measures. This approach can be solved in two different ways:

a) A minimum value is set above which the specificity must be and also a minimum value is set above which the sensitivity must be.

b) A desired specificity (Sp) and a desired sensitivity (Se) are fixed, and a weighted average of the corresponding estimated Sp and Se percentiles in the healthy and diseased populations, respectively, is selected as the optimal cutpoint (Schäfer, 1989).

In a similar way to the previous strategies, analogous criteria can be defined, in which the end-user, rather than setting a single minimum value for either or both measures, sets a range of values between those you want to place the measure Se or Sp, or a single target value for either the sensitivity or the specificity measures (Navarro et al., 1998; Rutter and Miglioretti, 2003).

4. Range of pre-established minimum values for sensitivity

Some diagnostic tests as for instance, screening tests pursue to detect a high percentage of patients with the disease (high sensitivity). Therefore, in these cases, it is usual to compute the optimal cutpoint by setting the sensitivity between certain limits and maximizing the specificity. For example, Navarro et al. (1998) studied the diagnostic accuracy of the "Child Behavior Checklist" (CBCL) questionnaire as a screening test for detecting the presence of phenomena of child psychopathology, and they computed the optimal cutoff by two criteria: a) Set the sensitivity between 0.75 and 0.85, and maximize the specificity, and b) set the sensitivity between 0.85 and 0.95 and maximize the specificity.

5. Range of pre-established minimum values for specificity

Sometimes the use of a specific diagnostic test needs to detect a high percentage of cases without the disease (high specificity). In this situation, a possible option is to chose the optimal cutoff by the criterion that sets a range of values for the specificity and maximizes the sensitivity subject to that condition on the specificity.

6. Pre-established value for specificity

Estimating optimal cutoff points requires the specification of optimal characteristics for the resulting binary test. Sometimes, the interest is to select a cutpoint that provides a binary test with a specific TPF or specificity. For example, Rutter and Miglioretti (2003) for estimating the accuracy of screening psychological tests, computed optimal cutpoints based on a specificity value of 80%.

7. Pre-established value for sensitivity

Similarly to the previous criterion, the optimal cutpoint in this case is computed based on a binary test that has a certain sensitivity. For instance, a sensitivity of 80% (Rutter and Miglioretti, 2003). Analogously to the criteria on which a minimum value was fixed for the sensitivity, specificity or both accuracy measures, the pre-established value is also usually a high value, based on the concept of statistical power in a hypothesis testing context.

The previous criteria have been established in terms of the specificity and/or the sensitivity. But they could be reformulated in terms of their complementaries, the FPF and FNF. For example, to set a false positive fraction (or rate) of 20% maximum is the same as to set a specificity of 80% minimum.

II. Criteria based on predictive values

As we pointed before in previous chapter, sometimes it is more interesting for a clinician to know what is the probability that an individual who has tested positive is actually diseased, and vice-versa, i.e., the probability that an individual who has tested negative is actually disease-free. There are indeed several strategies to select optimal cutpoints based on the positive and negative predicitive values, PPV and NPV (Vermont et al., 1991), similar to those previously discussed based on the sensitivity (Se) and specificity (Sp) indexes .

8. Required negative predictive value

This criterion for selecting the optimal value is adequate in diagnostic situations where it is desirable to have a higher probability of predicting a true negative result (negative predictive value) than a true positive one (positive predictive value). For this, a minimum value is set for the negative predictive value (Vermont et al., 1991) and the cutpoint c that maximizes the positive predictive value $PPV(c)$, subject to that restriction on the negative predictive value, $NPV(c) \geq x$, is defined as the optimal cutpoint, where x represents the pre-established minimum value for NPV.

9. Required positive predictive value

Similarly to the previous criterion, when it is desirable to have a higher probability of predicting a true positive result (positive predictive value) than a true negative one (negative predictive value), the optimal cutpoint must meet this objective and so a minimum value is set for the positive predictive value (Vermont et al., 1991) and the cutpoint c defined as optimal in this setting is that which maximizes $NPV(c)$ subject to the restriction that $PPV(c) \geq x$, where x represent the pre-established minimum value for PPV.

10. Required negative and positive predictive values

In some clinical scenarios, there may not be a clear preference to a high negative predictive value or a high positive predictive value, and that minimum values are desired for both measures, having thus an equilibrium between both predictive values. Under this situation, the optimal cutpoint is the value c that satisfies both restrictions $NPV(c) \geq x$ and $PPV(c) \geq y$, where x and y are the pre-established minimum values for each of the predictive values, NPV and PPV, respectively.

11. Set an arbitrary negative predictive value

In this criterion, a specific value is set for the negative predictive value, and the cutpoint c satisfying the condition $NPV(c) = x$ is selected as the optimal cutpoint (Hadzi-Pavlovic, 2008; Caraguel et al., 2011; Bartlett et al., 2012), where x is the target value previously set for the negative predictive value.

12. Set an arbitrary positive predictive value

Analogously to the previous criterion, in this case the optimal cutpoint is selected as the value c verifying that $PPV(c) = x$, where x is the target value previously set for the positive predictive value (Hadzi-Pavlovic, 2008; Caraguel et al., 2011; Bartlett et al., 2012).

Other strategies that also involve the predictive values are the following:

13. Posterior probabilities

This criterion defines the optimal cutpoint as the value c for which the maximum of the values of each of the two posterior to prior probability ratios is greater than one, that is, the quotient between the positive predictive value and the disease prevalence has to satisfy the restriction $PPV/p > 1$ and, similarly, the ratio between the negative predictive value and the prior probability of absence of disease has to satisfy the restriction $NPV/(1-p) > 1$ (Schuchard and Massof, 1990).

A probability ratio value of one indicates that there is no gain by performing the test, that is, the probability of having or not having the disease is the same with or without the test results.

14. Information theory (or mutual information)

If we consider the situation of transmitting a message (in this case the diagnostic information) through an imperfect communication channel (here the diagnostic test), since the communication chanel is imperfect (the diagnostic test is not error-free), the received information may (or not) be the same as the transmitted information. Thus, for a given clinical decision making process, the unique two possible received messages correspond to the diagnostic test decisions about the presence or absence of disease and the two

possible transmitted messages correspond to the actual status of the patient with two possible states (diseased or healthy). Therefore, the average information content of a diagnostic test can be found by evaluating the average reduction in uncertainty about the transmitted signal that is provided by the received signal (Schuchard and Massof, 1990), that is, it is the difference between *"average uncertainty about the disease before performing the test"* and *"average uncertainty about the disease after receiving the diagnostic information"*.

Therefore, the average information content (expressed as the fraction of information per patient in the case of diagnostic tests) can be expressed as (Metz et al., 1973):

$$\begin{aligned} I_{ave} &= pSe\log_2\frac{PPV}{p} + (1-p)(1-Sp)\log_2\frac{1-NPV}{1-p} \\ &\quad +p(1-Se)\log_2\frac{1-PPV}{p} + (1-p)Sp\log_2\frac{NPV}{1-p} \end{aligned} \tag{3.1}$$

Note that because the probabilistic events are binary, the base 2 logarithm is used in (3.1).

Under this setting, the value c that maximizes I_{ave} is selected as the optimal value to discriminate between diseased and healthy individuals (Schuchard and Massof, 1990; Zou et al., 2013). Note that if the information was perfectly transmitted by the diagnostic test, the average information content would be 1 (the maximum value).

We remind here that although the ROC curve also allows the graphical representation of the optimal cutpoints associated with these strategies based on the predictive values and these strategies are considered indeed within the ROC methodology, there is a more adequate curve in this case, the **predictive ROC curve** (PROC) (Vermont et al., 1991; Gallop et al., 2003) computed from the predictive values.

III. Criteria based on diagnostic likelihood ratios

Where the aim of the diagnostic test is predictive, cutpoints based on the diagnostic likelihood ratio (DLR) may be more useful (Boyko, 1994). So, optimal-cutpoint selection criteria based on pre-established values of these measures have been also proposed, similarly to those previously described for Se, Sp, PPV and NPV (Rutter and Miglioretti, 2003).

15. Set an arbitrary diagnostic negative likelihood ratio

In this criterion, a target value is set for the negative diagnostic likelihood ratio, that is, the cutpoint c that satisfies the condition $DLR-(c) = x$ is selected as the optimal cutpoint (Boyko, 1994; Rutter and Miglioretti, 2003; Hadzi-Pavlovic, 2008; Caraguel et al., 2011), where x is the prefixed value set for the diagnostic negative likelihood ratio.

16. Set an arbitrary diagnostic positive likelihood ratio

Analogously to the previous criterion, in this case the optimal cutpoint is selected as the value c that verifies the condition $DLR+(c) = x$, where x is the target value set for

the diagnostic positive likelihood ratio (Boyko, 1994; Rutter and Miglioretti, 2003; Hadzi-Pavlovic, 2008; Caraguel et al., 2011).

IV. Methodology based on cost-benefit analysis of the diagnosis

17. Cost-benefit ratio (minimum average cost or slope of isoutility) criterion

When undertaking a diagnostic procedure, a price is paid (in terms of money and/or risk of possible complications) to "gain" information that may be beneficial for the subsequent treatment and care of the patient. According to ROC methodology, this extra information on the patient can be measured and described in a statistical sense, by attempting to answer the following questions: 1) How can the benefits obtained from correct diagnostic decisions be balanced (offset) against the costs of incorrect decisions?; and 2) How can we judge where the additional information is worth the price "paid"?

Although a ROC curve describes the different balances that can be obtained between the relative frequencies of true positives, false negatives, true negatives and false positives, the particular compromise that turns to be more effective in practice depends on the disease prevalence and the "utilities" (benefits and costs) of the four different decisions that can be taken based on a diagnostic test. So, this optimal compromise between the relative frequencies of the different decisions can be studied in terms of *the expected net benefit* of the diagnostic system when such a system is applied to a population of patients (McNeill et al., 1975; Metz et al., 1975; Metz, 1978; Swets and Swets, 1979).

Therefore, from the combination of ROC analysis and *statistical decision theory*, we can study how to compute the mean cost of the consequences of performing a diagnostic test. In this case, the benefits and costs of each type of decision are combined with the disease prevalence to find the operating point on the ROC curve, that is, the coordinate point $(1-Sp(c), Se(c))$, that wild yield the minimum mean overall cost (maximum mean overall benefit) (McNeill et al., 1975; Metz et al., 1975; Metz, 1978; Swets and Swets, 1979), where the term "cost" can be construed as a combination of various aspects (for example, adverse health risks) and not exclusively as a monetary term (Edwards et al., 1975). This is the so called *cost-benefit ratio criterion or minimum average cost criterion* (Zweig and Campbell, 1993; Cantor et al., 1999; Zhou et al., 2002; Obuchowski, 2005; Hadzi-Pavlovic, 2008; Kaivanto, 2008). Note that "benefits" can be expressed as negative costs. However, in the following exposition of the cost-benefit ratio criterion discussion, we will express the consequences of all the decisions (either correct or incorrect) in terms of costs.

The expected net benefit or mean overall cost of the consequences of conducting a diagnostic test should include, firstly, the price that must be paid for performing the test ("overhead cost", C_0, that in general summarizes the direct costs and effects of testing), and the costs of the medical consequences derived from each type of diagnostic decision, weighted by its corresponding probability of occurrence. Hence, for a situation where there are two possible alternative decisions (though it may easily be extended to situations with a larger number of decisions), the expected cost C of the use of the diagnostic test can be expressed in terms of the threshold value c as follows:

$$C(c) = C_0 + C_{TP}\Pr(TP(c)) + C_{TN}\Pr(TN(c)) + C_{FP}\Pr(FP(c)) + C_{FN}\Pr(FN(c)),$$

where C_{TP}, C_{TN}, C_{FP}, and C_{FN} represent the mean costs of the medical consequences derived from each type of diagnostic decision (true positive, true negative, false positive

and false negative, respectively). So, C_{TP} and C_{TN} denote the classification costs when classifying correctly either a diseased or a healthy individual, whereas C_{FP} and C_{FN} are the costs when classifying incorrectly either a healthy or diseased individual. These costs weigh each diagnostic decision with respect to its clinical relevance.

It should be noted that, for example, the consequences of a FP decision generally represent a responsibility, although almost always a lower responsibility than the responsibility of a FN decision. The cost of a false positive decision C_{FP} represents often an inappropriate (over)treatment; and the cost of a false negative result C_{FN} is typically the worst outcome of missed case. C_{TN}, the cost of a true negative result, is typically the best outcome of not having the disease; and C_{TP}, the cost of a true positive misclassification, represents in general an appropriate treatment. Computing the probabilities of the four possible decisions, the expression of the expected cost $C(c)$ above, can be rewritten as follows:

$$\begin{aligned} C(c) &= C_0 + C_{TP}\Pr(D=1)\Pr(Y+|D=1) + C_{TN}\Pr(D=0)\Pr(Y-|D=0) \\ &\quad + C_{FP}P(D=0)\Pr(Y+|D=0) + C_{FN}\Pr(D=1)\Pr(Y-|D=1). \end{aligned}$$

Therefore:

$$\begin{aligned} C(c) &= C_0 + C_{TP}pSe(c) + C_{TN}(1-p)Sp(c) \\ &\quad + C_{FP}(1-p)(1-Sp(c)) + C_{FN}p(1-Se(c)) \\ &= C_0 + C_{TP}pSe(c) + C_{TN}(1-p)(1-(1-Sp(c)) \\ &\quad + C_{FP}(1-p)(1-Sp(c)) + C_{FN}p(1-Se(c)), \end{aligned} \tag{3.2}$$

where p is the disease prevalence. Rearranging terms, it follows that:

$$\begin{aligned} C(c) &= -(C_{FN}-C_{TP})pSe(c) + (C_{FP}-C_{TN})(1-p)(1-Sp(c)) \\ &\quad + C_0 + C_{TN}(1-p) + C_{FN}p. \end{aligned}$$

Hence the above expression is of the form:

$$C(c) = K_1 Se(c) + K_2(1-Sp(c)) + K_3, \tag{3.3}$$

where K_1, K_2, K_3 are real constants that do not depend on the selected cutpoint c,

$$\begin{aligned} K_1 &= -(C_{FN}-C_{TP})p, \\ K_2 &= (C_{FP}-C_{TN})(1-p) \\ K_3 &= C_0 + C_{TN}(1-p) + C_{FN}p. \end{aligned}$$

So, the cost-benefit ratio criterion selects the optimal cutpoint as the value that minimizes (3.3). We prove below that the solution of this minimization problem defines the optimal point where the slope of the ROC curve ("slope of isoutility") is given by (Lusted, 1968; Metz, 1978; Weinstein and Fineberg, 1980; England, 1988; Schuchard and Massof, 1990; John, 1992; Dwyer, 1996; Halpern et al., 1996; Martín-Andrés and Luna del Castillo, 2004):

$$S = \frac{1-p}{p}\frac{C_{FP}-C_{TN}}{C_{FN}-C_{TP}} = \frac{1-p}{p}\frac{C}{B} \tag{3.4}$$

Taking the first derivative of $C(c)$ with respect to $1 - Sp(c) = \Pr(Y + |D = 0)$ gives:

$$\frac{dC(c)}{d(1-Sp(c))} = K_1 \frac{dSe(c)}{d(1-Sp(c))} + K_2 = (C_{TP} - C_{FN})p \frac{dSe(c)}{d(1-Sp(c))} + (C_{FP} - C_{TN})(1-p).$$

Equating to zero and solving the resulting equation gives:

$$(C_{TP} - C_{FN})p \frac{dSe(c)}{d(1-Sp(c))} + (C_{FP} - C_{TN})(1-p) = 0,$$

that is,

$$\frac{dSe(c)}{d(1-Sp(c))} = \frac{C_{TN} - C_{FP}}{C_{TP} - C_{FN}} \frac{1-p}{p} = \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \frac{1-p}{p},$$

which proves equation (3.4).

So, other possible equivalent forms to compute the optimal value that minimizes the cost of diagnosis is to select the value where the slope of the ROC curve is equal to S. Besides, some authors instead of minimizing the expression of the mean cost given above in (3.3), they maximize the following expression:

$$Se(c) - \frac{1-p}{p} \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} (1 - Sp(c)),$$

which is equivalent. Based on these different perspectives, this optimal cutpoint is known in the literature as the *minimum average cost; slope of isoutility; or cost-benefit optimal criterion* (Lusted, 1968; Metz, 1978; England, 1988; Schuchard and Massof, 1990; John, 1992; Dwyer, 1996; Halpern et al., 1996; Zhou et al., 2002; Gallop et al., 2003; Pepe, 2004; Brown and Davis, 2006).

It is interesting to mention here that the S slope given in (3.4) (also called the Metz equation) actually weighs the decisions about non-diseased subjects versus diseased ones, highlighting their relative importance. This enables the benefits of correct diagnostic decisions to be balanced against the costs of incorrect decisions (Metz, 1978). The S term has to be a positive number in order to find a solution, and thus it is required that the costs corresponding to wrong decisions are greater than those corresponding to correct decisions, an assumption which is also supported by common sense. S could also be positive if the costs of the two correct decisions were higher than the costs of the corresponding wrong decisions, but when looking for a solution, this would result in a negative second derivative for the cost function, that is, in a maximum rather than a minimum of the cost function (Skaltsa et al., 2010). Therefore, the three key conditions for diagnosis to be cost-effective are that:

1. C_{TP} must be lower than C_{FN},
2. C_{TN} must be lower than C_{FP},
3. C_0 must not make the equation positive.

The first two conditions must be met because if we were indifferent as to the treatment of the disease-negative subgroup (that is the false positives, FP), we would simply classify everyone as positive, and vice versa.

The first term of the S slope given in equation (3.4), $((1-p)/p)$), is known as the "a priori" non-disease odds and it is the inverse of the "a priori" disease odds, generally denoted by Ω, $\Omega = p/(1-p)$. Thus, the prior probability of disease can be computed from the "a priori" disease odds by the following relationship: $p = \Omega/(1+\Omega)$ (Lusted, 1968). For diseases of low prevalence, that is, rare diseases (which typically result in more false positives than true positives) and for situations in which a false positive would result in painful or dangerous testing or treatment, the analyst should select a cutpoint that yields fewer false positives. Such a cutpoint should be chosen from the segment of the ROC curve in the lower-left quadrant of the ROC plot (Cantor et al., 1999). Conversely, for highly prevalent diseases and for conditions in which the number of false negative test results should be minimized, the user should select a cutpoint that limits the number of false negatives. Such a point should be chosen from the segment of the ROC curve in the upper-right quadrant of the ROC plot.

The second term of the S slope consists of the cost-benefit ratio C/B, which allows the cost ratio to be tuned to the non-disease odds as desired. Thus, S can be seen as a cost-benefit ratio modified by the ratio of "a priori" probabilities and, furthermore, is closely related to the sensitivity and specificity indexes, in the way that $S = 1$ implies that the minimum cost threshold gives the same relevance or weight to sensitivity and specificity; and conversely, values of S higher (lower) than 1 turn out on cutpoints that give higher (lower) specificity and lower (higher) sensitivity than the cutpoint corresponding to $S = 1$ (Skaltsa et al., 2010).

Assuming a binormal model for the diagnostic test, an explicit formula for the optimal cutpoint defined by the cost-benefit ratio criterion is given by (Jund et al., 2005; Skaltsa et al., 2010):

$$c = \frac{2\sigma^2 \log S - (\mu_{\bar{D}}^2 - \mu_D^2)}{2(\mu_D^2 - \mu_{\bar{D}}^2)},$$

where μ_D^2 is the diseased population mean, $\mu_{\bar{D}}^2$ is the non-diseased population mean, and σ is the common variance in both populations. When considering, however, unequal variances, a more realistic clinical setting (Hanley, 1988, 1996), the optimal cutpoint is the root of a second-degree univariate equation given by:

$$c = \frac{\sigma_D^2 \mu_{\bar{D}} - \sigma_{\bar{D}}^2 \mu_D + \sigma_{\bar{D}} \sigma_D \sqrt{(\mu_D - \mu_{\bar{D}})^2 + 2\log(S\sigma_D/\sigma_{\bar{D}})(\sigma_D^2 - \sigma_{\bar{D}}^2)}}{\sigma_D^2 - \sigma_{\bar{D}}^2}.$$

In the literature, this is called the *parametric minimum cost threshold with p, Se and Sp weights (binormal with unequal variance); or parametric maximum utility or expected value on the decision analysis criterion* (Jund et al., 2005; Skaltsa et al., 2010).

If the probability functions in both populations are unknown, an alternative is to use an empirical estimator, that is, for each possible cutpoint c (any of the observed values of Y), the empirical true and false positive fractions are computed and based on them an empirical cost function is obtained. Thus, the empirical cutpoint c will correspond to

the observed value of Y that provides a minimum of the empirical cost function. This cutpoint is known in the literature as the *empirical minimum cost threshold*.
Another alternative when distributions of the populations are unknown, it to base the estimation on the slope of isoutility S. For this, it is important to take into account the relationship between the ROC curve and the likelihood ratio (LR) (Greiner et al., 2000). The slope of the ROC curve takes values between 0 (at the right upper corner) and ∞ (at the lower left corner) and it is equivalent to the theoretical likelihood ratio (LR) function

$$LR(y) = \frac{\Pr(Y = y|D = 1)}{\Pr(Y = y|D = 0)}, \quad y \in \Re,$$

defined as the ratio of the probability of observing a test result y in the diseased individuals, $\Pr(Y = y|D = 1)$, and the probability of observing the same result in the healthy patients, $\Pr(Y = y|D = 0)$. So, a solution would be to compute the optimal cutpoint as the value for which the likelihood ratio is equal or closer to the S value. More specifically, Choi (1998) proved that the slope between two operating points on the ROC curve coincides with the likelihood ratio for a specific test value bounded by the test values associated to those operating points. Another possibility is to consider an appropriate statistical model for the LR function, for instance, the logistic regression model given by

$$\text{logit}\,\Pr(D = 1|Y = y) = a + by + \epsilon, \tag{3.5}$$

where a and b are the estimated coefficients, ϵ is the error term and $\Pr(D = 1|Y = y)$ denotes the "a posteriori" disease probability given the test value y. Under this model, the coefficient a depends on the sample prevalence $\tilde{p}$. Besides, if we consider the notation $\tilde{y}$ to refer to the value of the test Y that does not change the "a priori" disease probability, that is, $\tilde{y} = (\text{logit}(\tilde{p}) - a)/b$, then the LR function for the diagnostic test Y derived from model (3.5) is given by (Simel et al., 1993):

$$LR(y) = \exp\{b(y - \tilde{y})\}.$$

Therefore, based on the slope term S given in (3.4), the cost-benefit optimal criterion can be obtained as the root of the equation in y that sets $LR(y) = S$, that is,

$$\exp\{b(y - \tilde{y})\} = \frac{1 - p}{p} C/B,$$

where C/B denotes the ratio of the net cost of treating the healthy individuals ($C = C_{FP} - C_{TN}$) to the net benefit of treating the diseased individuals ($B = C_{FN} - C_{TP}$). The root of this equation is given by (Anderson, 1982):

$$c = \frac{\text{logit}(\tilde{p}) - \text{logit}(p) + ln(C/B) - a}{b}.$$

Coming back to the expression given in (3.4) for the "slope of isoutility" S, it requires that the users quantify in principle the consequences of each possible test result in order to allocate costs to the four different classification results, which is generally complex in practice. This allocation can be expressed in terms of financial or health costs and it can be seen from the perspective of patients, insurers, ... Besides, some judgment should be made about the relative costs of the different decisions. Cantor et al. (1999) reviewed

the medical literature on the use of cost ratio in ROC analysis, that is, how researchers assigned or determined a C/B ratio in the management of clinical problems in order to determine an optimal cutpoint to be used in practice, providing a rough guide on how to choose the corresponding weights/costs. These authors noted that most of the articles reviewed did not take into account either costs (the risks and benefits of over (and under) treatment) or disease prevalence, and even in some cases the threshold was chosen arbitrarily, without some justification or explanation given. In fact, of the 48 articles they reviewed, only 13 articles included a C/B ratio as part of the analysis based on the ROC curve for computing the optimal cutpoint. In these studies, the smallest C/B ratio found was $1/400 = 0.0025$ for tuberculosis screening (Lusted, 1971), because in this clinical setting, the consequences of a missed case are really serious (that may potentially result in death) and consequently, the costs of giving an overtreatment to a non-diseased patient are small relative to the benefits obtained after treating true disease patients; and the largest C/B ratio found was $260/96 = 2.7$ for teeth restoration necessitated by carious lesions (Kay and Knill-Jones, 1992).

However, the allocation of costs may not be as difficult as it may seems at first since the value of the ratio $[(C_{FP} - C_{TN})/(C_{FN} - C_{TP})]$ can often be determined without knowing the absolute values of the four costs involved. In fact, in only 2 of the studies reviewed (DeBaun and Sox, 1991; Hagen, 1995), the four costs of the four diagnostic outcomes (true positives, true negatives, false positives and false negatives) were identified. In the other 11 studies that considered a C/B ratio, in general, few details were provided in reference to how the C/B ratio was determined. As the number $C_{FP} - C_{TN}$ represents the loss that is incurred when a false positive decision is made rather than a true negative decision, and similarly, $C_{FN} - C_{TP}$ represents the loss involved when making a false negative decision rather than a true positive, it would only be necessary to estimate the ratio of these two losses, that is, their relative values with respect to each other. In the majority of these studies, the authors pointed out that "the costs of treatment of non-diseased patients were x times as great as the net benefits of treatment of diseased patients" and so in these cases, the C/B ratio was simple to calculate. For instance, in the study of Bergus (1993), the author states that the net benefits of treatment were 1.5 times as great as the net costs based on a holistic estimate, and therefore the C/B ratio was calculated as $1/1.5 = 0.667$. Moreover, in many of the reviewed studies (Hdez-Armas et al., 1982; England, 1988; Fombonne, 1991; Fujiyama et al., 1992), the authors pointed out that the net costs of treatment were equivalent to the net benefits and thus the C/B ratio was simply 1. In fact, without better information, the general tendency is to assume that the prevalence of disease is $p = 0.5$, $C_{FP} = C_{FN}$, and $C_{TP} = C_{TN}$ (or equivalently, $C_{FP} - C_{TN} = C_{FN} - C_{TP}$) and therefore a cutoff point is selected such that $S = 1$, that is, the cutpoint defined by the Youden index. However, it is important to stress that logically this cutpoint may not be optimal for other prevalence values and cost ratios and that the costs may change considerably between different clinical scenarios.

Lastly, it is also interesting to mention that the *decision theory (misclassification loss) criterion* (Kristjansson et al., 1996), defined for settings with two or more classes of interest, can be seen as a generalization of the cost-benefit ratio criterion. For the two-group case, the optimal cutpoint defined by the decision theory (misclassification loss) criterion is

given by

$$c = \arg\min_c \{(1-p)Sp(c)(C_{TN} - C_{FP}) + p(1 - Se(c))(C_{FN} - C_{TP}) + (1-p)C_{FP} + pC_{TP}\},$$

or, equivalently, by

$$c = \arg\min_c \{C_{FP}(1-p)(1-Sp(c)) + C_{FN}p(1-Se(c)) + C_{TP}p(1-1+Se(c)) + C_{TN}(1-p)Sp(c)\}.$$

So, it is clearly seen that it coincides with the solution given by the *cost-benefit ratio criterion*.

18. Particular cases of the cost-benefit ratio criterion

In the following we will show particular cases of the cost-benefit ratio criterion which are based on considering specific values for the costs and/or for the prevalence.

Minimization of misclassification cost term (MCT)
Minimum cost (Anderson's proposal)
Minimum loss function

Some authors (McNeill et al., 1975; Zweig and Campbell, 1993; Burgueño et al., 1995) only mention the ratio of a false positive against a false negative, because the costs of the correct decisions are assumed null and thus the slope of isoutility S can be simplified as follows:

$$S = \frac{1-p}{p}\frac{C_{FP}}{C_{FN}}.$$

The *minimization of misclassification cost term (MCT)*, or *minimum cost (Anderson's proposal)* (Berkson, 1947; Vizard et al., 1990; Smith, 1991; Greiner, 1995, 1996; Wang and Geisser, 2005; Caraguel et al., 2011), that defines the optimal cutpoint c as that which minimizes the MCT given by

$$MCT(c) = \frac{C_{FN}}{C_{FP}}p(1 - Se(c)) + (1-p)(1 - Sp(c))$$

can be seen as a particular case of the previous *cost-benefit criterion*, that considers that the costs of the correct classifications are null. Besides, the MCT term that measures the cost of the misclassifications coincides with the expected loss function for classifying a patient (Geisser, 1998; Greiner et al., 2000; Schisterman et al., 2005):

$$(1-p)(1 - Sp(c)) + ap(1 - Se(c)), \tag{3.6}$$

where $a = C_{FN}/C_{FP}$. So, the MCT based criterion can be also called the *minimum loss function criterion*.

Maximization of the Skill Plot

The previous particular cases of the cost-benefit ratio criterion are also equivalent to the criterion based on the ***Skill Plot*** $K_\theta(x)$ (Briggs and Zaretzki, 2008) defined by

$$K_\theta(x) = \begin{cases} \dfrac{p(1-\theta) - C(x)}{p(1-\theta)} = 1 - \dfrac{C(x)}{p(1-\theta)}, & \text{if } p < \theta, \\ \dfrac{\theta(1-p) - C(x)}{\theta(1-p)} = 1 - \dfrac{C(x)}{\theta(1-p)}, & \text{if } p \geq \theta, \end{cases}$$

which is simply a scaled version of the expected cost

$$C(c) = p(1 - Se(c))(1-\theta) + (1-p)(1 - Sp(c))\theta.$$

Note that when the parameter θ represents the relative cost or loss of a false positive decision, that is,

$$\theta = \frac{C_{FP}}{C_{FP} + C_{FN}},$$

and $1 - \theta$ represents the relative loss of a false negative, that is,

$$1 - \theta = \frac{C_{FN}}{C_{FP} + C_{FN}},$$

it follows that $K_\theta(x)$ can be rewritten as follows

$$K_\theta(x) = \begin{cases} \dfrac{p(1-\theta)Se - (1-p)\theta(1-Sp)}{p(1-\theta)} = Se - \dfrac{1-p}{p}\dfrac{\theta}{1-\theta}(1-Sp), & \text{if } p < \theta, \\ \dfrac{\theta(1-p)Sp - (1-\theta)p(1-Se)}{\theta(1-p)} = Sp - \dfrac{p}{1-p}\dfrac{1-\theta}{\theta}(1-Se), & \text{if } p \geq \theta, \end{cases}$$

Taking into account that the following optimization problems are equivalent:

$$\min\{p(1-\theta)Se - (1-p)\theta(1-Sp)\},$$

$$\min\{p(1-\theta)(1-Se) + (1-p)\theta(1-Sp)\},$$

$$\max\{\theta(1-p)Sp - (1-\theta)p(1-Se)\},$$

$$\max\{(1-p)\theta Sp + p(1-\theta)Se - (1-p)\theta\},$$

$$\min\{p(1-\theta)(1-Se) + (1-p)\theta(1-Sp)\},$$

it is easy to check that in both cases (whether $p < \theta$ or $p \geq \theta$), the Skill Plot defines the same cutpoint as the MCT method. Additionally, from the alternative expression

$$K_\theta(x) = \frac{p(1-\theta)Se - (1-p)\theta(1-Sp)}{p(1-\theta)} I_p + \frac{\theta(1-p)Sp - (1-\theta)p(1-Se)}{\theta(1-p)}(1 - I_p),$$

where $I_p = I(p < \theta)$, it is also obvious that maximizing the Skill Plot is exactly the optimal Bayes classification boundary.

The advantage of this method is that it provides an easy-to-interpret alternative to the ROC curve. Besides, the Skill Plot allows the analyst to immediately judge the quality

of a diagnostic test or forecast based on a particular cutoff value and to assess the range of useful cutoffs. However, when compared to the ROC curve, the disadvantages are: 1) the inability to compare multiple testing rules based on different diagnostic tests on the same graph (as each diagnostic test would have different x-axis units), and 2) the lack of an AUC-like overall measure.

Generalized Youden index

It is easy to see that minimizing the expected loss function over all possible cutoffs (3.6) is equivalent to maximize:

$$Se(c) + rSp(c) - 1$$

with $r = (1-p)/(ap) = ((1-p)C_{FP})/(pC_{FN})$, which is the optimization problem defined by the *Generalized Youden index* (Geisser, 1998; Greiner et al., 2000; Schisterman et al., 2005; McClish, 2012; Zou et al., 2013). Besides, for the particular case that $r = 1$, it coincides with the traditional Youden index. As we saw in the previous section, the traditional Youden index is one of the best known and used criterion in clinical practice for selecting the optimal cutpoint. However, an important limitation is that assigns the same weight to sensitivity and specificity, and sometimes different weights (based maybe on the cost of the different types of classification errors and the prevalence of disease) are adequate.

Maximization of Validity index (or efficiency)
Minimization of error rate

The Validity index (VI) or efficiency is defined as the proportion of individuals correctly classified ("Well Classified Frequency", WCF, or "accuracy"). Feinstein, SH (1975) demonstrated how this index depends, not only on measures of sensitivity and specificity, but also on the disease prevalence. Indeed it is given by:

$$\begin{aligned} VI(c) &= \frac{\#\text{of diseased individuals} \cdot Se(c) + \#\text{of healthy individuals} \cdot Sp(c)}{n} \\ &= pSe(c) + (1-p)Sp(c) = p(Se(c) - Sp(c)) + Sp(c). \end{aligned}$$

The geometric interpretation of this index is that it represents the equation of a straight line with intercept the specificity and slope equal to the difference between sensitivity and specificity.

The optimal cutpoint that *maximizes the Validity index* (Galen, 1986; Vermont et al., 1991; Greiner et al., 1995; Björk et al., 1996; Jung et al., 1996; Toubert et al., 1996; Woodward, 1999; Hoffman et al., 2000; Liu et al., 2005; McLaughlin et al., 2005; Chen et al., 2006; Smits et al., 2007; Freeman and Moisen, 2008; Caraguel et al., 2011; Bartlett et al., 2012; Zou et al., 2013), it is equivalent to the cost-benefit ratio criterion but considering the cost-benefit ratio equal to $C/B = 1$.
When $C/B = 1$, it follows that $C_{FP} - C_{TN} = C_{FN} - C_{TP}$, and the following optimization problems are equivalent:

$$\max\{pSe + (1-p)Sp\}$$

$$\max\{Se + (1-p)/pSp\}$$

$$\max\{Se + \frac{1-p}{p}\frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}Sp - 1\}$$

$$\max\{Se - \frac{1-p}{p}\frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}(1 - Sp)\}.$$

Therefore, the slope of the ROC curve at the optimal cutpoint obtained by this criterion is equal to $S = \frac{1-p}{p}$. Benowitz et al. (2009) and Sabatine et al. (2009) used this criterion for selecting the optimal cutpoint and since there is no a closed expression for computing the standard error of the optimal cutpoint, they used the bootstrap technique to validate the cutpoint and to generate 95% confidence intervals for both the cutpoint and its corresponding sensitivity and specificity indexes for several demographic groups (Rust and Rao, 1996).

The main disadvantage of this criterion is the fact that Se and Sp are weighted only by the disease prevalence and that the two incorrect decisions are considered equally hazardous. These assumptions are often inappropriate in practical applications (Brown and Davis, 2006).

It should be noted that maximizing the Validity index is also equivalent to minimize the proportion of incorrect diagnostic classifications (error rate):

$$\begin{aligned} ER(c) &= \frac{\#\text{of diseased individuals} \cdot (1 - Se(c))}{n} \\ &\quad + \frac{\#\text{of healthy individuals} \cdot (1 - Sp(c))}{n} \\ &= p(1 - Se(c)) + (1-p)(1 - Sp(c)) \\ &= p[(1 - Se(c)) - (1 - Sp(c))] + (1 - Sp(c)), \end{aligned}$$

that is, the so called *minimum error rate criterion* (Cummings and Richard, 1988). So, the method that maximizes efficiency or accuracy provides the same optimal cutpoint as the method that minimizes the classification error rate (Metz, 1978). Additionally, this criterion yields the same cutpoint as the Youden index if the disease prevalence is equal to 0.5.

19. Maximization of expected utility

In the following, we will present an approach based on the expected utility that is equivalent to *the minimum average cost criterion*, previously discussed in 17, *the maximization of expected utility criterion*, where utility is a measure of the strength of preference for an specific outcome (Cantor et al., 1999). In the literature, there exist two perspectives for evaluating the ratio of net costs to net benefits, depending on whether one wants to view outcomes as either "costs" (in a negative frame) or as "utilities" (in a positive frame). Based on the first perspective, we have seen so far that "cost" is a negative outcome measure, referring to the negative effects in terms of, for instance, monetary cost, adverse health risks, or a combination of both. From the second perspective, outcomes are seen in terms of utilities that refer, for example, to monetary savings, health benefits,

or a combination of the two. In essence, utility is a measure of the importance of the state of health for a given individual, such as life expectancy or quality of life. The paradigm of using utilities for evaluating the ratio of net costs to net benefits was developed by Krieg et al. (1986); Sox et al. (1988) and DeNeef and Kent (1993).

The expected utility of a classification threshold c, $E(U(c))$, for a randomly selected person is equal to the sum of the values of the four possible diagnostic decisions (true positive, false negative, false positive and true negative) weighted (that is, multiplied) each by its corresponding probability of occurrence (Krieg et al., 1986; Smits et al., 2007; Baker, 2009; Irwin and Irwin, 2011), that is,

$$\begin{aligned} E(U(c)) &= \Pr(D=1)\Pr(Y \geq c|D=1)U_{TP} + \Pr(D=0)\Pr(Y \geq c|D=0)U_{FP} \\ &\quad + \Pr(D=1)\Pr(Y < c|D=1)U_{FN} + \Pr(D=0)\Pr(Y < c|D=0)U_{TN} \\ &\quad + U_{test}, \end{aligned} \tag{3.7}$$

where U_{TP}, U_{TN}, U_{FP}, U_{FN} represent the mean utilities owing from each type of diagnostic decision to assess the state of health of a diseased-treated one, a healthy untreated one, a healthy subject mistakenly treated, and a disease-untreated one, respectively; and U_{test} is the utility (monetary cost or harm) from a test (it is often considered null). Similarly to costs, these utilities weigh each diagnostic decision with respect to its clinical relevance. Specifically, U_{TP}, the utility of a true positive result, represents in general an appropriate treatment (disease is present or will develop); and U_{TN}, the utility of a true negative misclassification, is typically the best outcome of not having disease. The utility of a false positive decision U_{FP} represents often an inappropriate treatment (because disease is absent or will not develop), and the utility of a false negative result U_{FN} is typically the worst outcome of missed case (no treatment is given but disease is present or will develop). So, in the majority of cases, these utilities are ordered in this way: $U_{FN} < U_{TP} < U_{FP} < U_{TN}$ (Jund et al., 2005).

Obviously, the optimal cutpoint defined by *the maximization of expected utility approach* is given by the threshold value c that has the highest expected utility.

The expected utility given in (3.7) can be also expressed in terms of the ROC method (Metz, 1978; Kraemer, 1992), rewriting the above conditional probabilities in terms of the sensitivity and specificity measures, as follows:

$$\begin{aligned} E(U(c)) &= Se(c)pU_{TP} + (1-Sp)(c)(1-p)U_{FP} \\ &\quad + (1-Se)(c)pU_{FN} + Sp(c)(1-p)U_{TN} + U_{test}. \end{aligned} \tag{3.8}$$

Note that the expression given in (3.8) for the expected utility is similar to that derived for the total cost given in (3.2), replacing here the costs of the different diagnostic classifications by the corresponding utilities (or profits). Due to the equivalence, but in an opposite direction, of these two perspectives to measure either cost or utility, there is a direct relationship between the values of the costs and utilities as follows:

$$C_{FP} = -U_{FP}, C_{FN} = -U_{FN}, C_{TP} = -U_{TP}, C_{TN} = -U_{TN}.$$

Consequently, the following optimization problems are equivalent:

$$\max\{Se(c)pU_{TP} + (1-Sp)(c)(1-p)U_{FP} + (1-Se)(c)pU_{FN} + Sp(c)(1-p)U_{TN}\},$$

$$\min\{C_{TP}Se(c)p + C_{FP}(1-Sp)(c)(1-p) + C_{FN}(1-Se)(c)p + C_{TN}Sp(c)(1-p)\}.$$

And thus, both approaches, *the maximization of expected utility approach* and *the minimum average cost criterion*, define the same classification threshold.

When considering, for instance, $U_{FP} = U_{FN} = 0$ and $U_{TP} = U_{TN} = 1$, (that is, the utilities of the incorrect decisions are null and the utilities of the correct decisions are equal to 1), maximizing (3.8) would be equivalent to maximizing efficiency or accuracy, which we had already treated as a particular case of the criterion that minimizes the cost.

We include here an illustrative example taken from the literature to clarify the meaning of utilities and their relationship with the costs. It should be noted that utilities can be negative if they are detrimental (Baker, 2009). For instance, in the context of a questionnaire to estimate the risk of colorectal cancer, Gail and Pfeiffer (2005) considered the following values for the utilities : U_{FN} = -100 for the possibility/risk of death and morbidity due to failing to detect colorectal cancer, that is, the cost of missing cancer cases is set equal to C_{FN} = 100 and it is a high cost because cancer is a very serious disease; U_{FP} = -1 for the risk of bleeding or perforation of the colon, because a non-diseased patient is suffering an unnecessary risk since treatment always has a risk and therefore it has a cost C_{FP} = 1, but anyway this cost of giving treatment to a non-diseased patient is much lower than the cost of a false negative decision (not detecting a patient with cancer); U_{TP} = -11 for the risk of bleeding or perforation of the colon and the lowered chance of death or morbidity from colorectal cancer due to early detection, that is, because a diseased patient also suffer the risk of treatment but this is better than a false negative because in this case the death could be avoided and so C_{TP} = 11 is lower than C_{FN} but greater than C_{FP} because the patient has the disease; and U_{TN} = 0 as a reference value because a healthy patient does not suffer any risk and therefore the corresponding cost C_{TN} is null.

The expected utility given in (3.7)-(3.8) can be also expressed in terms of predictive values (Greenland, 2008; Baker, 2009), that is,

$$\begin{aligned} E(U(c)) &= \Pr(Y \geq c)PPV(c)U_{TP} + \Pr(Y \geq c)(1 - NPV(c))U_{FP} \\ &\quad +(1 - \Pr(Y \geq c))(1 - PPV(c))U_{FN} + (1 - \Pr(Y \geq c))NPV(c)U_{TN} \\ &\quad +U_{test}, \end{aligned}$$

which is obtained from (3.8) by taking into account that:

$$\begin{aligned} Se(c)p &= \Pr(Y \geq c)PPV(c), \\ (1 - Sp(c))(1-p) &= \Pr(Y \geq c)(1 - NPV(c)), \\ (1 - Se(c))p &= (1 - \Pr(Y \geq c))(1 - PPV(c)), \\ Sp(c)(1-p) &= (1 - \Pr(Y \geq c))NPV(c). \end{aligned}$$

It is interesting to mention that the same expression for the expected utility given in (3.8) has been considered by several authors, for instance, Krieg et al. (1986); Smits et al. (2007); Baker (2009); Irwin and Irwin (2011) and Martínez-Camblor (2011), although such a expression has been developed in different, although equivalents, ways by the different authors. Besides, Smits et al. (2007) rearranged terms in (3.8) as follows:

$$E[U(c)] = (1-p)U_{FP} + pU_{FN} + (1-p)Sp(c)(U_{TN} - U_{FP}) + pSe(c)(U_{TP} - U_{FN}),$$

where the first two terms are fixed for all cutoff points c, $U_{TN} - U_{FP}$ reflects how much difference there is in utility whether healthy persons are correctly classified or not, and $U_{TP} - U_{FN}$ how much difference there is whether diseased persons are correctly classified or not. In the terminology used by Baker (2009), $L = U_{TN} - U_{FP}$ is the *loss* for a positive prediction among those without disease and $P = U_{TP} - U_{FN}$ is the *profit* for a positive prediction among those with disease. This terminology, however, had been previously introduced by Peirce (1884), who defined the profit P as the difference in utilities from making a positive prediction instead of a negative prediction (of disease) among patients with disease, and, in a similar way, the loss L was defined as the negative of the difference in utilities from making a positive prediction instead of a negative one but in patients without disease (Baker, 2009). According to DeNeef and Kent (1993), the net benefit P of treating a diseased individual depends on the treatment efficacy, the net cost L of treating a healthy individual depends on the treatment adverse effects, and the ratio of P over L may be considered as the number of healthy individuals a physician would accept to treat mistakenly to not have a diseased individual without treatment. Again, this is equivalent to the C/B term previously discussed for *the cost-benefit ratio criterion*. Considering the aforementioned biomedical example of Gail and Pfeiffer (2005), previously introduced to clarify the relationship between utilities and costs, the profit P is $P = -11 - (-100) = 89$, the loss is $L = 0 - (-1) = 1$ and the ratio $P/L = C/B = 89/1 = 89$.

Similarly to the *cost-benefit ratio criterion*, it is easy to prove that *the maximum expected utility criterion* defines as optimal cutpoint the operating point on the ROC curve where the slope of the ROC curve (equivalently, the likelihood ratio function) is given by (McNeill et al., 1975)

$$LR(c) = \frac{(1-p)(U_{TN} - U_{FP})}{p(U_{TP} - U_{FN})}, \tag{3.9}$$

which is indeed the fundamental formula of Peterson et al. (1954).

Considering the terminology introduced by all these authors, *the maximum expected utility criterion* is also known in the literature as *the highest net utility approach* (Krieg et al., 1986; Irwin and Irwin, 2011), or *the indifference curve approach*. This last terminology is closely related to the expression given above in (3.9) for the slope of the ROC curve at the maximum utility threshold. In fact, indifference curves are straight lines which are tangent to the ROC curves and their slopes equal the slope of the ROC curve at the point that yields the maximum utility.

20. Criterion based on the relative utility (RU) curve

Baker (2009) considered a new function of expected utilities, the so called *relative utility* (RU), which is basically defined as the maximum fraction of the expected utility achieved by risk prediction as compared with perfect prediction.

On one hand, the expected utility of perfect prediction is given by

$$U_{perfect} = pP + pU_{FN} + (1-p)U_{TN}.$$

This expression is simply obtained by considering that $Se(c) = 1$, $1 - Sp(c) = 0$ and $U_{test} = 0$ into the expression for the expected utility given in (3.8). Similarly, we can also define the expected utility of treating none (that is, all individuals are classified as negative),

$$U_{none} = pU_{FN} + (1-p)U_{TN},$$

and the expected utility of treating all individuals (that is, all are classified as positive),

$$U_{all} = pU_{TP} + (1-p)U_{FP}.$$

From these, the expected utility of not prediction is equal to the maximum of U_{none} and U_{all}. In the absence of prediction if treatment is given then $U_{all} > U_{none}$, and if not then $U_{all} \leq U_{none}$.

On the other hand, let R be the so called *risk threshold*, a summary measure of harms and benefits, defined in terms of the loss L and the profit P :

$$R = \frac{L}{L+P} = \frac{1}{1+P/L},$$

where P/L can be understood as the number of false positives that a person would trade for each true positive. In terms of this risk threshold, when $U_{all} > U_{none}$, the relevant region is given by $R < p$ and the slope of the ROC curve is lower than 1, that is, $LR < 1$. Conversely, when $U_{all} \leq U_{none}$, the relevant region is $R \geq p$ and the slope of the ROC curve is higher than or equal to 1, that is, $LR \geq 1$. Based on these relevant regions, the relative utility RU is defined as the ratio of the maximum utility of prediction (versus no prediction) to the utility of perfect prediction (versus no prediction), that is:

$$RU(R) = \begin{cases} \dfrac{U_{c(R)} - U_{all}}{U_{perfect} - U_{all}}, & \text{if } R < p, \\ \dfrac{U_{c(R)} - U_{all}}{U_{perfect} - U_{all}}, & \text{if } R \geq p, \end{cases}$$

where $c(R)$ is the smallest value c such that the probability of disease at the cutpoint c is $R = \Pr(D = 1|Y \geq c)$, that is,

$$c(R) = \inf\{c : R = \Pr(D = 1|Y \geq c)\},$$

and $U_{c(R)}$ denotes the expected utility associated to the risk prediction based on classifying as disease an individual with test value Y larger than or equal to $c(R)$ and as healthy otherwise. Equivalently, $RU(R)$ can be rewritten as follows

$$RU(R) = \begin{cases} Sp - \dfrac{1-R}{R}\dfrac{p}{1-p}(1-Se) - \dfrac{1-R}{R} - \dfrac{C}{1-p} \\ \quad = Sp - \dfrac{U_{TP} - U_{FN}}{U_{TN} - U_{FP}}\dfrac{p}{1-p}(1-Se) - \dfrac{1-R}{R} - \dfrac{C}{1-p}, & \text{if } R < p, \\ Sp - \dfrac{1-R}{R}\dfrac{p}{1-p}(1-Se) - \dfrac{1-R}{R} - \dfrac{C}{1-p} \\ \quad = Sp - \dfrac{U_{TP} - U_{FN}}{U_{TN} - U_{FP}}\dfrac{p}{1-p}(1-Se) - \dfrac{1-R}{R} - \dfrac{C}{1-p}, & \text{if } R \geq p, \end{cases}$$

where $C = -U_{test}/P$ is the cost of the test per unit profit. Besides, similarly to the expected utility, RU can also be expressed using the predictive values, see Baker (2009).

Therefore, *the criterion based on the relative utility curve* defines as optimal the threshold value $c(R)$ that maximizes the relative utility $RU(R)$. Relative utility curves can be seen as an improvement of the decision curves (Vickers and Elkin, 2006) because there is no need to set any utility equal to a reference value. A relative utility curve is simply a plot of the relative utility $RU(R)$ versus the risk threshold R and it shows information on how much "the risk prediction" contributes to clinical utility relative to "perfect prediction" (Baker, 2009).

It is easy to prove that either if $R \geq p$ or $R < p$, the maximization of $RU(R)$ yields the same solution as *the maximum expected utility criterion* and, consequently, also the same as *the cost-benefit ratio criterion*. Besides, it should be noted that maximizing the relative utility curve is equivalent to the maximization of the Skill Plot when considering the utilities of correct classifications null.

21. F-measure (or F approach) criterion

The optimal cutpoint c defined by this criterion maximizes the F-measure (Liu et al., 2005), given by

$$F_\alpha = \frac{1}{\frac{\alpha}{PPV(c)} + \frac{1-\alpha}{Se(c)}} = \frac{TP(c)}{TP(c) + (1-\alpha)FN(c) + \alpha FP(c)},$$

that is,

$$c = \arg\max_c \left\{ \frac{1}{\frac{\alpha}{PPV(c)} + \frac{1-\alpha}{Se(c)}} \right\} = \arg\max_c \left\{ \frac{TP(c)}{TP(c) + (1-\alpha)FN(c) + \alpha FP(c)} \right\},$$

with $0 \leq \alpha \leq 1$, a user defined parameter to penalize the two types of errors (false negatives and false positives). Note that the F-measure combines both recall (true positive rate or sensitivity) and precision (positive predictive value) into a single utility function.

3.4 Practical considerations

We reviewed here numerous techniques or methods for selecting optimal cutoff points for quantitative diagnostic tests. From a practical point of view, it is not always clear when one should use this method or the other. Each has its own benefits. Besides, the superiority or equivalence of these approaches depends on the specific clinical scenario (Wheeler et al., 2007).

As we have previously mentioned, the selection of the appropriate optimal criterion should depend on the ultimate goal of the diagnostic test, taking into account mainly the

prevalence of the disease and the severity of false positives and false negatives misclassifications. In Table 3.1 we present the criteria that would be more appropriate in terms of these two aspects and the main characteristic of the corresponding optimal cutpoint.

Table 3.1: Most appropriate criteria in practice.

Prevalence	Severity of misclassifications	Criteria recommended	Cutoff resulting
High	FN $>$ FP	Maximizing Sp for a fixed Se Maximizing efficiency (or accuracy)	Very low
	FN $\approx$ FP	Maximizing efficiency (or accuracy)	Low
	FN $<$ FP	Maximizing Se for a fixed Sp	Medium-high
Intermediate	FN $>$ FP	Maximizing Sp for a fixed Se Maximizing efficiency (or accuracy)	Low
	FN $\approx$ FP	A weighted combination of Se and Sp Maximizing efficiency (or accuracy)	Intermediate
	FN $<$ FP	Maximizing Se for a fixed Sp Maximizing efficiency (or accuracy)	High
Low	FN $>$ FP	Maximizing Sp for a fixed Se	Low-medium
	FN $\approx$ FP	Maximizing efficiency (or accuracy)	High
	FN $<$ FP	Maximizing Se for a fixed Sp Maximizing efficiency (or accuracy)	Very high

When a value for the specificity is specified in advance, we pursue to ensure a higher confidence in the prediction of a negative result (that is, the absence of disease). Similarly, when a sensitivity value is set a priori, we pursue to guarantee a higher confidence in the prediction of a positive result (that is, presence of disease). When the consequences of an incorrect classification in either direction are approximately equal, it is appropriate to maximize efficiency, especially in the case of a disease with an intermediate prevalence (neither too high nor too low), or to simultaneous maximize both sensitivity and specificity (that is, the Symmetry point).

In situations where the false negative decisions are more serious than the false positive ones, maximizing specificity for a fixed high sensitivity is a recommended method for selecting the optimal threshold. This occurs when not diagnosing the disease can be fatal for the diseased individuals, such as for serious but treatable diseases, as tuberculosis or lymphoma, or diseases in which a false positive does not cause serious psychological or economic problems for the patient (for example, performing a mammography in breast cancer). Moreover, in the case of high or intermediate prevalence, it is also convenient to maximize the efficiency (or validity index). If instead the consequences of a false positive are more serious than the consequences of a false negative result, for example, in some diseases which are incurable or rapidly lethal, you can select the optimal cutoff based on maximizing the sensitivity for a fixed value of the specificity, and in the event that the prevalence of the disease under study is intermediate or low, the method of maximum efficiency or proportion of correctly classified is also an appropriate choice. Briefly, we can say that a higher accuracy or efficiency is desirable when the disease is important but treatable, and both false positives and false negatives involve trauma and entail serious consequences.

The method based on costs and benefits incorporates the costs and prevalence into

the estimation of each of the above methods. The main issue is, therefore, thinking about the implications for a false positive diagnosis and a false negative diagnosis. If the impacts are similar, the method of maximum efficiency or simultaneously maximization of sensitivity and specificity (Symmetry point) should be used. When estimates of the cost ratio and the prevalence are known, which is not always easy in practice, the method based on costs (and benefits) should be used.

For instance, in the case of coronary artery disease (CAD), if one wished to use leukocyte elastase determination as a screening test prior to performing a coronary angiography for detecting CAD, one would seek high sensitivity so as to be able to identify all the diseased patients. Thus, using the criterion that sets a minimum value for sensitivity with $Se \geq 0.95$, the cutpoint obtained would be 22 μgl^{-1}, and so coronary angiography would be performed on any patient having a leukocyte elastase level ≥ 22 μgl^{-1}. Using this optimal cutpoint, 96% of CAD patients would be correctly classified, whereas only 38% of patients without CAD would be correctly identified (28 false positive classifications). If, however, one were seeking an equilibrium between sensitivity and specificity, an optimal cutpoint of 38 would be obtained by using the Symmetry point, and 68% of the patients with CAD and 67% of the patients without CAD would be correctly detected. This value is very close to that obtained using the Youden index, which would afford a value of 37, similarly yielding an equilibrium between the two measures of sensitivity and specificity (Se = 0.69 and Sp = 0.67). However, if we select the highest possible value of specificity, for instance, $Sp \geq 0.95$, a value of 54 would be obtained, a value that, as will be readily appreciated, is very much higher than the previous ones. In our example, however, this approach would not be appropriate, in view of the fact that angioplasty, with or without a stent, is usually a successful treatment in CAD.

Greiner et al. (2000) suggested that the method based on minimizing the misclassification term (MCT) and the logistic regression model based on the likelihood ratio are useful methods for selecting the cutoff point. Nevertheless, further studies are required to investigate the behavior of both criteria under different distributional assumptions of the diagnostic test.

Coming back to the CAD example, if one assumed that, despite being a severe disease, coronary stenosis is usually treated successfully with minimal risk for the patient, clinically speaking it would make more sense to consider that the false negative results have higher "cost" than the false positive results. If the case that a FN had three times the cost of a FP, the MCT method with a ratio of C_{FN}/C_{FP} = 3 could be used. One would thus obtain an optimal value of 21, so that patients with elastase higher than or equal to 21 would be classified as patients with CAD, so minimizing the false negative classifications.

If we consider, instead, criteria based on predictive values, from an applied point of view, it is usual to seek elevated positive predictive values in any case where treating false positives may have serious consequences, be these psychological, physical or economic (for example, chemotherapy in cancer or AIDS). Taking the CAD example, in view of the fact that 1) coronary disease is potentially curable (there is a treatment), 2) a false positive does not produce serious disorders for the patient, and 3) coronary angiography enjoys good results with low risk, one would seek elevated negative predictive values. This is related to the ability to rule out the disease with a greater degree of certainty. For the purpose, one could, for instance, use the method which sets a minimum value

for negative predictive value. Thus, for a $NPV \geq 0.95$, the cutpoint obtained would be 13, with a PPV = 0.72 and an NPV = 1 (the maximum value). This means that all patients with elastase below 13 are identified as patients without CAD and can be correctly classified as healthy (i.e., there are no false negative results).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Symmetry is what we see at a glance; based on the fact that there is no reason for any difference ...

- Blaise Pascal

Chapter 4

The Symmetry point and its cost-based generalization

In Chapter 3 we have presented several methods previously proposed in the literature for identifying optimal cutpoints in continuous diagnostic tests, in the sense of a specific optimality criterion such as the North-West corner (Metz, 1978; Vermont et al., 1991; Perkins and Schisterman, 2006), the Youden index (Youden, 1950; Aoki et al., 1997; Greiner et al., 2000), the Concordance probability (Liu, 2012; Lewis et al., 2008) and the Symmetry point (Greiner, 1995; Defreitas et al., 2004; Adlhoch et al., 2011), among others. Under the assumption that the marker in healthy and diseased populations Y_0 and Y_1 are normally distributed with the same standard deviation, these criteria yield the same operating point, $(1-p(c), q(c))$, on the ROC curve, that is, the same pair of specificity (p) and sensitivity (q) indexes, and consequently the same threshold value on the marker scale. However, this is not the case in general, as can be seen in Figure 4.1, where we have plotted the four different operating points obtained with the above-mentioned optimal criteria using a specific binormal example in which the marker Y follows a normal distribution with mean $\mu_0 = 6.5$ and standard deviation $\sigma_0 = 0.3$ in the healthy population and with mean $\mu_1 = 7.25$ and standard deviation $\sigma_1 = 0.5$ in the diseased population (this is the binormal model a) with $AUC = 0.90$ which will be considered in the simulation study shown in the next Subsection 4.4, see Table 4.1). Anyway, it is interesting to have in mind that any of the different threshold values derived from these optimal criteria can be interpreted similarly in the sense that they can be seen as two specific quantiles, the $p(c)$-th quantile of the healthy population and the $(1-q(c))$-th quantile of the diseased population. In this chapter we will focus on the Symmetry point and a generalization of it that takes into account the misclassification costs.

The Symmetry point, also known in the literature as the point of equivalence, see Greiner et al. (1995); Defreitas et al. (2004); Adlhoch et al. (2011), is the point c_S where $p(c_S) = q(c_S)$, that is, where the ROC curve and the line $y = 1 - x$ (the perpendicular to the positive diagonal line) intersect. The Symmetry point can also be seen as the point that maximizes simultaneously both types of correct classifications (Riddle and Stratford, 1999; Gallop et al., 2003), that is, it balances the two types of correct classifications and therefore it corresponds to the probability of correctly classifying any subject, whether it is healthy or diseased (Jiménez-Valverde, 2012, 2014). As Riddle and Stratford (1999) mention, from a clinical perspective, there may be situations where it is more appro-

priate to maximize simultaneously both types of correct classifications rather than the sum or the product of them, and consequently the Symmetry point should be preferred to the Youden index and the Concordance probability criteria. Recently, the Symmetry point has been recognized as an useful discrimination measure to select between several species distribution models (Jiménez-Valverde, 2012, 2014). In addition, the lower and upper half-AUCs, recently defined as alternative summary indexes to the classical AUC, see Bradley (2014), are given in terms of the Symmetry point. However, up to our knowledge, there are no theoretical studies concerning the inference of the Symmetry point. Based on these facts, this chapter will be devoted to propose several inference procedures for this specific optimal cutpoint, the Symmetry point, and its cost-based generalization that we call the Generalized Symmetry point.

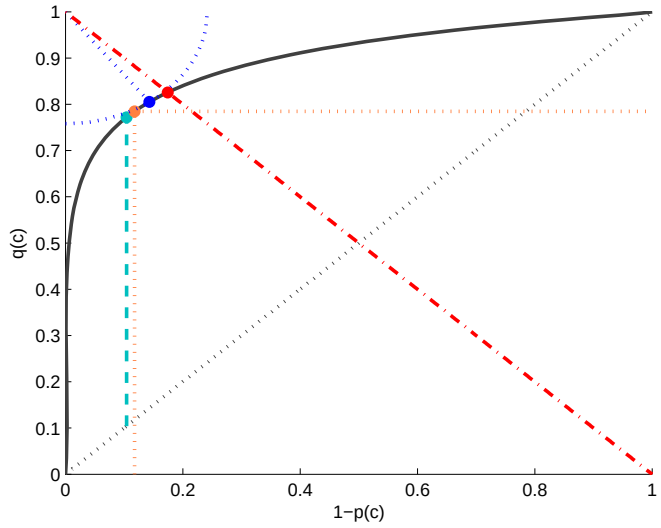


Figure 4.1: ROC curve associated to the binormal model a) with AUC = 0.90, and operating points, $(1-p(c), q(c))$, corresponding to the Symmetry point (S) (0.1743, 0.8257) (*red*), the North-West corner (NW) (0.1431, 0.8052) (*orange*), the Concordance probability (CP) (0.1177, 0.7847) (*blue*), and the Youden index (J) (0.1041, 0.7718) (*green*).

In the following, we first prove that the Symmetry point can be interpreted from a Bayesian decision theory perspective, and we introduce a generalization of it, the Generalized Symmetry point, that takes into account the misclassification costs. We then construct confidence intervals for the Generalized Symmetry point, c_{GS}, and its associated accuracy measures $p(c_{GS})$ and $q(c_{GS})$ using two approaches: the first one based on the **Generalized Pivotal Quantity** (GPQ) and the second one on **Empirical Likelihood** (EL). These two methods have been recently applied for constructing confidence intervals for the Youden index and its corresponding optimal threshold. For instance, in Lai et al. (2012), the GPQ method has been considered under the normality assumption, and in Molanes-López and Letón (2011), the nonparametric EL method has been used instead. Finally, we check the performance of the two methods through an extensive simulation study, and we analyze three real biomedical examples to illustrate the methodology pre-

viously discussed.

4.1 The Symmetry point

First of all, it is interesting to mention that the Symmetry point can be interpreted from a Bayesian decision theory perspective. Specifically, we will prove below that it coincides with the solution given by the minimax rule (Webb, 2002) that aims to minimize the maximum of the misclassification error probability when the classification or decision regions are specified in terms of a single cut-off point as it is usual in classical ROC analysis.

Proposition 4.1. The Symmetry point satisfies the following minimax condition

$$c_S = arg\min_c\{\max_\pi\{e(\pi, c)\}\}, \tag{4.1}$$

where $e(\pi, c)$ denotes the misclassification error probability given by

$$e(\pi, c) = (1-\pi)(1-F_0(c)) + \pi F_1(c) = (1-p(c)) + \pi \cdot ((1-q(c)) - (1-p(c))),$$

π denotes the disease prevalence and F_i is the cumulative distribution function (cdf) of the marker in the i-th population, for $i = 0, 1$.

Proof. On one hand, the Bayes decision rule for the minimum error (Webb, 2002) minimizes the misclassification error probability (also called the Bayes minimum error) given by:

$$\Pr(error) = (1-\pi)\int_{\Omega_1} f_0(y)dy + \pi\int_{\Omega_0} f_1(y)dy,$$

where $\Omega_0 = \{y : f_0(y)(1-\pi) > f_1(y)\pi\}$ and $\Omega_1 = \{y : f_0(y)(1-\pi) \leq f_1(y)\pi\}$ represent the classification regions based on which each individuals is assigned either to group 0 or group 1, respectively, and f_i denotes the density function of the marker in the i-th population, for $i = 0, 1$. This rule can lead to disjoint regions. However, in classical ROC analysis, the decision regions are given by $\Omega_0 = \{y : y < c\}$ and $\Omega_1 = \{y : y \geq c\}$, that is, defined in terms of a single cut-off point. So, introducing this restriction on the classification regions, $\Pr(error)$ can be rewritten as follows:

$$\begin{aligned} \Pr(error) &= (1-\pi)(1-F_0(c)) + \pi F_1(c) \\ &= (1-\pi)(1-p(c)) + \pi(1-q(c)) \\ &= (1-p(c)) + \pi \cdot ((1-q(c)) - (1-p(c))). \end{aligned} \tag{4.2}$$

From here on, to emphasize that $\Pr(error)$ depends on both parameters π and c, we will use the notation $e(\pi, c) = \Pr(error)$.

On the other hand, the minimax rule that aims to minimize the maximum of the misclassification error probability given in (4.2) solves the following optimization problem:

$$\min_c\{\max_\pi\{e(\pi, c)\}\} = \min_c\{\max\{1-p(c), 1-q(c)\}\} \tag{4.3}$$

where we have considered the fact that for a fixed c, the misclassification error probability, $e(\pi, c)$, is linear in the prevalence π, and therefore its maximum value is attained at

$\pi = 0$ or $\pi = 1$, depending on the slope sign. Consequently, the minimax solution of (4.3) is obtained by solving the following equation:

$$1 - p(c) = 1 - q(c) \Leftrightarrow p(c) = q(c)$$

which coincides with the Symmetry point, c_S. ■

It is important to emphasize here that when c is not fixed, the dependence of $e(\pi, c)$ in the prevalence π is not necessarily monotone because c also depends on π, and it is actually a more complex relationship (Webb, 2002). In fact, the Symmetry point yields the minimum error solution with respect to the least favourable prior distribution or, in other words, it maximizes the minimum error curve.

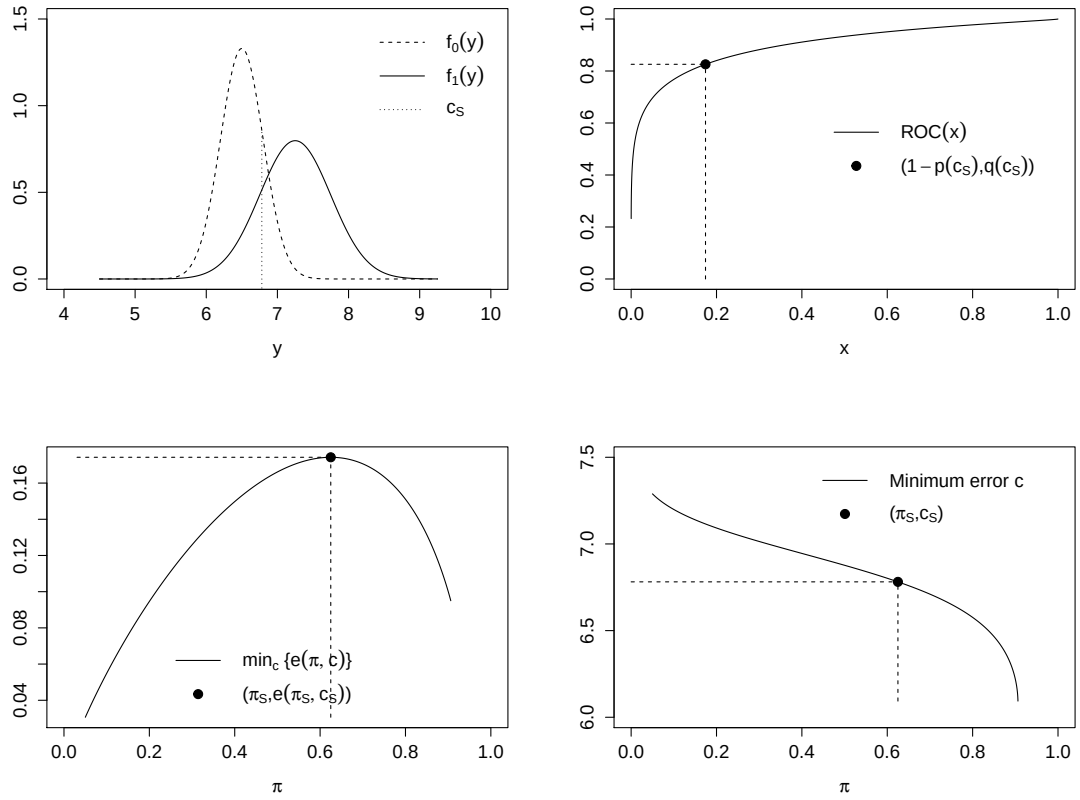


Figure 4.2: Binormal model a) with $AUC = 0.90$. *Top left panel:* the two population densities and the location of $c_S = 6.7816$. *Top right panel:* the corresponding ROC curve with the operating point associated to c_S, $(1 - p(c_S), q(c_S)) = (0.1743, 0.8257)$. *Bottom left panel:* the minimum error versus the prevalence π and the location of its maximum (the least favourable prior distribution $\pi_S = 0.6246$) for which the minimax error is $e(\pi_S, c_S) = 0.1743$. *Bottom right panel:* the minimum error threshold versus the prevalence and the location of (π_S, c_S).

In order to illustrate this fact, we have considered the same binormal example used previously for Figure 4.1. Figure 4.2 clearly illustrates that by plotting the minimum error probability as a function of the prevalence π, the maximum of this error provides the minimax solution. In the top left panel, the two population densities are shown and also a vertical line indicating the location of the Symmetry point or minimax cutpoint in the marker scale. In the top right panel, the corresponding ROC curve is shown, and also the operating point associated to the minimax cutpoint (that is, the operating point $(1 - p(c), q(c))$ where $p(c) = q(c)$). In the bottom left panel, the minimum error probability versus the prevalence is plotted and also a vertical line highlighting the location of its maximum (which provides the least favourable prior distribution π). Finally, in the bottom right panel, we show the minimum error threshold curve in terms of π, a horizontal line indicating the minimax cutpoint and a vertical line indicating the corresponding least favourable prior distribution. For this binormal example, we obtain in particular that the least favourable prior distribution is $\pi = 0.6246$, the Symmetry point or minimax cutpoint is $c_S = 6.7816$, the maximum of the minimum error curve or minimax error is $e(\pi, c_S) = 0.1743$, and the associated sensitivity and specificity indexes are $p(c_S) = q(c_S) = 0.8257$.

4.2 The Generalized Symmetry point

When selecting the optimal threshold, an important issue in order to assess the clinical effectiveness of a marker is to take into account the costs associated with the decisions taken (see, for instance, Remaley et al. (1999); Rutter and Miglioretti (2003)). For example, in a medical diagnosis problem in which a patient has back pain, it is far worse to classify a patient with severe abnormality as healthy than the other way around. In this case, when it is more important not to miss a diagnosis than misclassifying a healthy individual, a diagnostic test with high sensitivity is preferred over one with high specificity. Consequently, the cost associated to a false negative should be higher than the cost associated to a false positive. As we already mentioned in previous chapter, it may be very difficult to assign costs in practice. Specially, when the costs are a combination of several different factors measured in different units, such as money, time, and quality of life. In these situations, it may be crucial the subjective opinion of an expert.

Despite this practical difficulty, an interesting generalization of the Symmetry point (4.1) that incorporates the costs associated to false positives (c_{F+}), false negatives (c_{F-}), true positives (c_{T+}) and true negatives (c_{T-}) could be considered. In the following we prove that such a generalization is given by the minimax rule (Webb, 2002) that aims to minimize the maximum overall expected cost, loss or risk, when the classification or decision regions are specified in terms of a single cut-off point as it is usual in classical ROC analysis.

Proposition 4.2. The threshold value that solves the following minimax condition

$$\min_{c}\{\max_{\pi}\{\text{risk}(\pi, c)\}\}, \tag{4.4}$$

where risk(π, c) denotes the overall expected cost given by

$$\begin{aligned} \text{risk}(\pi, c) &= c_{T-} + (c_{F+} - c_{T-})(1 - p(c)) \\ &\quad +\pi \cdot ((c_{T+} - c_{T-}) + (c_{T-} - c_{F+})(1 - p(c)) + (c_{F-} - c_{T+})(1 - q(c))), \end{aligned} \tag{4.5}$$

yields a cost-based generalization of the Symmetry point, c_{GS}, that takes into account the misclassification costs.

Proof. First, we give below the specific details to obtain the formula of the overall expected cost given in (4.5). On one hand, the conditional cost, loss or risk of assigning an individual y to group 0 is defined as:

$$l^0(y) = c_{T-} \Pr(H|y) + c_{F-} \Pr(D|y),$$

and consequently, the average cost over region Ω_0 is:

$$r^0 = \int_{\Omega_0} l^0(y) f(y) dy$$

Using Bayes's theorem, it is straightforward to prove that

$$r^0 = \int_{\Omega_0} (c_{T-} f_0(y)(1 - \pi) + c_{F-} f_1(y)\pi) dy$$

On the other hand, the conditional cost of assigning an individual y to group 1 is defined as:

$$l^1(y) = c_{T+} \Pr(D|y) + c_{F+} \Pr(H|y),$$

and consequently, the average cost over region Ω_1 is:

$$r^1 = \int_{\Omega_1} l^1(y) f(y) dy$$

Analogously, using Bayes's theorem, it is straightforward to prove that

$$r^1 = \int_{\Omega_1} (c_{T+} f_1(y)\pi + c_{F+} f_0(y)(1 - \pi)) dy$$

Therefore, the overall expected cost is given by:

$$\begin{aligned} \text{risk}(\pi, c) &= r^0 + r^1 \\ &= \int_{\Omega_0} (c_{T-} f_0(y)(1 - \pi) + c_{F-} f_1(y)\pi) dy + \int_{\Omega_1} (c_{T+} f_1(y)\pi + c_{F+} f_0(y)(1 - \pi)) dy \end{aligned}$$

The Bayes decision rule for the minimum risk (Webb, 2002) minimizes this overall expected cost.

Under the restriction that $\Omega_0 = \{y : y < c\}$ and $\Omega_1 = \{y : y \geq c\}$, the overall expected cost is given by:

$$\begin{aligned} risk(\pi, c) &= c_{T-}p(c)(1-\pi) + c_{F-}(1-q(c))\pi \\ &\quad + c_{T+}q(c)\pi + c_{F+}(1-p(c))(1-\pi) \\ &= c_{T-}(1-1+p(c))(1-\pi) + c_{F-}(1-q(c))\pi \\ &\quad + c_{T+}(1-1+q(c))\pi + c_{F+}(1-p(c))(1-\pi) \\ &= c_{T-} - c_{T-}\pi - c_{T-}(1-p(c)) + c_{T-}(1-p(c))\pi + c_{F-}(1-q(c))\pi \\ &\quad + c_{T+}\pi - c_{T+}(1-q(c))\pi + c_{F+}(1-p(c)) - c_{F+}(1-p(c))\pi \\ &= c_{T-} + (c_{F+} - c_{T-})(1-p(c)) \\ &\quad + \pi \cdot (c_{T+} - c_{T-} + (c_{T-} - c_{F+})(1-p(c)) + (c_{F-} - c_{T+})(1-q(c))) \end{aligned}$$

Therefore, the minimax solution of (4.4) is obtained by solving the following equation:

$$\begin{aligned} risk(0, c) &= risk(1, c) \\ \Updownarrow & \\ c_{T-} + (c_{F+} - c_{T-})(1-p(c)) &= c_{T+} + (c_{F-} - c_{T+})(1-q(c)) \end{aligned}$$

From here on, we will refer to this risk-based minimax solution as the **Generalized Symmetry point** and, without loss of generality, we will assume that $c_{T+} = c_{T-} = 0$, that is, c_{GS} will be defined as the point that satisfies the following equation (López-Ratón et al., 2015a)

$$\rho(1-p(c_{GS})) = (1-q(c_{GS})), \tag{4.6}$$

where $\rho = \dfrac{c_{F+}}{c_{F-}}$ is the relative cost or loss of a false-positive classification as compared with a false-negative classification. ■

Similarly to the Symmetry point, now c_{GS} in (4.6) is graphically obtained by intersecting the ROC curve and the line $y = 1 - \rho x$. As ρ decreases, that is, as c_{F+} is smaller than c_{F-}, the operating point on the ROC curve associated to the Generalized Symmetry point c_{GS} moves from the bottom left side to the top right side, increasing the sensitivity at the cost of decreasing the specificity (equivalently, decreasing the false negative rate at the cost of increasing the false positive rate). When $\rho = 1$, the Generalized Symmetry point given in (4.6) yields the traditional Symmetry point, a particular case where the costs are not taken into account and consequently they are assumed to be equal, which may be far from reality in clinical practice.

In Figures 4.3 and 4.4, we illustrate the fact that the Generalized Symmetry point c_{GS} maximizes the Bayes minimum risk curve when $\Omega_0 = \{y : y < c\}$, and $\Omega_1 = \{y : y \geq c\}$, considering the same binormal example previously used for illustrating that the Symmetry point maximizes the Bayes minimum error curve (that is, the binormal model a) with $AUC = 0.90$ which will be used later on in the simulation study, see Table 4.1). Additionally, we have considered two different values of ρ, when the cost of a FP decision is four times higher than the cost of a FN (ρ=4) and the opposite, when the cost

associated to a FN misclassification is four times than the cost associated to a FP (that is, ρ=0.25). By graphing the Bayes minimum risk (that is, the overall expected cost $risk(\pi, c)$ corresponding to the threshold value c given by the Bayes rule for the minimum risk) as a function of the prevalence π, the maximum of this risk provides the minimax solution (see Figures 4.3 and 4.4).

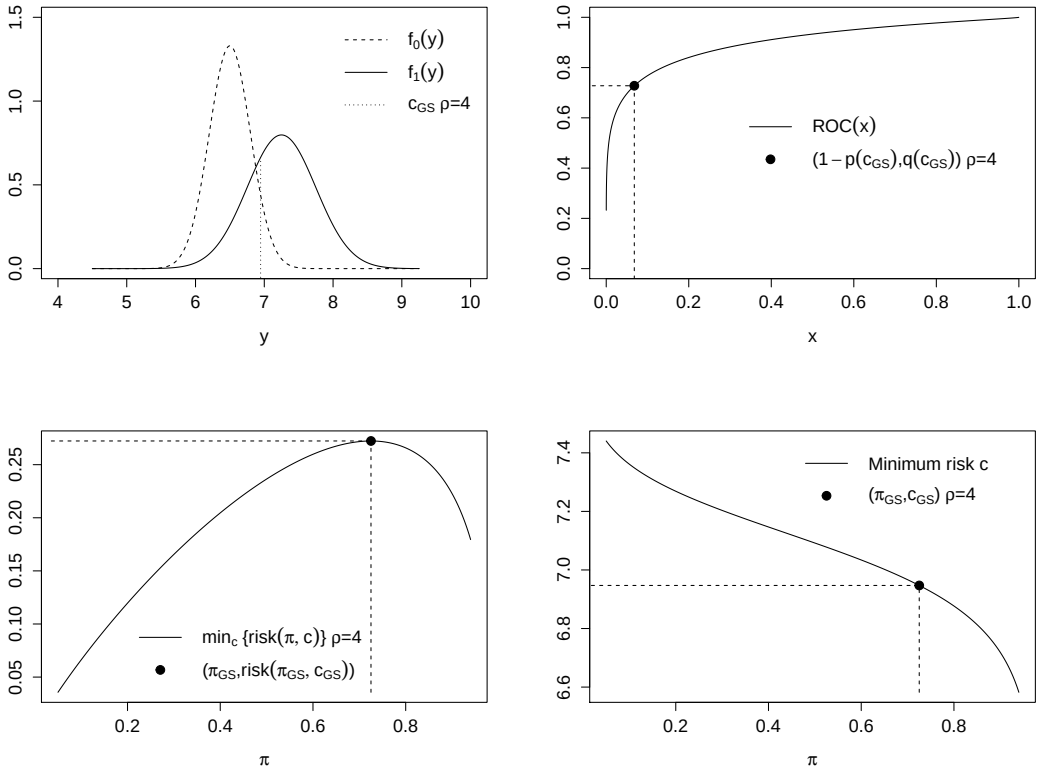


Figure 4.3: Binormal model a) with $AUC = 0.90$. *Top left panel:* the two population densities and the location of $c_{GS} = 6.9471$ for $\rho = 4$. *Top right panel:* the corresponding ROC curve with the operating point associated to c_{GS}, $(1 - p(c_{GS}), q(c_{GS})) = (0.0681, 0.7277)$ for $\rho = 4$. *Bottom left panel:* the minimum risk versus the prevalence π and the location of its maximum (the least favourable prior distribution $\pi_{GS} = 0.7251$ for $\rho = 4$) for which the minimax risk is $risk(\pi_{GS}, c_{GS}) = 0.2723$. *Bottom right panel:* the minimum risk threshold versus the prevalence and the location of (π_{GS}, c_{GS}).

In the top left panel of the Figures, the two healthy and diseased population densities are shown and also two vertical lines indicating the location of the Generalized Symmetry point or minimax cutpoint in the marker scale for the two values of ρ considered, respectively. We can see when increasing ρ the Generalized Symmetry point moves to the left. In the top right panel, the corresponding ROC curve is shown, and also the operating point $(1 - p(c_{GS}), q(c_{GS}))$ corresponding to the minimax cutpoint when ρ = 4 and ρ =

0.25, respectively. In the bottom left panel, the Bayes minimum risk versus the prevalence π is plotted for both values of ρ and also for each plot, a vertical line highlighting the location of its maximum (which provides the least favourable prior distribution π). Finally, in the bottom right panel, we show the Bayes rule for the minimum risk for ρ= 4 and for ρ=0.25, and also for each graph, a horizontal line indicating the minimax cutpoint and a vertical line indicating the corresponding least favourable prior distribution. In this case, for ρ= 4 the least favourable prior distribution is $\pi = 0.7251$, the maximum of the Bayes minimum risk or minimax risk is $risk(\pi, c_{GS}) = 0.2723$, and the Generalized Symmetry point or minimax cutpoint $c_{GS} = 6.9471$, with their associated accuracy measures $p(c_{GS}) = 0.9319$, and $q(c_{GS}) = 0.7277$. If we consider, instead, a value ρ= 0.25 we obtain a lower minimax cutpoint $c_{GS} = 6.5933$, with a lower prevalence $\pi = 0.4847$, and a higher minimax risk $risk(\pi, c_{GS}) = 0.3780$.

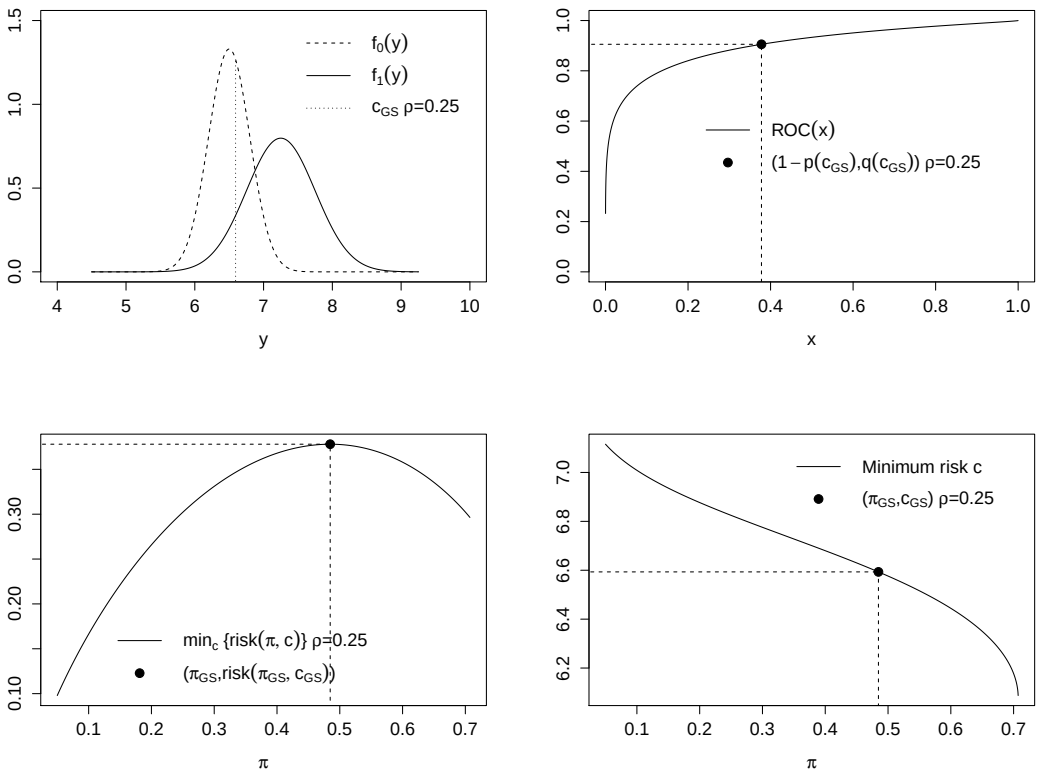


Figure 4.4: Binormal model a) with $AUC = 0.90$. *Top left panel:* the two population densities and the location of $c_{GS} = 6.5933$ for $\rho = 0.25$. *Top right panel:* the corresponding ROC curve with the operating point associated to c_{GS}, $(1 - p(c_{GS}), q(c_{GS})) = (0.3780, 0.9055)$ for $\rho = 0.25$. *Bottom left panel:* the minimum risk versus the prevalence π and the location of its maximum (the least favourable prior distribution $\pi_{GS} = 0.4847$ for $\rho = 0.25$) for which the minimax risk is $risk(\pi_{GS}, c_{GS}) = 0.3780$ for $\rho = 0.25$. *Bottom right panel:* the minimum risk threshold versus the prevalence and the location of (π_{GS}, c_{GS}).

In Figure 4.5 we have plotted the ROC curve in a 3-dimensional space that includes an extra axis with the relative cost values, and we have plotted the operating points associated to the North-West corner, the Youden index, the Concordance probability, the Symmetry point, the Generalized Youden index and the Generalized Symmetry point. This graph is useful to illustrate two aspects when selecting the optimal cutoff point: how the Generalized Symmetry point is affected by varying the relative cost value and how the Generalized Youden index is affected by varying both the relative cost and prevalence.

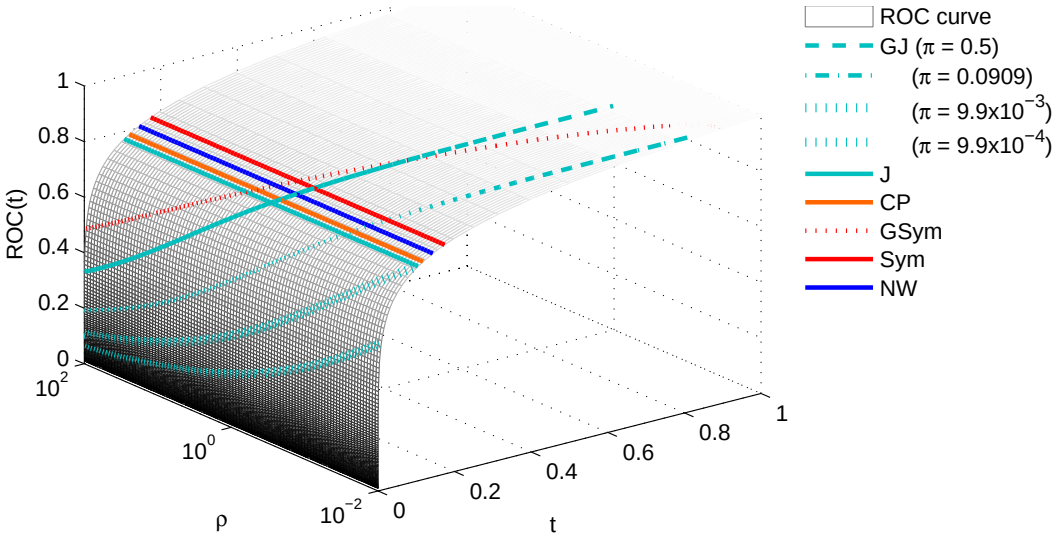


Figure 4.5: ROC curve for the binormal model a) with AUC=0.90, represented in a 3-dimensional space that includes the relative cost axis, and operating points associated to the North-West corner (*blue solid line*), the Youden index (*aquamarine solid line*), the Concordance probability (*orange solid line*), the Symmetry point (*red solid line*), the Generalized Youden index (*aquamarine dashed, dotted-dashed and dotted lines* for several values of π, $\pi = 0.5, 0.0909, 9.9 \times 10^{-3}, 9.9 \times 10^{-4}$) and the Generalized Symmetry point (*dotted red line*).

4.3 Two inference methods for the Generalized Symmetry point

As we said in the introduction of this chapter, for estimating the Generalized Symmetry point and their associated specificity and sensitivity measures, we have considered two approaches: one based on the **Generalized Pivotal Quantity** (GPQ) (Weerahandi, 1993, 1995) and the other based on **Empirical Likelihood** (EL) (Thomas and Grunkemeier, 1975). These two methods have been recently applied for computing confidence intervals for the Youden index and its corresponding optimal value, see, for instance, Lai et al. (2012), where the GPQ method has been considered under the normality assumption, and Molanes-López and Letón (2011), where the nonparametric EL method has been used instead.

In this section, we describe how to obtain percentile confidence intervals for c_{GS} and its accuracy measures $p(c_{GS})$ and $q(c_{GS})$, using both the GPQ and EL methods. Before we explain the two methods in detail, to make inference on these parameters of interest: the Generalized Symmetry point c_{GS} and its specificity $p(c_{GS})$ and sensitivity $q(c_{GS})$ accuracy measures, we consider two independent samples of i.i.d. (independently and identically distributed) observations, $\{Y_{0k_0}\}_{k_0=1}^{n_0}$ and $\{Y_{1k_1}\}_{k_1=1}^{n_1}$, taken from the healthy and diseased populations, Y_0 and Y_1, respectively, with sample sizes n_0 and n_1. Consequently, this means that for each group all the individuals follow the same parametric model but it does not mean that the same parametric model has to be considered for both groups.

4.3.1 Generalized pivotal quantity

The methodology of generalized confidence intervals was first introduced by Weerahandi (1993, 1995) and it is a parametric method based on the normality assumption which was proved to have correct asymptotic frequentist coverage under mild conditions (Hanning et al., 2006). Recently, it has been applied in the context of diagnostic studies to the Youden index by Lai et al. (2012) and Zhou and Qin (2013). In this case, we have applied this same methodology to compute generalized confidence intervals for the Generalized Symmetry point c_{GS} and the corresponding pair of its sensitivity and specificity accuracy measures ($q(c_{GS})$ and $p(c_{GS})$). But unlike Lai et al. (2012) we also consider that a monotone transformation of Box-Cox type follows a normal distribution. Assuming that the diagnostic test in healthy and diseased populations Y_i follows a normal distribution with mean μ_i and standard deviation σ_i, for $i = 0, 1$, it follows that the ROC curve has an explicit expression:

$$ROC(x) = \Phi(a + b\Phi^{-1}(x)), \tag{4.7}$$

where $a = \frac{\mu_1 - \mu_0}{\sigma_1}$, $b = \frac{\sigma_0}{\sigma_1}$, $x = 1 - p(c_{GS})$ and Φ denotes the standard normal cdf.

So, under the normality assumption, using if necessary a monotone transformation of Box-Cox type to achieve normality, the Generalized Symmetry point c_{GS} can be estimated from the following equation

$$\Phi(a + b\Phi^{-1}(x)) = 1 - \rho x \quad \Leftrightarrow \quad \Phi\left(\frac{\Phi^{-1}(1 - \rho x) - a}{b}\right) - x = 0, \tag{4.8}$$

where now a and b are replaced by their sample versions, that is, $\tilde{a} = \frac{m_1 - m_0}{s_1}$ and $\tilde{b} = \frac{s_0}{s_1}$ with m_i and s_i denoting the sample mean and sample standard deviation of each population, for $i = 0, 1$. Once the root of (4.8) is obtained, $\tilde{x}$, then the parametric point estimates of c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ are given by $\tilde{c}_{GS} = m_0 + s_0\Phi^{-1}(1 - \tilde{x})$, $\tilde{p}(c_{GS}) = 1 - \tilde{x}$ and $\tilde{q}(c_{GS}) = 1 - \rho\tilde{x}$, respectively.

For computing the GPQ-based confidence intervals for c_{GS}, $p(c_{GS})$, and $q(c_{GS})$, we follow the same reasoning as in Lai et al. (2012), based on considering their corresponding generalized pivotal quantities. For instance, if $R_{c_{GS}}$ denotes the generalized pivotal quantity for estimating c_{GS}, and $R_{c_{GS},\alpha}$ denotes the αth quantile of the distribution of

$R_{c_{GS}}$, then the $100(1-\alpha)\%$ CI for c_{GS} based on $R_{c_{GS}}$ is given by the percentile method, that is, by $(R_{c_{GS},\alpha/2}, R_{c_{GS},1-\alpha/2})$. Therefore, we propose the following four-step Monte Carlo algorithm to approximate by simulation the distribution of the generalized pivotal quantities of interest, $R_{c_{GS}}$, $R_{p(c_{GS})}$ and $R_{q(c_{GS})}$. More specifically, we repeat I times the Steps 1–3 below, for $k = 1, \ldots, I$, in order to obtain I repetitions of each of the generalized pivotal quantities involved.

- Step 1. Generate $V_i^{(k)} \sim \chi^2_{n_i-1}$ and $t_i^{(k)} \sim t_{n_i-1}$, for $i = 0, 1$.
- Step 2. Compute the generalized pivotal values of a and b:

$$R_a^{(k)} = \frac{R_{\mu_1}^{(k)} - R_{\mu_0}^{(k)}}{R_{\sigma_1}^{(k)}}, \quad R_b^{(k)} = \frac{R_{\sigma_0}^{(k)}}{R_{\sigma_1}^{(k)}},$$

  where $R_{\mu_i}^{(k)}$ and $R_{\sigma_i}^{(k)}$, for $i = 0, 1$, denote, respectively, the generalized pivotal values of μ_i and σ_i, for $i = 0, 1$, that is,

$$R_{\mu_i}^{(k)} = m_i - t_i^{(k)} \frac{s_i}{\sqrt{n_i}}, \quad R_{\sigma_i}^{(k)} = \sqrt{\frac{(n_i - 1)s_i^2}{V_i^{(k)}}}.$$

- Step 3. Obtain the root of equation (4.8) using $R_a^{(k)}$ and $R_b^{(k)}$ instead of a and b, that is, obtain the generalized pivotal quantity of x. Let $R_x^{(k)}$ be this root and set the generalized pivotal quantities of interest for estimating c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ by, respectively,

$$R_{\tilde{c}_{GS}}^{(k)} = R_{\mu_0}^{(k)} + R_{\sigma_0}^{(k)} \Phi^{-1}(1 - R_x^{(k)}), \quad \tilde{p}(c_{GS}) = 1 - R_x^{(k)}, \quad \tilde{q}(c_{GS}) = 1 - \rho R_x^{(k)}.$$

- Step 4. Finally, with the I repetitions of $R_{c_{GS}}$, $R_{p(c_{GS})}$ and $R_{q(c_{GS})}$, we obtain the corresponding GPQ-based confidence intervals for c_{GS}, $p(c_{GS})$, and $q(c_{GS})$ by the percentile method.

In the following we prove that $R_{c_{GS}} = R_{\mu_0} + R_{\sigma_0}\Phi^{-1}(1 - R_x)$, $R_{p(c_{GS})} = 1 - R_x$, and $R_{q(c_{GS})} = 1 - \rho R_x$ are the generalized pivotal quantities for estimating c_{GS}, $p(c_{GS})$, or $q(c_{GS})$, respectively, according to the definition given by Weerahandi (1993).

Proposition 4.3. According to the definition given by Weerahandi (1993), $R_{c_{GS}}$, $R_{p(c_{GS})}$, and $R_{q(c_{GS})}$ are the generalized pivotal quantities for estimating c_{GS}, $p(c_{GS})$, and $q(c_{GS})$, respectively.

Proof. From Lai et al. (2012), we know that the generalized pivotal quantity for estimating μ_i, with $i = 0, 1$, is given by

$$R_{\mu_i} = \overline{y_i} - t_i \frac{s_i}{\sqrt{n_i}},$$

where

$$\begin{aligned} t_i &= \frac{Z_i}{\sqrt{V_i/(n_i-1)}} \sim t_{n_i-1}, \\ V_i &= \frac{(n_i-1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1}, \\ Z_i &= \frac{\overline{Y}_i - \mu_i}{\sigma_i/\sqrt{n_i}}, \end{aligned}$$

and the generalized pivotal quantity for estimating σ_i, with $i = 0, 1$, is given by

$$R_{\sigma_i} = \sqrt{\frac{(n_i-1)s_i^2}{V_i}}.$$

Consequently, since $R_{c_{GS}}$, $R_{p(c_{GS})}$, and $R_{q(c_{GS})}$ have been defined from the expressions of c_{GS}, $p(c_{GS})$, and $q(c_{GS})$ by substituting the parameters μ_i and σ_i by their generalized pivotal quantities previously detailed, it follows that $R_{c_{GS}}$, $R_{p(c_{GS})}$, and $R_{q(c_{GS})}$satisfy the two necessary conditions for them to be generalized pivotal quantities for estimating c_{GS}, $p(c_{GS})$, and $q(c_{GS})$, respectively, that is, the distribution of each one is free of any unknown parameter and it coincides with the parameter of interest, c_{GS}, $p(c_{GS})$, and $q(c_{GS})$, respectively, when evaluated at $\overline{Y}_i = \overline{y}_i$ and $S_i = s_i$, with $i = 0, 1$. ■

For $\rho = 1$, that is, for the Symmetry point c_S, there is an explicit solution to the equation (4.8). We give the proof in the following proposition.

Proposition 4.4. Under the binormal model with ROC curve given by (4.7), the root of the equation (4.8) when $\rho = 1$, is given by

$$x = 1 - \Phi\left(\frac{a}{1+b}\right) = \Phi\left(-\frac{a}{1+b}\right). \tag{4.9}$$

and the Symmetry point c_S is given by

$$c_S = \mu_0 + \sigma_0 \frac{a}{1+b} = \frac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_0 + \sigma_1}. \tag{4.10}$$

Proof. First, taking into account that the standard normal density is symmetric about zero, it follows that $\Phi^{-1}(x) = -\Phi^{-1}(1-x)$.

On one hand, from (4.8) and the fact that $\Phi^{-1}(x) = -\Phi^{-1}(1-x)$, it follows that

$$\begin{aligned} a + b\Phi^{-1}(x) &= \Phi^{-1}(1-x), \\ a - b\Phi^{-1}(1-x) &= -\Phi^{-1}(x), \\ -a + b\Phi^{-1}(1-x) &= \Phi^{-1}(x), \\ x &= \Phi(-a + b\Phi^{-1}(1-x)). \end{aligned}$$

On the other hand, from (4.8) it follows that

$$\begin{aligned} a + b\Phi^{-1}(x) &= \Phi^{-1}(1-x), \\ \Phi^{-1}(x) &= \frac{\Phi^{-1}(1-x) - a}{b}, \\ x &= \Phi\left(\frac{\Phi^{-1}(1-x) - a}{b}\right). \end{aligned}$$

Consequently,

$$\Phi\left(\frac{\Phi^{-1}(1-x) - a}{b}\right) = x = \Phi(-a + b\Phi^{-1}(1-x))$$

and then, it is straightforward to obtain (4.9) because it follows that

$$\begin{aligned} \frac{\Phi^{-1}(1-x) - a}{b} &= -a + b\Phi^{-1}(1-x) \\ \Phi^{-1}(1-x)(1-b^2) &= a(1-b) \\ -\Phi^{-1}(x) = \Phi^{-1}(1-x) &= \frac{a}{1+b} \\ x = \Phi\left(-\frac{a}{1+b}\right) &= 1 - \Phi\left(\frac{a}{1+b}\right). \end{aligned}$$

Now, taking into account that

$$\Phi\left(\frac{a}{1+b}\right) = 1 - x = p(c_S) = F_0(c_S) = \Pr\left(\frac{Y_0 - \mu_0}{\sigma_0} \leq \frac{c_S - \mu_0}{\sigma_0}\right) = \Phi\left(\frac{c_S - \mu_0}{\sigma_0}\right),$$

it is easy to prove (4.10) because it follows that

$$\frac{a}{1+b} = \frac{c_S - \mu_0}{\sigma_0} \Rightarrow c_S = \mu_0 + \sigma_0 \frac{a}{1+b} = \mu_0 + \sigma_0 \frac{(\mu_1 - \mu_0)/\sigma_1}{1 + \sigma_0/\sigma_1} = \frac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_0 + \sigma_1}.$$

■

Similarly to the Generalized Symmetry point, replacing a and b by their sample versions, $\tilde{a} = \frac{m_1 - m_0}{s_1}$ and $\tilde{b} = \frac{s_0}{s_1}$ where m_i and s_i denote the sample mean and sample standard deviation of each population, for $i = 0, 1$, we propose to parametrically estimate c_S and $p(c_S) = q(c_S)$ by $\tilde{c}_S = \frac{s_1 m_0 + s_0 m_1}{s_0 + s_1}$ and $\tilde{p}(c_S) = \tilde{q}(c_S) = 1 - \tilde{x}$, respectively, where $\tilde{x} = \Phi(-\frac{\tilde{a}}{1+\tilde{b}})$.

In this case, for computing the generalized confidence intervals for c_S and $p(c_S) = q(c_S)$, we use almost the same four-step Monte Carlo algorithm previously introduced for the Generalized Symmetry point, to approximate by simulation the distribution of the generalized pivotal quantities of interest, R_{c_S} and $R_{p(c_S)} = R_{q(c_S)}$. We simply replace ρ, and c_{GS}, by $\rho = 1$, and c_S, respectively and reformulate Step 3 as follows:

- Step 3. Obtain the root of equation (4.8) using $R_a^{(k)}$ and $R_b^{(k)}$ instead of a and b in equation (4.9), that is, obtain the generalized pivotal quantity of x,

$$R_x^{(k)} = \Phi\left(-\frac{R_a^{(k)}}{1+R_b^{(k)}}\right),$$

and set the generalized pivotal quantities of interest for estimating c_S and $p(c_S) = q(c_S)$ by, respectively,

$$R_{c_S}^{(k)} = \frac{R_{\sigma_1}^{(k)} R_{\mu_0}^{(k)} + R_{\sigma_0}^{(k)} R_{\mu_1}^{(k)}}{R_{\sigma_0}^{(k)} + R_{\sigma_1}^{(k)}}, \quad R_{p(c_S)}^{(k)} = R_{q(c_S)}^{(k)} = 1 - R_x^{(k)}.$$

4.3.2 Empirical likelihood

The origins of the **Empirical likelihood** (EL) methodology go back to 1975 with the construction of EL-based confidence intervals for the Kaplan-Meier estimator (Thomas and Grunkemeier, 1975). Nowadays, this methodology is an active area of research in several fields due to the good properties presented by the EL-based confidence intervals and confidence regions, see for example, Molanes-López et al. (2009), among others. Taking into account the fact that the optimal threshold defined by the Youden index can be seen as two specific quantiles on the healthy and diseased populations, a bootstrap-based EL approach was recently proposed to make inference on the Youden index and its associated optimal threshold (Molanes-López and Letón, 2011). Specifically, the EL methodology for estimating a population quantile (Zhou and Jing, 2003) was extended to the two-sample case. In this subsection, we apply these same ideas to the Generalized Symmetry point with the purpose of obtaining confidence intervals for c_{GS} and its accuracy measures $p(c_{GS})$ and $q(c_{GS})$. This approach has the advantages of easy implementation and not requiring any particular parametric assumption.

Let (x, s) be an arbitrary operating point on the ROC curve and let c be the corresponding threshold value on the marker scale, that is, satisfying that $x = 1 - p(c) = 1 - F_0(c)$ and $s = q(c) = 1 - F_1(c)$. If we consider that x is known in advance and that s plays the role of a nuisance parameter, it follows from Molanes-López and Letón (2011) that the adjusted empirical log-likelihood ratio function to estimate c is given by

$$\begin{aligned}\ell(c,s) &= -2\log(R(c,s)) \qquad (4.11)\\ &= 2n_0\hat{F}_{0,g_0}(c)\log\left(\frac{\hat{F}_{0,g_0}(c)}{1-x}\right) + 2n_0(1-\hat{F}_{0,g_0}(c))\log\left(\frac{1-\hat{F}_{0,g_0}(c)}{x}\right)\\ &\quad + 2n_1\hat{F}_{1,g_1}(c)\log\left(\frac{\hat{F}_{1,g_1}(c)}{1-s}\right) + 2n_1(1-\hat{F}_{1,g_1}(c))\log\left(\frac{1-\hat{F}_{1,g_1}(c)}{s}\right),\end{aligned}$$

where $R(c, s)$ refers to the EL ratio function for the parameter of interest c, and $\hat{F}_{i,g_i}$ is a kernel-type estimate of the cdf F_i, for $i = 0, 1$, that is,

$$\hat{F}_{i,g_i}(y) = \frac{1}{n_i}\sum_{k_i=1}^{n_i}\mathbb{K}\left(\frac{y - Y_{ik_i}}{g_i}\right), \qquad (4.12)$$

where g_i refers to the smoothing parameter, for $i = 0, 1$, $\mathbb{K}(y) = \int_{-\infty}^{y} K(z)\mathrm{d}z$ and K denotes a kernel function.

In the following proposition we will give a more detailed derivation of (4.11).

Proposition 4.5. The adjusted empirical log-likelihood function to estimate c is given by (4.11).

Proof. Under the setting considered in Subsection 4.3.2, the parameter of interest, $c \in \Re$, can be seen as two specific quantiles, the $(1-x)$-th quantile of the healthy population and the $(1-s)$-th quantile of the diseased population. Therefore, it follows from Molanes-López and Letón (2011) that the EL ratio function for estimating c can be formulated by

$$R(c,s) = \sup_{p_{0k_0}, p_{1k_1}} \frac{\prod_{k_0=1}^{n_0} p_{0k_0} \prod_{k_1=1}^{n_1} p_{1k_1}}{n_0^{-n_0} n_1^{-n_1}},$$

subject to the following restrictions,

$$\begin{aligned} p_{0k_0} \geq 0, k_0 = 1, \ldots, n_0, \qquad & p_{1k_1} \geq 0, k_1 = 1, \ldots, n_1, \\ \sum_{k_0=1}^{n_0} p_{0k_0} = 1, \qquad & \sum_{k_1=1}^{n_1} p_{1k_1} = 1, \\ \sum_{k_0=1}^{n_0} p_{0k_0} \delta_0 \left(Y_{0k_0} - c\right) = 0, \qquad & \sum_{k_1=1}^{n_1} p_{1k_1} \delta_1 \left(Y_{1k_1} - c\right) = 0, \end{aligned}$$

where

$$\delta_0(z) = \begin{cases} -1, & \text{if } z \leq 0 \\ \frac{1-x}{x}, & \text{otherwise,} \end{cases}$$

and

$$\delta_1(z) = \begin{cases} -1, & \text{if } z \leq 0 \\ \frac{1-s}{s}, & \text{otherwise.} \end{cases}$$

By the Lagrange multiplier method, it is easy to prove that the maximization of $R(c,s)$ occurs with

$$\begin{aligned} p_{0k_0} &= \frac{1}{n_0} \frac{1}{1 + \lambda_0(c,s)\delta_0(Y_{0k_0} - c)}, \quad \text{for } k_0 = 1, \ldots, n_0, \\ p_{1k_1} &= \frac{1}{n_1} \frac{1}{1 + \lambda_1(c,s)\delta_1(Y_{1k_1} - c)}, \quad \text{for } k_1 = 1, \ldots, n_1, \end{aligned}$$

where $\lambda_0(c,s)$ and $\lambda_1(c,s)$ satisfy the equations below

$$\frac{1}{n_0} \sum_{k_0=1}^{n_0} \frac{\delta_0(Y_{0k_0} - c)}{1 + \lambda_0(c,s)\delta_0(Y_{0k_0} - c)} = 0, \tag{4.13}$$

$$\frac{1}{n_1} \sum_{k_1=1}^{n_1} \frac{\delta_1(Y_{1k_1} - c)}{1 + \lambda_1(c,s)\delta_1(Y_{1k_1} - c)} = 0. \tag{4.14}$$

When $c \in [Y_{0[1]}, Y_{0[n_0]}]$ and $c \in [Y_{1[1]}, Y_{1[n_1]}]$ simultaneously, with $Y_{i[k]}$ denoting the kth ordered statistic of $\{Y_{i1}, \ldots, Y_{in_i}\}$, for $i = 0, 1$, then the solutions to the equations (4.13) and (4.14), $\lambda_0(c, s)$ and $\lambda_1(c, s)$, are given by, respectively,

$$\lambda_0(c, s) = \frac{1 - x - F_{0n_0}(c)}{1 - x}, \qquad \lambda_1(c, s) = \frac{1 - s - F_{1n_1}(c)}{1 - s},$$

where F_{in_i} stands for the empirical cdf of Y_i, with $i = 0, 1$.

Consequently, the adjusted log-likelihood ratio can be rewritten as given in (4.11):

$$\begin{aligned}
\ell(c, s) &= -2\log(R(c, s)) \\
&= 2n_0\hat{F}_{0,g_0}(c)\log\left(\frac{\hat{F}_{0,g_0}(c)}{1 - x}\right) + 2n_0(1 - \hat{F}_{0,g_0}(c))\log\left(\frac{1 - \hat{F}_{0,g_0}(c)}{x}\right) \\
&\quad + 2n_1\hat{F}_{1,g_1}(c)\log\left(\frac{\hat{F}_{1,g_1}(c)}{1 - s}\right) + 2n_1(1 - \hat{F}_{1,g_1}(c))\log\left(\frac{1 - \hat{F}_{1,g_1}(c)}{s}\right),
\end{aligned}$$

where we have replaced F_{0n_0} and F_{1n_1} appearing in $\lambda_0(c, s)$ and $\lambda_1(c, s)$, respectively, by the kernel-type estimates given in (4.12), following the recommendation of Zhou and Jing (2003). ■

Since the Generalized Symmetry point defines an operating point on the ROC curve satisfying $x = 1 - p(c_{GS})$ and $s = 1 - \rho(1 - p(c_{GS})) = 1 - \rho x$, the adjusted empirical log-likelihood ratio function given above in (4.11), $\ell(c, s)$, is reduced to:

$$\begin{aligned}
\ell(c) &= 2n_0\hat{F}_{0,g_0}(c)\log\left(\frac{\hat{F}_{0,g_0}(c)}{1 - x}\right) + 2n_0(1 - \hat{F}_{0,g_0}(c))\log\left(\frac{1 - \hat{F}_{0,g_0}(c)}{x}\right) \\
&\quad + 2n_1\hat{F}_{1,g_1}(c)\log\left(\frac{\hat{F}_{1,g_1}(c)}{\rho x}\right) + 2n_1(1 - \hat{F}_{1,g_1}(c))\log\left(\frac{1 - \hat{F}_{1,g_1}(c)}{1 - \rho x}\right),
\end{aligned} \tag{4.15}$$

where the nuisance parameter s has vanished.

Therefore, assuming that x is known in advance, a point estimate of the corresponding Generalized Symmetry point could be found by minimizing $\ell(c)$, given in (4.15), over c, and a confidence interval for it could be obtained based on the usual χ^2 limiting distribution of $\ell(c)$. However, this is not possible here because x is not known in advance. So, since x is unknown we propose the procedure detailed below in Steps 1–2 to obtain nonparametric point estimates of c_{GS}, $p(c_{GS})$ and $q(c_{GS})$, and the bootstrap-based strategy specified in Steps 3–4 to construct EL-based CIs. The complete four-step procedure is as follows:

- Step 1. We first estimate the ROC curve from the original pair of samples by the kernel type estimator given by

$$\widehat{ROC}(x) = 1 - \frac{1}{n_1}\sum_{k_1=1}^{n_1}\mathbb{K}\left(\frac{1 - x - \hat{F}_{0,g_0}(Y_{1k_1})}{h_0}\right),$$

where h_0 is a smoothing parameter. A point estimate of the FPR x, denoted by $\hat{x}$, is obtained by intersecting $\widehat{ROC}(x)$ with the line $y = 1 - \rho x$, and consequently, the corresponding specificity and sensitivity can be estimated by $\hat{p}_1 = 1 - \hat{x}$ and $\hat{q}_1 = 1 - \rho\hat{x}$, respectively.

We now interchange the axes of the ROC curve and estimate the corresponding graph, that is, we reflect the ROC curve about the positive diagonal line, and get the inverse graph of the ROC curve that we denote by ROC^{-1}. This graph is formed by representing the points $(q(c), 1-p(c))$, with $-\infty < c < \infty$, which, analogously to the ROC curve, can be reparameterized in the interval $(0, 1)$ as follows $ROC^{-1}(s) = 1-F_0(F_1^{-1}(1-s))$, where $s = q(c) = 1-F_1(c) = TPR(c) \in (0, 1)$. Then, we estimate $ROC^{-1}(s)$ using the following kernel type estimator

$$\widehat{ROC^{-1}}(s) = 1 - \frac{1}{n_0}\sum_{k_1=1}^{n_0} \mathbb{K}\left(\frac{1 - s - \hat{F}_{1,g_1}(Y_{0k_0})}{h_1}\right),$$

where h_1 is a smoothing parameter. A point estimate of the TPR s, denoted by $\hat{s}$, is obtained by intersecting $\widehat{ROC^{-1}}(s)$ with the line $y = \frac{1}{\rho}(1 - s)$, and consequently, the corresponding specificity and sensitivity can be alternatively estimated by $\hat{p}_2 = 1 - \frac{1}{\rho}(1 - \hat{s})$ and $\hat{q}_2 = \hat{s}$, respectively.

Finally, we propose to nonparametrically estimate the specificity $p(c_{GS})$ by averaging the two point estimates previously obtained, that is, $\hat{p}(c_{GS}) = \frac{1}{2}(\hat{p}_1 + \hat{p}_2)$. Analogously, the sensitivity is estimated by averaging $\hat{q}_1$ and $\hat{q}_2$, that is, $\hat{q}(c_{GS}) = \frac{1}{2}(\hat{q}_1 + \hat{q}_2)$.

- Step 2. With the previous point estimate of $p(c_{GS})$, we then minimize in c the log-likelihood ratio $\ell(c)$ given above in (4.15) and propose the minimizer found, $\hat{c}_{GS}$, as the nonparametric point estimate of c_{GS}.

- Step 3. In order to obtain confidence intervals for c_{GS}, $p(c_{GS})$, and $q(c_{GS})$, we resample independently from the original pair of samples a large number of times, let's say B, and repeat Steps 1–2 given above for each pair of bootstrapped resamples.

- Step 4. Finally, the bootstrap EL-based confidence intervals for the Generalized Symmetry point and its corresponding specificity and sensitivity indexes are constructed by the percentile method using the B bootstrap estimates of c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ previously obtained in Step 3.

4.4 Simulation study

In this section, we carry out a simulation study in R (R Core Team, 2013) to compare the performance of the two approaches previously introduced in the former section. In this simulation study, we focus on the mean squared error (MSE), bias ($Bias$) and standardized bias ($SBias$) of the point estimators previously proposed for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$, and on the coverage probability ($cover$) and average width or expected length

($width$) of the GPQ-based and EL-based confidence intervals previously proposed for these parameters of interest. For instance, for the parametric point estimator $\tilde{c}_{GS}$ specified in Subsection 4.3.1, we have approximated by Monte Carlo (using here 2000 trials) the following performance measures:

$$\begin{aligned} MSE(\tilde{c}_{GS}) &= E\left[(\tilde{c}_{GS} - c_{GS})^2\right] = \mathbb{V}ar[\tilde{c}_{GS}] + Bias^2(\tilde{c}_{GS}), \\ Bias(\tilde{c}_{GS}) &= E\left[\tilde{c}_{GS}\right] - c_{GS}, \\ SBias(\tilde{c}_{GS}) &= \frac{Bias(\tilde{c}_{GS})}{\sqrt{\mathbb{V}ar[\tilde{c}_{GS}]}}. \end{aligned}$$

An interesting interpretation of $SBias^2$ is that it gives the increase of MSE over $\mathbb{V}ar$ due to $Bias^2$. Besides, the proportion of MSE due to $Bias^2$ is just $\frac{Bias^2}{MSE} = \frac{1}{1+(SBias^{-2})}$.

Regarding the GPQ-based CI for c_{GS}, $(R_{c_{GS},\alpha/2}, R_{c_{GS},1-\alpha/2})$, specified in Subsection 4.3.1, we have approximated by Monte Carlo (using 2000 trials) the following performance measures:

$$\begin{aligned} cover((R_{c_{GS},\alpha/2}, R_{c_{GS},1-\alpha/2})) &= \Pr(c_{GS} \in (R_{c_{GS},\alpha/2}, R_{c_{GS},1-\alpha/2})), \\ width((R_{c_{GS},\alpha/2}, R_{c_{GS},1-\alpha/2})) &= E[R_{c_{GS},1-\alpha/2} - R_{c_{GS},\alpha/2}]. \end{aligned}$$

In terms of $cover$, good performance means that this value is as close as possible to its nominal value. In terms of $width$, MSE or $Bias$, the lower their value, the better. However, for a confidence interval it is more important to have good performance in terms of $cover$ albeit its $width$ is wider.

In the following subsections, we first describe the scenarios considered in our simulation study and then we discuss the results obtained.

4.4.1 Scenarios

Inspired in Fluss et al. (2005), we specify below and in Table 4.1 the parameters of the six models considered in our simulation study:

a) $Y_0 \sim Normal(\mu_0 = 6.5, \sigma_0^2 = 0.09)$ and $Y_1 \sim Normal(\mu_1, \sigma_1^2 = 0.25)$.

b) $Y_0^{-1/3} \sim Normal(\mu_0 = 3.5, \sigma_0^2 = 0.09)$ and $Y_1^{-1/3} \sim Normal(\mu_1, \sigma_1^2 = 0.25)$.

c) $\ln(Y_0) \sim Normal(\mu_0 = 2.5, \sigma_0^2 = 0.09)$ and $\ln(Y_1) \sim Normal(\mu_1, \sigma_1^2 = 0.25)$.

d) $Y_0 \sim Gamma(\alpha_0 = 2, \beta_0 = 2)$ and $Y_1 \sim Gamma(\alpha_1 = 2, \beta_1)$.

e) $Y_0 \sim 0.50Normal(\mu_{0,1} = 10, \sigma_{0,1}^2 = 1) + 0.50Normal(\mu_{0,2} = 13, \sigma_{0,2}^2 = 1)$ and $Y_1 \sim 0.50Normal(\mu_{1,1}, \sigma_{1,1}^2 = 1) + 0.50Normal(\mu_{1,2}, \sigma_{1,2}^2 = 5)$.

f) $Y_0 \sim 0.75Beta(\alpha_{0,1} = 1, \beta_{0,1} = 3) + 0.25Beta(\alpha_{0,2} = 5, \beta_{0,2} = 1.75)$ and $Y_1 \sim 0.75Beta(\alpha_{1,1}, \beta_{1,1} = 2) + 0.25Beta(\alpha_{1,2}, \beta_{1,2} = 4.5)$.

Table 4.1: Parameters of the scenarios considered in the simulation study of Section 4.4.

a)	μ_0	σ_0^2	σ_1^2	AUC	μ_1	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	6.5	0.09	0.25	0.70	6.81	0.1	6.136	0.11	0.91
						0.25	6.343	0.30	0.83
						0.5	6.487	0.48	0.74
						1	6.615	0.65	0.65
						2	6.732	0.78	0.56
						4	6.837	0.87	0.48
						10	6.962	0.94	0.38
				0.80	6.99	0.1	6.278	0.23	0.92
						0.25	6.452	0.44	0.86
						0.5	6.573	0.60	0.80
						1	6.684	0.73	0.73
						2	6.786	0.83	0.66
						4	6.879	0.90	0.59
						10	6.993	0.95	0.50
				0.90	7.25	0.1	6.454	0.44	0.94
						0.25	6.592	0.62	0.91
						0.5	6.689	0.74	0.87
						1	6.782	0.83	0.83
						2	6.866	0.89	0.78
						4	6.947	0.93	0.73
						10	7.046	0.97	0.66

b)	μ_0	σ_0^2	σ_1^2	AUC	μ_1	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	3.5	0.09	0.25	0.70	3.19	0.1	0.017	0.11	0.91
						0.25	0.020	0.30	0.82
						0.5	0.023	0.48	0.74
						1	0.026	0.65	0.65
						2	0.029	0.78	0.56
						4	0.032	0.87	0.48
						10	0.036	0.94	0.38
				0.80	3.01	0.1	0.019	0.23	0.92
						0.25	0.022	0.44	0.86
						0.5	0.025	0.60	0.80
						1	0.027	0.73	0.73
						2	0.030	0.83	0.66
						4	0.033	0.90	0.59
						10	0.037	0.95	0.50
				0.90	2.75	0.1	0.022	0.44	0.94
						0.25	0.025	0.62	0.91
						0.5	0.028	0.74	0.87
						1	0.030	0.83	0.83
						2	0.032	0.89	0.78
						4	0.035	0.93	0.73
						10	0.039	0.97	0.66

Table 4.1: *Continued*.

c)	μ_0	σ_0^2	σ_1^2	AUC	μ_1	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	2.5	0.09	0.25	0.70	2.81	0.1	8.436	0.11	0.91
						0.25	10.383	0.30	0.82
						0.5	11.997	0.48	0.74
						1	13.664	0.65	0.65
						2	13.353	0.78	0.56
						4	17.055	0.87	0.48
						10	19.318	0.94	0.38
				0.80	2.99	0.1	9.760	0.23	0.92
						0.25	11.617	0.44	0.86
						0.5	13.109	0.60	0.80
						1	14.645	0.73	0.73
						2	16.213	0.83	0.66
						4	17.805	0.90	0.59
						10	19.945	0.95	0.50
				0.90	3.25	0.1	11.633	0.44	0.94
						0.25	13.359	0.62	0.91
						0.5	14.723	0.74	0.87
						1	16.127	0.83	0.83
						2	17.567	0.89	0.78
						4	19.040	0.93	0.73
						10	21.035	0.97	0.66

d)	β_0	α_0	α_1	AUC	β_1	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	2	2	2	0.70	3.51	0.1	1.635	0.20	0.92
						0.25	2.505	0.36	0.84
						0.5	3.364	0.50	0.75
						1	4.388	0.64	0.64
						2	5.554	0.76	0.53
						4	6.832	0.85	0.42
						10	8.646	0.93	0.29
				0.80	4.97	0.1	2.153	0.29	0.93
						0.25	3.149	0.47	0.87
						0.5	4.070	0.60	0.80
						1	5.117	0.73	0.73
						2	6.269	0.82	0.64
						4	7.505	0.89	0.55
						10	9.236	0.94	0.45
				0.90	8.23	0.1	3.072	0.45	0.95
						0.25	4.221	0.62	0.91
						0.5	5.212	0.73	0.87
						1	6.289	0.82	0.82
						2	7.437	0.89	0.77
						4	8.641	0.93	0.72
						10	10.303	0.96	0.64

Table 4.1: *Continued*.

e)	$(\mu_{0,1}, \mu_{0,2})$	$(\sigma^2_{0,1}, \sigma^2_{0,2})$	$(\sigma^2_{1,1}, \sigma^2_{1,2})$	AUC	$(\mu_{1,1}, \mu_{1,2})$	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	(10.00,13.00)	(1.00,1.00)	(1.00,5.00)	0.70	(11.43,15.43)	0.1	10.288	0.31	0.93
						0.25	10.834	0.41	0.85
						0.5	11.378	0.48	0.74
						1	12.123	0.59	0.59
						2	12.924	0.73	0.47
						4	13.538	0.85	0.41
						10	14.137	0.94	0.36
				0.80	(12.39,16.39)	0.1	11.114	0.45	0.94
						0.25	11.633	0.52	0.88
						0.5	12.105	0.58	0.79
						1	12.639	0.68	0.68
						2	13.178	0.78	0.57
						4	13.666	0.87	0.49
						10	14.210	0.94	0.43
				0.90	(13.51,17.51)	0.1	12.079	0.58	0.96
						0.25	12.515	0.65	0.91
						0.5	12.864	0.72	0.86
						1	13.220	0.79	0.79
						2	13.575	0.86	0.72
						4	13.921	0.91	0.64
						10	14.353	0.96	0.56

f)	α_0	β_0	β_1	AUC	α_1	ρ	c_{GS}	$p(c_{GS})$	$q(c_{GS})$
	(1.00,5.00)	(3.00,1.75)	(2.00,4.50)	0.70	(3.70,2.00)	0.1	0.175	0.33	0.93
						0.25	0.268	0.46	0.86
						0.5	0.361	0.56	0.78
						1	0.472	0.66	0.66
						2	0.595	0.75	0.49
						4	0.713	0.83	0.31
						10	0.832	0.91	0.13
				0.80	(7.25,3.00)	0.1	0.258	0.44	0.94
						0.25	0.360	0.56	0.89
						0.5	0.462	0.65	0.82
						1	0.577	0.73	0.73
						2	0.681	0.81	0.61
						4	0.766	0.87	0.46
						10	0.851	0.93	0.27
				0.90	(16.50,7.00)	0.1	0.455	0.64	0.96
						0.25	0.535	0.70	0.93
						0.5	0.609	0.76	0.88
						1	0.692	0.81	0.81
						2	0.768	0.87	0.74
						4	0.827	0.91	0.64
						10	0.882	0.95	0.47

For those parameters whose values have been left unspecified, see Table 4.1, where we collect the three possible values considered for them (with the corresponding values of c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ within parentheses) in order to get three different values of $AUC = 0.70, 0.80, 0.90$, and consequently three different discriminating abilities.

The first three scenarios a)–c) correspond to the binormal model, that is, either Y_0 and Y_1 or a Box-Cox transformation of them follow normal distributions (see, for example, Fluss et al. (2005) and Molodianovitch et al. (2006)). Under these binormal scenarios, it should be noted that there are closed form expressions for the Symmetry point (that is, only in the case $\rho = 1$), which could be expressed as a function of the corresponding means and standard deviations:

a) $c_S = \dfrac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_0 + \sigma_1}$, for model a),

b) $c_S = \left(\dfrac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_0 + \sigma_1}\right)^{-3}$, for model b),

c) $c_S = e^{\dfrac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_0 + \sigma_1}}$, for model c).

In the following we will present the mathematical relationship between AUC and c_{GS} for the binormal model.

Proposition 4.6. Under the binormal model it follows that

$$\begin{aligned} AUC &= \Phi_2^{\rho_1}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_{GS}))\right) \\ &\quad + \Phi_2^{\rho_2}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_{GS}))\right) \\ &\quad + p(c_{GS})(1-\rho(1-p(c_{GS}))), \end{aligned}$$

where Φ^{-1} denotes the quantile function of the standard normal random variable, $\Phi_2^{\rho_i}$, for $i = 1, 2$, refers to the cdf of a bivariate normal random vector with marginal means equal to zero, marginal variances equal to one and correlation coefficient ρ_i given by

$$\rho_1 = -\frac{b}{\sqrt{1+b^2}}, \qquad \rho_2 = -\frac{1}{\sqrt{1+b^2}}.$$

Proof. From Hillis and Metz (2012), we know that the partial AUC corresponding to the area under the binormal ROC curve over the interval of FPF's given by $(0, 1-p(c_{GS}))$ is equal to

$$pAUC(0, 1-p(c_{GS})) = \Phi_2^{\rho_1}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_{GS}))\right),$$

and that the partial AUC corresponding to the area to the right of the binormal ROC curve in the interval of TPF's given by $(q(c_{GS}), 1)$ is equal to

$$pAUC(q(c_{GS}), 1) = \Phi_2^{\rho_2}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-q(c_{GS}))\right).$$

Besides, taking into account that

$$AUC = pAUC(0, 1-p(c_{GS})) + pAUC(q(c_{GS}), 1) + p(c_{GS})q(c_{GS}),$$

with $\rho(1-p(c_{GS}))=1-q(c_{GS})$, we can conclude that

$$\begin{aligned} AUC &= \Phi_2^{\rho_1}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_{GS}))\right) \\ &\quad +\Phi_2^{\rho_2}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_{GS}))\right) \\ &\quad +p(c_{GS})(1-\rho(1-p(c_{GS}))) \end{aligned}$$

under the binormal model. ■

For the Symmetry point c_S (a particular case of c_{GS} with ρ=1), the mathematical relationship between AUC and c_S for the binormal model is reduced to:

$$AUC = \Phi_2^{\rho_1}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_S))\right) + \Phi_2^{\rho_2}\left(\frac{a}{\sqrt{1+b^2}}, \Phi^{-1}(1-p(c_S))\right) + p(c_S)^2.$$

The last three scenarios d)–f) are outside the Box-Cox transformation family and have been included to study the robustness of the GPQ approach. Note that when we consider/apply the GPQ method, we will always transform the data by a transformation of Box-Cox type. Under these scenarios Y_0 and Y_1 are either gamma distributed or follow mixtures of two normal random variables or mixtures of two beta random variables. In Figure 4.6, the two gamma densities are plotted in the left panel and the associated ROC curve in the right panel. Since for these scenarios there are no closed form expressions for the Generalized Symmetry point, we have used the R function *uniroot* from the *Stats* package to numerically approximate the value of this parameter. The same for the three binormal scenarios a)–c) for the Generalized Symmetry point when $\rho \neq 1$. It is worth mentioning that the scenarios e)–f) give rise to very different parametric shapes between groups as can be seen in Figure 4.7.

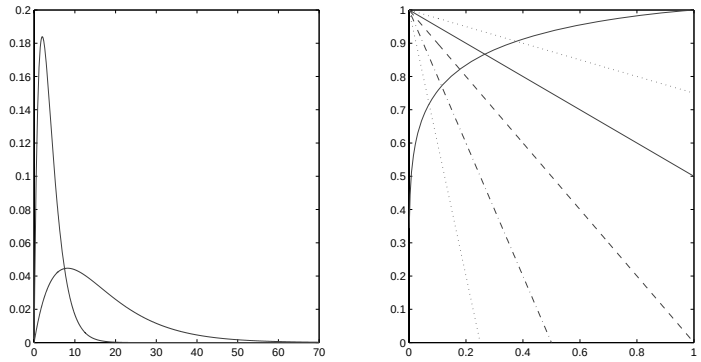


Figure 4.6: *Left panel*: gamma density with parameters $\alpha_0 = 2$ and $\beta_0 = 2$, and gamma density with parameters $\alpha_1 = 2$ and $\beta_1 = 8.23$. *Right panel*: associated ROC curve (*thick black solid line*) and lines given by equations $y = 1 - \rho t$ with $\rho = 4, 2, 1, 0.5, 0.25$ (*dotted, dashed-dotted, dashed, solid and dotted black lines*, respectively). The corresponding Generalized Symmetry points are given by 8.641, 7.437, 6.289, 5.212, and 4.221, respectively.

For every scenario we have considered three possible values of $AUC = 0.70, 0.80, 0.90$ (as we pointed out before) and seven different values of $\rho = 0.1, 0.25, 0.5, 1, 2, 4, 10$. In Figure 4.7, we have plotted the two density functions associated to the six scenarios previously described in the simulation study a)–f) with $\rho = 1$ and $AUC = 0.90$.

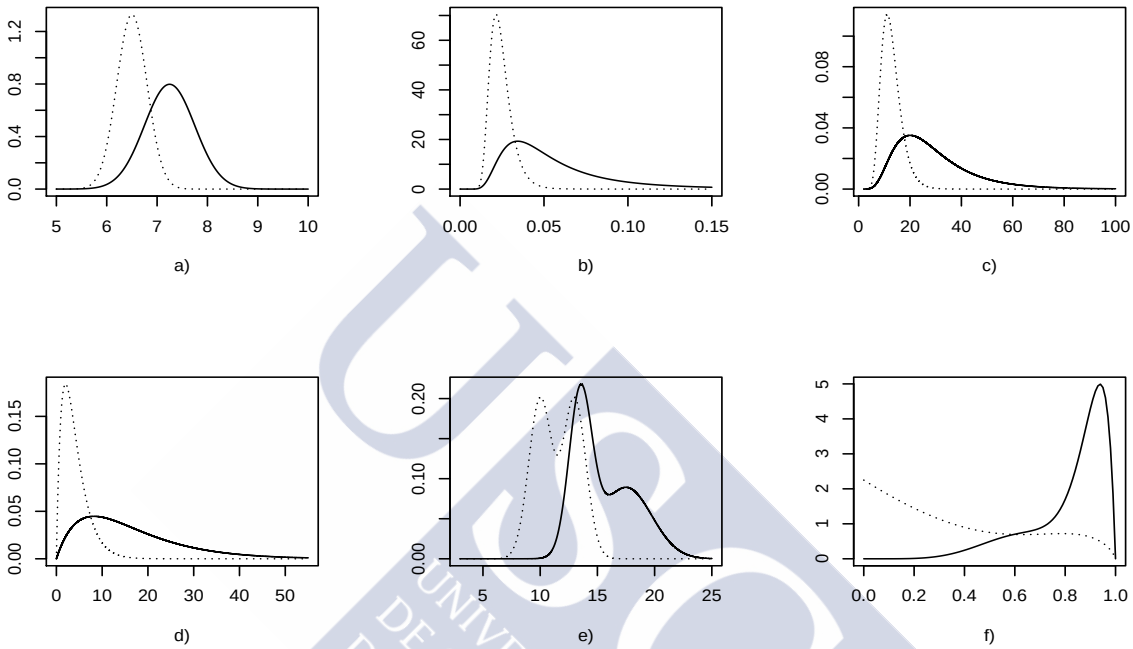


Figure 4.7: Density functions, f_0 (*dotted lines*) and f_1 (*solid lines*), of the six models considered in Table 4.1 when $AUC = 0.90$.

For every model and AUC value specified in Table 4.1, 2000 trials were considered. For each trial, a sample of n_0 i.i.d. observations, $\{Y_{01}, \ldots, Y_{0n_0}\}$, and a sample of n_1 i.i.d. observations, $\{Y_{11}, \ldots, Y_{1n_1}\}$, were independently drawn from Y_0 and Y_1, respectively. Specifically we have considered the following pairs of sample sizes (considering balanced and unbalanced designs): $(n_0, n_1) = (30, 30), (30, 50), (50, 50), (100, 50), (100, 100)$. These sample sizes and the three values previously considered for AUC mimic real situations that are frequently seen in clinical practice (see, for instance, the biomedical datasets considered in Section 4.5).

For the GPQ approach, since it is based on the normality assumption, before constructing the GPQ-based CIs for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$, we have previously transformed the data in order to make them as normally distributed as possible by using a monotone Box-Cox transformation estimated by maximum likelihood. Besides, we have run $I = 2500$ times the Steps 1-3 previously presented in Subsection 4.3.1.

For the EL method, the Gaussian kernel, K, was considered in the required kernel

type estimates. For these kernel type estimates, we considered the following bandwidths as recommended by Wand and Jones (1995) and Zhou and Jing (2003),

$$g_i = 0.25\,\text{std}(Y_{i1},\ldots,Y_{in_i})n_i^{\gamma}, \quad \text{for } i=0,1, \tag{4.16}$$

$$h_0 = 0.25\,\text{std}(\hat{F}_{0,g_0}(Y_{11}),\ldots,\hat{F}_{0,g_0}(Y_{1n_1}))n_1^{\delta}, \tag{4.17}$$

$$h_1 = 0.25\,\text{std}(\hat{F}_{1,g_1}(Y_{01}),\ldots,\hat{F}_{1,g_1}(Y_{0n_0}))n_0^{\delta}, \tag{4.18}$$

with std referring to the sample standard deviation. Specifically, we have considered $\gamma = -1/3$ to estimate F_0 and $\delta = -1/3$ in Step 1, $\gamma = -1/2$ to estimate F_0 and F_1 in the adjusted empirical log-likelihood ratio function in Step 2, and $\gamma = -1/5$ to draw the bootstrap resamples in Step 3. From each pair of samples, we generated $B = 499$ bootstrapped resamples to obtain 95%-confidence intervals for c_{GS} and its accuracy measures $p(c_{GS})$ and $q(c_{GS})$ through the percentile method. When Steps 1–2 are carried out with the bootstrap resamples, then the original samples $Y_{i1},\ldots,Y_{in_i}$, for $i=0,1$ appearing in (4.16)–(4.18) are replaced by their bootstrap analogues $Y_{i1}^{*,b},\ldots,Y_{in_i}^{*,b}$, for $i=0,1$, with $b=1,\ldots,B$.

4.4.2 Results

For every model, we collect in the same table the results obtained for the Generalized Symmetry point, c_{GS}, and the associated specificity $p(c_{GS})$ and sensitivity $q(c_{GS})$ measures. Specifically, in Tables 4.2–4.19 (see pages 115–169) we collect the results obtained for models a)–f) considering the five pair of sample sizes (n_0, n_1), the seven different values of ρ and the three possible values of AUC, previously specified. It should be noted that depending on the marker scale the CIs for c_{GS} may turn out wider or narrower, and consequently we have used different scientific notation ($\times 10^0$, $\times 10^{-1}$, $\times 10^{-2}$, etc.) in these tables to better represent significant figures. For instance, the $width$ of the CIs for c_{GS} is expressed in $\times 10^{-1}$ for model a) but in $\times 10^{-3}$ for model b).

From Tables 4.2–4.13, we observe the following: In terms of interval coverage, both methods have a good behaviour when estimating the Generalized Symmetry point, and the corresponding specificity and sensitivity indexes. With respect to average width, the EL approach gives slightly wider confidence intervals than the GPQ approach. This was expected since the EL approach is a nonparametric method and the GPQ approach is a parametric one. Besides, for both methods the interval width decreases as the sample sizes increase. For a fixed pair of sample sizes when $\rho = 1$, the interval width increases for c_{GS} as the AUC increases but it decreases for the corresponding $p(c_{GS})$ and $q(c_{GS})$. Regarding the point estimators, the MSE decreases as the sample sizes increase for both methods. However, the nonparametric EL approach always produces slightly higher values for the MSE than the parametric GPQ approach. Besides, for a fixed pair of sample sizes when $\rho = 1$, the MSE increases for c_{GS} as the AUC increases but it decreases for the corresponding $p(c_{GS})$ and $q(c_{GS})$. In terms of $Bias$, the nonparametric point estimators proposed for $p(c_{GS})$ and $q(c_{GS})$ behave in most of the cases better than the parametric proposal. However, there is not a clear winner between the two point estimators of c_{GS} in terms of $Bias$.

Table 4.2: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model a) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.75	4.4792	11.8502	9.7834	9.0215	92.20	4.8250	16.4288	11.3058	8.8529
	0.25	94.50	2.9948	5.4270	6.8992	9.4045	94.55	3.4283	7.8588	-0.0018	-0.0020
	0.5	94.45	2.2903	3.4760	4.3613	7.4161	95.05	2.7188	4.8662	3.3006	4.7356
	1	93.75	2.0089	2.7333	1.0445	1.9978	96.20	2.3860	3.5772	1.2848	2.1481
	2	94.40	2.0821	2.7900	-1.4756	-2.7940	95.10	2.4379	3.8052	-0.7004	-1.1352
	4	94.95	2.3717	3.4237	-3.0444	-5.2088	93.95	2.7349	4.7932	-1.9184	-2.7713
	10	94.60	2.9151	5.2732	-5.3197	-7.3436	87.60	3.2350	7.9755	-3.1519	-3.5307
(30,50)	0.1	93.30	3.3613	7.0380	7.5520	9.0364	93.65	3.8642	9.8799	6.4165	6.4673
	0.25	95.35	2.3338	3.4896	7.3682	12.5681	94.95	2.7705	4.8304	3.5716	5.1444
	0.5	93.15	1.8831	2.4998	4.8451	9.7342	95.00	2.2764	3.4022	3.2341	5.5517
	1	93.35	1.7873	2.2360	1.4100	2.9825	95.60	2.1495	3.0313	1.2093	2.1964
	2	94.85	1.9661	2.4846	-1.4909	-2.9916	95.15	2.3050	3.3826	-0.4870	-0.8372
	4	95.95	2.3048	3.1752	-3.5762	-6.3577	93.50	2.6431	4.5161	-2.2292	-3.3181
	10	94.95	2.8598	4.8059	-5.5683	-8.0563	87.50	3.1671	7.5648	-3.3769	-3.8845
(50,50)	0.1	93.15	3.3844	7.4299	3.9023	4.5307	93.70	3.8510	10.2554	0.7748	0.7649
	0.25	94.60	2.2821	3.3147	3.6732	6.3915	94.80	2.7293	4.5964	-0.0418	-0.0617
	0.5	94.00	1.7496	2.1347	3.0784	6.6759	94.95	2.1270	2.8882	1.4304	2.6620
	1	94.65	1.5415	1.6710	0.9495	2.3228	95.95	1.8777	2.2532	1.9222	4.0518
	2	95.30	1.5931	1.6497	-0.8839	-2.1762	95.50	1.9072	2.2607	0.5177	1.0886
	4	95.75	1.8028	1.9761	-2.1603	-4.8643	95.50	2.1391	2.9276	-1.1207	-2.0711
	10	94.80	2.1890	2.9224	-2.9892	-5.5366	92.95	2.6509	4.5080	-2.0737	-3.0893
(100,50)	0.1	94.15	3.3417	7.0579	2.9714	3.5382	93.40	3.7816	10.0789	4.6811	4.6666
	0.25	93.90	2.2398	3.2150	2.0749	3.6612	95.15	2.6544	4.5657	-0.0476	-0.0705
	0.5	93.90	1.6418	1.8256	1.2972	3.0367	95.20	1.9840	2.4120	0.4648	0.9462
	1	94.30	1.3351	1.2701	-0.1187	-0.3329	95.70	1.6350	1.6961	0.3714	0.9015
	2	95.00	1.2600	1.0827	-1.0625	-3.2298	95.00	1.5310	1.5293	0.1575	0.4026
	4	94.95	1.3405	1.1450	-1.4465	-4.2779	95.40	1.6220	1.7050	-0.4517	-1.0935
	10	94.80	1.5648	1.5714	-1.2006	-3.0294	94.75	1.9899	2.4280	-1.6091	-3.2664
(100,100)	0.1	93.75	2.3159	3.6661	6.8071	11.3114	94.85	2.8039	5.3400	4.9786	6.8274
	0.25	95.05	1.5886	1.5816	2.6648	6.7139	95.60	1.9191	2.3259	1.3848	2.8719
	0.5	94.35	1.2274	1.0220	1.5027	4.7057	96.05	1.5012	1.4416	0.7943	2.0920
	1	93.60	1.0840	0.8334	0.3527	1.2214	95.70	1.3376	1.1056	0.3308	0.9946
	2	94.10	1.1196	0.8582	-0.5084	-1.7353	95.10	1.3618	1.1719	0.2666	0.7785
	4	94.90	1.2636	1.0448	-1.0116	-3.1303	94.55	1.5352	1.5779	-0.5306	-1.3356
	10	94.40	1.5249	1.5251	-1.1848	-3.0345	94.70	1.9361	2.5530	-0.8669	-1.7155

Table 4.2: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.90	2.8468	7.3108	19.2587	23.1122	95.90	3.4993	9.5310	22.3283	23.4875
	0.25	94.85	3.4871	8.4170	9.8417	10.7869	96.10	4.0871	11.0966	6.6755	6.3483
	0.5	95.20	2.9443	5.8373	2.2194	2.9054	96.55	3.5268	7.9385	-1.3441	-1.5083
	1	94.55	2.0259	2.7683	1.8854	0.3585	96.40	2.4491	3.7583	-0.4842	-0.7896
	2	94.50	1.2448	1.0813	0.8018	2.4384	96.40	1.4994	1.4188	-0.4302	-1.1418
	4	93.50	0.7109	0.3676	0.3476	1.8130	96.15	0.8498	0.4634	-0.2337	-1.0853
	10	93.05	0.3110	0.0772	0.2484	2.8287	94.50	0.3660	0.0886	-0.0629	-0.6681
(30,50)	0.1	94.65	2.3967	4.5698	11.2309	16.8437	96.90	2.9213	5.9675	11.7739	15.4177
	0.25	94.75	2.9388	6.2982	8.8626	11.2350	95.60	3.5121	8.6037	8.8599	9.5933
	0.5	94.95	2.4551	3.9941	5.5740	8.8520	95.95	2.9761	5.6859	3.1369	4.1626
	1	94.50	1.6970	1.9117	3.6318	0.8333	96.05	2.0687	2.6444	1.9458	3.7856
	2	94.25	1.0585	0.7583	2.1329	7.7667	96.10	1.2788	1.0064	0.7789	2.4554
	4	94.15	0.6151	0.2650	1.1837	7.2934	96.05	0.7330	0.3421	0.9608	5.2002
	10	93.90	0.2746	0.0558	0.5079	6.8155	94.30	0.3209	0.0700	0.2472	2.9552
(50,50)	0.1	94.35	2.2438	4.1159	1.1305	17.8968	95.95	2.7543	5.3319	12.0297	16.6985
	0.25	95.30	2.7412	5.2362	5.4529	7.5552	95.50	3.2775	7.0458	3.4476	4.1096
	0.5	95.20	2.2810	3.4744	3.2127	5.4572	96.10	2.7543	4.7705	1.3516	1.9568
	1	94.70	1.5609	1.6349	1.8509	0.4581	96.20	1.9017	2.3371	0.1506	0.3115
	2	94.50	0.9577	0.6273	1.0096	4.0329	96.15	1.1646	0.8605	-0.0210	-0.0716
	4	94.40	0.5476	0.2104	0.5384	3.7140	96.05	0.6633	0.2715	0.2307	1.3998
	10	94.50	0.2413	0.0425	0.2534	3.8867	95.80	0.2921	0.0515	0.1104	1.5377
(100,50)	0.1	94.15	2.1109	3.6001	10.6356	18.0065	94.45	2.5784	4.9891	15.0759	21.8419
	0.25	94.75	2.5847	4.6460	5.0013	7.3554	95.50	3.0638	6.4889	4.5277	5.6282
	0.5	95.20	2.1375	2.9072	2.4976	4.6360	95.65	2.5750	4.1890	0.9420	1.4553
	1	95.30	1.4505	1.3477	0.7235	0.1971	95.80	1.7619	1.9739	-0.2725	-0.6132
	2	94.90	0.8776	0.5078	-0.0290	-0.1288	95.40	1.0701	0.7298	-0.7447	-2.7572
	4	94.60	0.4939	0.1672	-0.2201	-1.7020	96.30	0.6014	0.2298	-0.4417	-2.9142
	10	94.15	0.2137	0.0331	-0.1617	-2.8123	96.20	0.2611	0.0424	-0.1967	-3.0214
(100,100)	0.1	94.15	1.6336	2.0227	8.1027	18.3113	95.70	2.0030	2.9412	8.9145	16.6600
	0.25	95.15	1.9575	2.5608	3.2369	6.4080	96.20	2.3654	3.3878	2.1162	3.6372
	0.5	95.25	1.6094	1.6968	1.7886	4.3452	96.65	1.9631	2.3453	0.6720	1.3873
	1	95.15	1.0989	0.8007	0.9338	0.3301	96.35	1.3506	1.1123	0.0483	0.1448
	2	95.25	0.6741	0.3079	0.4442	2.5315	96.50	0.8269	0.4099	0.1626	0.8030
	4	94.55	0.3858	0.1033	0.2035	2.0025	96.75	0.4719	0.1343	0.1362	1.1752
	10	94.00	0.1706	0.0209	0.0834	1.8264	96.10	0.2111	0.0276	-0.0191	-0.3640

Table 4.2: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.90	0.2847	0.0731	1.9259	23.1122	95.90	0.3499	0.0953	2.2328	23.4875
	0.25	94.85	0.8718	0.5261	2.4604	10.7869	96.10	1.0218	0.6935	1.6689	6.3483
	0.5	95.20	1.4722	1.4593	1.1097	2.9054	96.55	1.7634	1.9846	-0.6720	-1.5083
	1	94.55	2.0259	2.7683	1.8854	0.3585	96.40	2.4491	3.7583	-0.4842	-0.7896
	2	94.50	2.4895	4.3250	1.6035	2.4384	96.40	2.9988	5.6752	-0.8603	-1.1418
	4	93.50	2.8434	5.8811	1.3905	1.8130	96.15	3.3992	7.4140	-0.9347	-1.0853
	10	93.05	3.1102	7.7160	2.4844	2.8287	94.50	3.6596	8.8621	-0.6291	-0.6681
(30,50)	0.1	94.65	0.2397	0.0457	1.1231	16.8437	96.90	0.2921	0.0597	1.1774	15.4177
	0.25	94.75	0.7347	0.3936	2.2157	11.2350	95.60	0.8780	0.5377	2.2150	9.5933
	0.5	94.95	1.2275	0.9985	2.7870	8.8520	95.95	1.4881	1.4215	1.5684	4.1626
	1	94.50	1.6970	1.9117	3.6318	0.8333	96.05	2.0687	2.6444	1.9458	3.7856
	2	94.25	2.1170	3.0333	4.2658	7.7667	96.10	2.5576	4.0258	1.5579	2.4554
	4	94.15	2.4606	4.2347	4.7347	7.2934	96.05	2.9321	5.4743	3.8433	5.2002
	10	93.90	2.7458	5.5756	5.0786	6.8155	94.30	3.2090	7.0004	2.4721	2.9552
(50,50)	0.1	94.35	0.2244	0.0412	1.1305	17.8968	95.95	0.2754	0.0533	1.2030	16.6985
	0.25	95.30	0.6853	0.3273	1.3632	7.5551	95.50	0.8194	0.4404	0.8619	4.1096
	0.5	95.20	1.1405	0.8686	1.6064	5.4572	96.10	1.3772	1.1926	0.6758	1.9568
	1	94.70	1.5609	1.6349	1.8509	0.4581	96.20	1.9017	2.3371	0.1506	0.3115
	2	94.50	1.9155	2.5095	2.0191	4.0329	96.15	2.3292	3.4421	-0.0420	-0.0716
	4	94.40	2.1905	3.3659	2.1538	3.7140	96.05	2.6534	4.3442	0.9227	1.3997
	10	94.50	2.4126	4.2541	2.5337	3.8867	95.80	2.9207	5.1545	1.1042	1.5377
(100,50)	0.1	94.15	0.2111	0.0360	1.0636	18.0065	94.45	0.2578	0.0499	1.5076	21.8419
	0.25	94.75	0.6462	0.2904	1.2503	7.3554	95.50	0.7660	0.4056	1.1319	5.6282
	0.5	95.20	1.0688	0.7268	1.2488	4.6360	95.65	1.2875	1.0473	0.4710	1.4553
	1	95.30	1.4505	1.3477	0.7235	0.1971	95.80	1.7619	1.9739	-0.2725	-0.6132
	2	94.90	1.7552	2.0311	-0.0581	-0.1288	95.40	2.1402	2.9190	-1.4895	-2.7572
	4	94.60	1.9755	2.6747	-0.8803	-1.7020	96.30	2.4055	3.6774	-1.7669	-2.9142
	10	94.15	2.1373	3.3086	-1.6174	-2.8123	96.20	2.6107	4.2385	-1.9666	-3.0214
(100,100)	0.1	94.15	0.1634	0.0202	0.8103	18.3113	95.70	0.2003	0.0294	0.8914	16.6600
	0.25	95.15	0.4894	0.1600	0.8092	6.4080	96.20	0.5913	0.2117	0.5290	3.6372
	0.5	95.25	0.8047	0.4242	0.8943	4.3452	96.65	0.9815	0.5863	0.3360	1.3873
	1	95.15	1.0989	0.8007	0.9338	0.3301	96.35	1.3506	1.1123	0.0483	0.1448
	2	95.25	1.3482	1.2316	0.8883	2.5315	96.50	1.6539	1.6398	0.3253	0.8030
	4	94.55	1.5433	1.6526	0.8141	2.0025	96.75	1.8875	2.1495	0.5450	1.1752
	10	94.00	1.7058	2.0870	0.8344	1.8264	96.10	2.1110	2.7600	-0.1913	-0.3640

Table 4.3: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model a) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.25	4.1792	10.6770	10.2237	9.9405	89.90	4.6597	15.2116	8.5291	6.9303
	0.25	94.10	2.8725	5.1641	4.2258	5.8892	94.20	3.3708	7.5585	-1.2069	-1.3880
	0.5	94.05	2.3088	3.5445	3.0931	5.2012	95.05	2.7555	4.8150	-1.4476	-2.0862
	1	94.90	2.1114	2.9350	-0.6158	-1.1365	95.65	2.5271	3.9813	-1.5255	-2.4178
	2	95.40	2.1860	2.9536	-3.5285	-6.5046	96.00	2.5839	4.2776	-2.8142	-4.3058
	4	94.40	2.5851	4.3833	-4.6929	-7.1044	90.05	2.9618	6.3716	-6.4472	-8.1013
	10	93.90	2.9931	5.9241	-6.0585	-7.8940	78.40	3.0328	9.2158	-8.3300	-8.7078
(30,50)	0.1	94.35	3.1487	5.8763	7.2939	9.5559	93.80	3.7337	8.5761	3.3412	3.6094
	0.25	94.65	2.2583	3.3154	5.3182	9.2735	94.90	2.7040	4.7366	2.4400	3.5467
	0.5	94.30	1.9281	2.5424	5.088	10.14	94.80	2.3022	3.4865	3.899	6.6161
	1	93.95	1.8809	2.4066	2.0681	4.2184	95.20	2.2316	3.1798	2.1271	3.7739
	2	95.75	2.2213	2.8756	-5.0987	-9.549	94.40	2.5779	4.2279	-5.8218	-8.9874
	4	95.25	2.4855	3.6386	-6.4410	-10.7366	88.60	2.8827	5.8323	-8.1300	-10.7038
	10	95.05	2.9286	5.1934	-7.4732	-10.4236	74.35	2.9901	9.2693	-9.9656	-10.4043
(50,50)	0.1	94.10	3.1316	5.9701	2.6260	3.3997	94.65	3.7315	8.4909	2.1382	2.3205
	0.25	94.30	2.1767	3.2235	6.3017	11.1655	93.10	2.5958	4.5934	3.9417	5.8243
	0.5	94.20	1.7747	2.1154	3.1849	6.9395	95.95	2.1519	2.9632	2.1046	3.8681
	1	95.25	1.6204	1.7586	1.1537	2.7515	96.20	1.9550	2.4006	1.5297	3.1228
	2	94.90	1.8321	2.1575	1.6327	3.5162	94.65	2.1979	3.2429	-0.7462	-1.3101
	4	94.85	1.9838	2.5176	1.0801	2.1525	94.10	2.4434	4.0872	-0.6620	-1.0353
	10	94.35	2.2815	3.4233	1.1708	2.0009	88.45	2.8661	5.9648	-1.4309	-1.8526
(100,50)	0.1	95.30	3.0776	5.8075	4.8376	6.3592	93.95	3.6890	8.7252	5.3427	5.7276
	0.25	94.95	2.1204	2.8113	3.3880	6.4013	95.05	2.5452	4.0327	1.5003	2.3626
	0.5	94.35	1.6517	1.8096	1.9713	4.6378	95.60	2.0004	2.4206	0.8636	1.7551
	1	94.50	1.4025	1.3428	0.5935	1.6195	95.80	1.7060	1.8183	1.0196	2.3912
	2	94.95	1.4753	1.4334	0.2272	0.5999	95.55	1.7926	2.1060	0.2268	0.4940
	4	95.20	1.5081	1.5015	-0.5441	-1.4039	95.35	1.8891	2.2492	-1.0366	-2.1857
	10	93.95	1.6499	1.8891	-0.9910	-2.2800	92.50	2.2265	3.2308	-1.8131	-3.1906
(100,100)	0.1	95.15	2.1358	2.8686	3.7093	6.9405	93.95	2.6487	4.4927	0.3994	0.5958
	0.25	95.35	1.5217	1.4368	2.7904	7.3796	96.30	1.8551	2.1650	1.7432	3.7481
	0.5	94.80	1.2448	1.0066	1.6608	5.2406	96.20	1.5171	1.4658	1.1596	3.0293
	1	94.70	1.1392	0.8682	0.5638	1.9132	94.95	1.3875	1.2302	0.6170	1.7591
	2	95.30	1.2837	1.0434	0.1570	0.4860	94.95	1.5733	1.6215	-0.4498	-1.1167
	4	95.45	1.3844	1.2278	-0.4025	-1.1486	95.15	1.7601	2.0245	-0.9132	-2.0295
	10	94.60	1.5803	1.6657	-0.7447	-1.8245	93.60	2.1392	3.1728	-0.5359	-0.9512

Table 4.3: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.70	3.8596	13.2386	20.1747	17.8056	94.05	4.7084	16.8518	21.7887	17.0217
	0.25	94.65	3.7082	9.7960	10.4721	10.6376	95.65	4.3938	12.8023	4.1023	3.6271
	0.5	95.40	2.8189	5.0786	8.4484	11.9362	96.00	3.3911	7.1867	2.2608	2.6672
	1	95.50	1.8866	2.3122	4.9525	1.0352	96.90	2.2865	3.1320	0.9148	1.6345
	2	95.25	1.1728	0.9227	2.5933	8.5666	96.00	1.4156	1.2058	0.8744	2.5182
	4	94.95	0.6130	0.2577	0.8755	5.4600	95.55	0.7376	0.3520	-0.0474	-0.2526
	10	93.95	0.2998	0.0659	0.2417	2.9782	93.80	0.3448	0.0764	-0.2877	-3.2927
(30,50)	0.1	94.00	3.2827	8.4556	13.6829	15.0438	95.75	4.0227	11.2469	12.4046	11.7747
	0.25	95.15	3.0734	6.3697	7.7160	9.7110	96.60	3.7108	8.5282	3.0174	3.2684
	0.5	95.85	2.3396	3.6077	6.3484	10.6262	96.90	2.8435	5.0526	1.9281	2.7128
	1	95.90	1.5873	1.6628	4.0217	0.9908	96.50	1.9274	2.2833	1.5929	3.3345
	2	94.80	0.8717	0.4951	0.4172	1.8750	96.15	1.0554	0.6699	-0.1865	-0.7202
	4	93.95	0.5348	0.1924	0.2271	1.6370	94.85	0.6441	0.2605	-0.1235	-0.7651
	10	92.30	0.2658	0.0509	0.0859	1.2036	90.95	0.2978	0.0616	-0.3914	-4.9908
(50,50)	0.1	93.25	3.0808	7.6646	9.8759	11.3503	95.75	3.7872	9.7429	11.0847	11.2987
	0.25	93.05	2.8859	6.3090	5.9578	7.5201	94.50	3.4606	8.3936	1.9770	2.1579
	0.5	95.60	2.1875	3.1657	3.8271	6.8160	95.65	2.6414	4.4528	0.7651	1.1464
	1	95.30	1.4604	1.4196	2.3976	0.6375	96.60	1.7745	2.0095	0.0031	0.0069
	2	94.80	0.7832	0.4242	0.3910	1.8985	95.65	0.9592	0.5958	-1.0159	-4.1645
	4	94.05	0.4726	0.1593	0.0000	0.0002	95.85	0.5821	0.2189	-0.5470	-3.6988
	10	92.90	0.2306	0.0407	-0.1500	-2.3523	95.20	0.2857	0.0519	-0.4153	-5.7723
(100,50)	0.1	94.65	2.9628	6.9075	12.4737	15.1766	95.10	3.6148	9.9432	15.7800	16.023
	0.25	94.80	2.7494	4.9786	5.6858	8.0825	95.60	3.2905	7.1693	3.2130	3.7964
	0.5	95.25	2.0591	2.6856	2.9363	5.6737	94.95	2.4767	3.8302	0.3422	0.5528
	1	94.60	1.3570	1.1734	1.2764	0.3728	95.75	1.6433	1.6742	-0.4087	-0.9986
	2	94.80	0.7113	0.3243	0.8710	4.8410	95.90	0.8680	0.4519	-0.0317	-0.1491
	4	94.30	0.4190	0.1185	0.3734	3.4310	96.55	0.5173	0.1607	0.0869	0.6852
	10	93.80	0.1989	0.0293	0.0798	1.4739	96.25	0.2515	0.0383	-0.0875	-1.4134
(100,100)	0.1	94.85	2.2528	3.5672	6.8812	11.5955	95.30	2.7870	4.9796	5.7818	8.2190
	0.25	94.70	2.0563	2.8479	3.5870	6.7350	95.85	2.5057	3.8683	2.2107	3.5557
	0.5	95.25	1.5436	1.5659	2.2290	5.6403	96.55	1.8811	2.1309	0.5546	1.2013
	1	95.50	1.0298	0.7026	1.2688	0.4791	95.90	1.2595	0.9786	-0.0426	-0.1360
	2	94.80	0.5505	0.2037	0.6851	4.8044	95.45	0.6764	0.2889	0.2023	1.1902
	4	94.50	0.3308	0.0766	0.3233	3.6943	95.95	0.4139	0.1080	0.0640	0.6158
	10	93.25	0.1608	0.0195	0.0989	2.2397	95.60	0.2069	0.0275	-0.1106	-2.1099

Table 4.3: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.70	0.3860	0.1324	2.0175	17.8056	94.05	0.4708	0.1685	2.1789	17.0217
	0.25	94.65	0.9271	0.6123	2.6180	10.6376	95.65	1.0985	0.8001	1.0256	3.6271
	0.5	95.40	1.4095	1.2697	4.2242	11.9362	96.00	1.6955	1.7967	1.1304	2.6672
	1	95.50	1.8866	2.3122	4.9525	1.0352	96.90	2.2865	3.1320	0.9148	1.6345
	2	95.25	2.3456	3.6908	5.1867	8.5666	96.00	2.8312	4.8233	1.7488	2.5183
	4	94.95	2.4518	4.1238	3.5019	5.4600	95.55	2.9506	5.6316	-0.1896	-0.2526
	10	93.95	2.9985	6.5870	2.4166	2.9782	93.80	3.4480	7.6383	-2.8769	-3.2927
(30,50)	0.1	94.00	0.3283	0.0846	1.3683	15.0438	95.75	0.4023	0.1125	1.2405	11.7747
	0.25	95.15	0.7683	0.3981	1.9290	9.7110	96.60	0.9277	0.5330	0.7544	3.2684
	0.5	95.85	1.1698	0.9019	3.1742	10.6262	96.90	1.4218	1.2632	0.9640	2.7128
	1	95.90	1.5873	1.6628	4.0217	0.9908	96.50	1.9274	2.2833	1.5929	3.3345
	2	94.80	1.7433	1.9804	0.8345	1.8750	96.15	2.1109	2.6797	-0.3729	-0.7202
	4	93.95	2.1393	3.0776	0.9083	1.6370	94.85	2.5764	4.1677	-0.4940	-0.7651
	10	92.30	2.6577	5.0867	0.8585	1.2036	90.95	2.9783	6.1626	-3.9140	-4.9908
(50,50)	0.1	93.25	0.3081	0.0766	0.9876	11.3503	95.75	0.3787	0.0974	1.1085	11.2987
	0.25	93.05	0.7215	0.3943	1.4895	7.5201	94.50	0.8652	0.5246	0.4943	2.1579
	0.5	95.60	1.0937	0.7914	1.9135	6.8160	95.65	1.3207	1.1132	0.3826	1.1464
	1	95.30	1.4604	1.4196	2.3976	0.6375	96.60	1.7745	2.0095	0.0031	0.0069
	2	94.80	1.5665	1.6966	0.7820	1.8985	95.65	1.9184	2.3832	-2.0318	-4.1645
	4	94.05	1.8903	2.5480	0.0001	0.0002	95.85	2.3286	3.5021	-2.1879	-3.6988
	10	92.90	2.3060	4.0692	-1.5005	-2.3523	95.20	2.8573	5.1907	-4.1529	-5.7723
(100,50)	0.1	94.65	0.2963	0.0691	1.2474	15.1766	95.10	0.3615	0.0994	1.5780	16.0230
	0.25	94.80	0.6874	0.3112	1.4215	8.0825	95.60	0.8226	0.4481	0.8032	3.7964
	0.5	95.25	1.0296	0.6714	1.4682	5.6737	94.95	1.2383	0.9576	0.1711	0.5528
	1	94.60	1.3570	1.1734	1.2764	0.3728	95.75	1.6433	1.6742	-0.4087	-0.9986
	2	94.80	1.4226	1.2972	1.7420	4.8410	95.90	1.7360	1.8077	-0.0634	-0.1491
	4	94.30	1.6760	1.8959	1.4934	3.4310	96.55	2.0690	2.5716	0.3475	0.6852
	10	93.80	1.9887	2.9321	0.7982	1.4739	96.25	2.5154	3.8288	-0.8747	-1.4134
(100,100)	0.1	94.85	0.2253	0.0357	0.6881	11.5955	95.30	0.2787	0.0498	0.5782	8.2190
	0.25	94.70	0.5141	0.1780	0.8967	6.7350	95.85	0.6264	0.2418	0.5527	3.5557
	0.5	95.25	0.7718	0.3915	1.1145	5.6403	96.55	0.9405	0.5327	0.2773	1.2013
	1	95.50	1.0298	0.7026	1.2688	0.4791	95.90	1.2595	0.9786	-0.0426	-0.1360
	2	94.80	1.1011	0.8149	1.3703	4.8044	95.45	1.3528	1.1554	0.4046	1.1902
	4	94.50	1.3231	1.2262	1.2931	3.6943	95.95	1.6556	1.7286	0.2561	0.6158
	10	93.25	1.6079	1.9489	0.9887	2.2397	95.60	2.0691	2.7500	-1.1065	-2.1099

Table 4.4: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model a) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.85	3.8519	9.0144	5.7682	6.0851	86.70	4.5131	13.6995	5.2606	4.4979
	0.25	94.45	2.8425	5.1255	7.1739	10.0687	92.50	3.3829	7.6349	4.6116	5.2839
	0.5	94.00	2.4463	4.0675	3.2455	5.0942	94.20	2.8858	5.7846	2.5910	3.4078
	1	93.95	2.3148	3.6324	0.0643	0.1066	94.30	2.6947	4.9971	-0.3607	-0.5101
	2	94.55	2.3794	3.7696	-2.6755	-4.3608	93.00	2.7528	5.1611	-2.4626	-3.4290
	4	94.40	2.5851	4.3833	-4.6929	-7.1044	90.05	2.9618	6.3716	-6.4472	-8.1013
	10	93.90	2.9931	5.9241	-6.0585	-7.8940	78.40	3.0328	9.2158	-8.3300	-8.7078
(30,50)	0.1	93.80	2.8845	5.3486	5.2384	7.1794	93.40	3.5979	8.4467	3.3800	3.6793
	0.25	95.65	2.2717	2.9375	-0.0261	-0.0481	95.90	2.7788	4.5487	-1.1499	-1.7048
	0.5	95.85	2.0642	2.4284	-1.5423	-3.1305	96.55	2.4732	3.6245	-1.4150	-2.3503
	1	95.65	2.0638	2.4572	-3.3861	-6.8452	95.70	2.4185	3.5619	-3.0320	-5.0856
	2	95.75	2.2210	2.8756	-5.0986	-9.5490	94.40	2.5779	4.2279	-5.8218	-8.9874
	4	95.25	2.4855	3.6386	-6.4410	-10.7366	88.60	2.8827	5.8323	-8.1300	-10.7038
	10	95.05	2.9286	5.1934	-7.4732	-10.4236	74.35	2.9901	9.2693	-9.9656	-10.4043
(50,50)	0.1	92.95	2.8326	5.4438	6.6865	9.0977	92.60	3.5542	8.5841	5.8410	6.3153
	0.25	94.30	2.1641	3.1331	5.3850	9.6630	94.30	2.6205	4.6764	4.5696	6.6956
	0.5	94.65	1.8848	2.3489	4.0237	8.3290	94.80	2.2482	3.3764	2.3780	4.0949
	1	94.90	1.7858	2.0806	2.6652	5.8515	95.25	2.1206	3.0291	-0.1698	-0.3084
	2	94.90	1.8321	2.1575	1.6327	3.5162	94.65	2.1979	3.2429	-0.7462	-1.3101
	4	94.85	1.9838	2.5176	1.0801	2.1525	94.10	2.4434	4.0872	-0.6620	-1.0353
	10	94.35	2.2815	3.4233	1.1708	2.0009	88.45	2.8661	5.9648	-1.4309	-1.8526
(100,50)	0.1	93.20	2.8090	5.1967	4.9030	6.8154	93.65	3.5370	8.1815	2.8398	3.1404
	0.25	94.80	2.0792	2.7815	3.8473	7.3124	94.35	2.5756	4.2655	2.5301	3.8758
	0.5	94.65	1.7347	1.9717	2.6193	5.9077	95.15	2.1211	2.9288	2.8205	5.2175
	1	94.50	1.5405	1.5709	1.3242	3.3421	95.35	1.8622	2.2178	1.4915	3.1680
	2	94.95	1.4753	1.4334	0.2272	0.5999	95.55	1.7926	2.1060	0.2268	0.4940
	4	95.20	1.5081	1.5015	-0.5441	-1.4039	95.35	1.8891	2.2492	-1.0366	-2.1857
	10	93.95	1.6499	1.8891	-0.9910	-2.2800	92.50	2.2265	3.2308	-1.8131	-3.1906
(100,100)	0.1	94.55	1.9527	2.4830	3.8693	7.7867	94.70	2.5346	4.0289	3.8180	6.0245
	0.25	95.40	1.5103	1.4148	2.9415	7.8423	96.55	1.8845	2.1887	2.3702	5.0715
	0.5	95.50	1.3215	1.0872	1.9346	5.8759	95.95	1.6151	1.6230	1.8960	4.7104
	1	95.50	1.2529	0.9877	0.9578	3.0483	95.85	1.5177	1.4455	0.5135	1.3503
	2	95.30	1.2837	1.0434	0.1570	0.4860	94.95	1.5733	1.6215	-0.4498	-1.1167
	4	95.45	1.3844	1.2278	-0.4025	-1.1486	95.15	1.7601	2.0245	-0.9132	-2.0295
	10	94.60	1.5803	1.6657	-0.7447	-1.8245	93.60	2.1392	3.1728	-0.5359	-0.9512

Table 4.4: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.00	4.5701	16.514	10.0506	7.8431	91.85	5.4764	22.8020	9.9962	6.6328
	0.25	93.20	3.4902	8.2964	9.5401	10.5291	94.55	4.1660	11.7779	5.0361	4.6443
	0.5	94.25	2.4571	4.1126	5.7765	9.0421	95.75	2.9517	5.5202	1.0326	1.3896
	1	95.10	1.6150	1.7410	3.5211	0.8467	95.15	1.9455	2.4292	-0.1033	-0.2096
	2	95.25	1.0134	0.6871	1.8877	7.2187	96.15	1.2260	0.9594	-0.3184	-1.0276
	4	94.95	0.6130	0.2577	0.8755	5.4600	95.55	0.7376	0.3520	-0.0474	-0.2526
	10	93.95	0.2998	0.0659	0.2417	2.9782	93.80	0.3448	0.0764	-0.2877	-3.2927
(30,50)	0.1	93.80	3.7823	10.9911	10.0032	9.5829	95.45	4.7760	15.1832	7.8364	6.3710
	0.25	94.00	2.8760	5.6875	0.2375	0.3148	96.45	3.5750	8.0883	-4.1704	-4.6410
	0.5	94.8	2.05	2.757	0.5995	1.1416	96.15	2.5161	3.8798	-2.0258	-3.2532
	1	94.90	1.3678	1.2080	0.6080	0.1749	96.00	1.6613	1.7033	-0.9277	-2.2477
	2	94.80	0.8717	0.4951	0.4172	1.8750	96.15	1.0554	0.6699	-0.1865	-0.7202
	4	93.95	0.5348	0.1924	0.2271	1.6370	94.85	0.6441	0.2605	-0.1235	-0.7651
	10	92.30	0.2658	0.0509	0.0859	1.2036	90.95	0.2978	0.0616	-0.3914	-4.9908
(50,50)	0.1	91.90	3.6049	10.9007	11.9065	11.4760	93.80	4.5272	14.7954	10.0795	8.3131
	0.25	93.90	2.6931	5.3822	4.3463	5.9332	95.10	3.3017	7.7110	0.2049	0.2333
	0.5	94.55	1.9007	2.5397	2.4476	4.8612	95.60	2.3238	3.5692	-1.1045	-1.8486
	1	95.30	1.2501	1.0746	1.1723	0.3577	95.50	1.5260	1.5442	-1.6655	-4.2410
	2	94.80	0.7832	0.4242	0.3910	1.8985	95.65	0.9592	0.5958	-1.0159	-4.1645
	4	94.05	0.4726	0.1593	0.0000	0.0002	95.85	0.5821	0.2189	-0.5470	-3.6988
	10	92.90	0.2306	0.0407	-0.1500	-2.3523	95.20	0.2857	0.0519	-0.4153	-5.7723
(100,50)	0.1	92.60	3.4974	9.5568	8.4437	8.6675	94.20	4.3430	13.5996	7.6651	6.5855
	0.25	94.85	2.5851	4.3895	4.7234	7.1457	95.70	3.1559	6.5604	1.9871	2.4534
	0.5	95.10	1.7998	2.0369	2.9796	6.6147	96.25	2.1835	2.9805	0.4911	0.8993
	1	95.35	1.1619	0.8432	1.7194	0.5930	96.10	1.4122	1.1944	0.0860	0.2489
	2	94.80	0.7113	0.3243	0.8710	4.8410	95.90	0.8680	0.4519	-0.0317	-0.1491
	4	94.30	0.4190	0.1185	0.3734	3.4310	96.55	0.5173	0.1607	0.0869	0.6852
	10	93.80	0.1989	0.0293	0.0798	1.4739	96.25	0.2515	0.0383	-0.0875	-1.4134
(100,100)	0.1	93.00	2.5968	5.2110	5.5290	7.6799	94.95	3.3406	7.6338	5.6973	6.5330
	0.25	94.30	1.9103	2.5280	3.3174	6.6106	95.70	2.3857	3.6681	1.6848	2.7822
	0.5	95.10	1.3437	1.2006	2.1669	6.2645	95.85	1.6527	1.6973	0.8582	2.0831
	1	94.95	0.8817	0.5126	1.2893	0.5703	95.85	1.0812	0.7132	0.3199	1.1976
	2	94.80	0.5505	0.2037	0.6851	4.8044	95.45	0.6764	0.2889	0.2023	1.1902
	4	94.50	0.3308	0.0766	0.3233	3.6943	95.95	0.4139	0.1080	0.0640	0.6158
	10	93.25	0.1608	0.0195	0.0989	2.2397	95.60	0.2069	0.0275	-0.1106	-2.1099

Table 4.4: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.00	0.4570	0.1651	1.0051	7.8431	91.85	0.5476	0.2280	0.9996	6.6328
	0.25	93.20	0.8725	0.5185	2.3850	10.5291	94.55	1.0415	0.7361	1.2590	4.6443
	0.5	94.25	1.2285	1.0281	2.8883	9.0421	95.75	1.4759	1.3801	0.5163	1.3896
	1	95.10	1.6150	1.7410	3.5211	0.8467	95.15	1.9455	2.4292	-0.1033	-0.2096
	2	95.25	2.0268	2.7483	3.7755	7.2187	96.15	2.4520	3.8378	-0.6367	-1.0276
	4	94.95	2.4518	4.1238	3.5019	5.4600	95.55	2.9506	5.6316	-0.1896	-0.2526
	10	93.95	2.9985	6.5870	2.4166	2.9782	93.80	3.4480	7.6383	-2.8769	-3.2927
(30,50)	0.1	93.80	0.3782	0.1099	1.0003	9.5829	95.45	0.4776	0.1518	0.7836	6.3710
	0.25	94.00	0.7190	0.3555	0.0594	0.3148	96.45	0.8937	0.5055	-1.0426	-4.6410
	0.5	94.80	1.0250	0.6893	0.2998	1.1416	96.15	1.2580	0.9700	-1.0129	-3.2532
	1	94.90	1.3678	1.2080	0.6080	0.1749	96.00	1.6613	1.7033	-0.9277	-2.2477
	2	94.80	1.7433	1.9804	0.8345	1.8750	96.15	2.1109	2.6797	-0.3729	-0.7202
	4	93.95	2.1393	3.0776	0.9083	1.6370	94.85	2.5764	4.1677	-0.4940	-0.7651
	10	92.30	2.6577	5.0867	0.8585	1.2036	90.95	2.9783	6.1626	-3.9140	-4.9908
(50,50)	0.1	91.90	0.3605	0.1090	1.1907	11.4760	93.80	0.4527	0.1480	1.0080	8.3131
	0.25	93.90	0.6733	0.3364	1.0866	5.9332	95.10	0.8254	0.4819	0.0512	0.2333
	0.5	94.55	0.9503	0.6349	1.2238	4.8612	95.60	1.1619	0.8923	-0.5522	-1.8486
	1	95.30	1.2501	1.0746	1.1723	0.3577	95.50	1.5260	1.5442	-1.6655	-4.2410
	2	94.80	1.5665	1.6966	0.7820	1.8985	95.65	1.9184	2.3832	-2.0318	-4.1645
	4	94.05	1.8903	2.5480	0.0001	0.0002	95.85	2.3286	3.5021	-2.1879	-3.6988
	10	92.90	2.3060	4.0692	-1.5005	-2.3523	95.20	2.8573	5.1907	-4.1529	-5.7723
(100,50)	0.1	92.60	0.3497	0.0956	0.8444	8.6675	94.20	0.4343	0.1360	0.7665	6.5855
	0.25	94.85	0.6463	0.2743	1.1808	7.1457	95.70	0.7890	0.4100	0.4968	2.4534
	0.5	95.10	0.8999	0.5092	1.4898	6.6147	96.25	1.0917	0.7451	0.2455	0.8993
	1	95.35	1.1619	0.8432	1.7194	0.5930	96.10	1.4122	1.1944	0.0860	0.2489
	2	94.80	1.4226	1.2972	1.7420	4.8410	95.90	1.7360	1.8077	-0.0634	-0.1491
	4	94.30	1.6760	1.8959	1.4934	3.4310	96.55	2.0690	2.5716	0.3475	0.6852
	10	93.80	1.9887	2.9321	0.7982	1.4739	96.25	2.5154	3.8288	-0.8747	-1.4134
(100,100)	0.1	93.00	0.2597	0.0521	0.5529	7.6799	94.95	0.3341	0.0763	0.5697	6.5330
	0.25	94.30	0.4776	0.1580	0.8293	6.6106	95.70	0.5964	0.2293	0.4212	2.7822
	0.5	95.10	0.6718	0.3001	1.0835	6.2645	95.85	0.8264	0.4243	0.4291	2.0831
	1	94.95	0.8817	0.5126	1.2893	0.5703	95.85	1.0812	0.7132	0.3199	1.1976
	2	94.80	1.1011	0.8149	1.3703	4.8044	95.45	1.3528	1.1554	0.4046	1.1902
	4	94.50	1.3231	1.2262	1.2931	3.6943	95.95	1.6556	1.7286	0.2561	0.6158
	10	93.25	1.6079	1.9489	0.9887	2.2397	95.60	2.0691	2.7500	-1.1065	-2.1099

Table 4.5: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model b) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	95.15	5.9206	2.1721	1.9022	1.3012	97.30	7.0392	2.6530	1.9952	1.2339
	0.25	94.50	5.0576	1.6497	1.9168	1.5089	96.70	5.9601	2.0522	0.8898	0.6222
	0.5	94.35	4.5354	1.3612	1.2409	1.0694	96.30	5.4157	1.7955	0.6169	0.4608
	1	93.70	4.6356	1.5004	0.8163	0.6678	95.65	5.5523	1.9943	0.6382	0.4523
	2	94.35	5.5192	2.0051	0.1897	0.1340	94.80	6.5018	2.7029	0.3302	0.2008
	4	95.15	7.1961	3.1120	-0.3730	-0.2114	93.55	8.4069	4.4673	0.0775	0.0367
	10	94.80	10.5569	6.0577	-0.2959	-0.1202	86.75	11.6426	9.5532	1.0693	0.3461
(30,50)	0.1	93.40	4.5129	1.4031	1.2691	1.0773	96.55	5.5449	1.7606	1.2183	0.9218
	0.25	95.05	3.9421	0.9618	1.2692	1.3048	96.80	4.7319	1.2866	0.6292	0.5555
	0.5	94.05	3.7193	0.9478	0.6725	0.6922	96.05	4.4912	1.2697	0.2189	0.1942
	1	93.35	4.1008	1.1848	0.1644	0.1510	94.90	4.8965	1.6039	0.0633	0.0500
	2	93.75	5.2010	1.8133	-0.3994	-0.2967	94.20	6.0978	2.4143	-0.5098	-0.3282
	4	94.35	7.0062	3.0589	-0.8879	-0.5082	93.00	8.0588	4.3102	-0.2822	-0.1359
	10	94.80	10.3818	6.0675	-1.4926	-0.6069	86.40	11.6690	9.3550	-0.1766	-0.0577
(50,50)	0.1	94.20	4.4852	1.2757	0.9317	0.8275	97.10	5.5036	1.6880	0.8190	0.6315
	0.25	94.90	3.8237	0.9141	0.8064	0.8463	97.15	4.6109	1.2489	0.3262	0.2919
	0.5	93.95	3.4547	0.7770	0.6809	0.7745	96.65	4.1908	1.0810	0.3277	0.3152
	1	94.00	3.5380	0.8547	0.2881	0.3117	96.10	4.3087	1.1261	0.2592	0.2442
	2	94.80	4.2081	1.1758	-0.1487	-0.1371	96.25	5.0423	1.6262	0.1106	0.0867
	4	95.25	5.4451	1.8745	-0.5285	-0.3862	94.35	6.4672	2.6964	-0.2474	-0.1507
	10	94.60	7.8485	3.8017	-0.8079	-0.4146	92.50	9.5194	6.1588	0.0463	0.0187
(100,50)	0.1	94.60	4.4439	1.2477	1.2434	1.1198	97.00	5.4614	1.5852	1.2758	1.0183
	0.25	94.45	3.7493	0.8726	0.6982	0.7493	96.50	4.5068	1.1621	0.0013	0.0012
	0.5	94.20	3.2528	0.6976	1.0034	1.2098	96.05	3.9347	0.9230	0.7010	0.7314
	1	94.05	3.0694	0.6617	0.7526	0.9289	95.40	3.7454	0.8918	0.5832	0.6186
	2	94.20	3.3233	0.7580	0.4595	0.5284	95.65	4.0577	1.0719	0.5200	0.5028
	4	94.65	4.0242	1.0359	0.1970	0.1936	95.45	4.9273	1.5997	0.5539	0.4382
	10	94.50	5.5260	1.8991	-0.0116	-0.0085	94.05	7.0954	3.3219	0.8566	0.4704
(100,100)	0.1	93.80	3.1088	0.6733	0.4320	0.5271	97.00	3.9432	0.8956	0.3264	0.3450
	0.25	94.80	2.6709	0.4732	0.4205	0.6122	96.90	3.2625	0.6579	0.2392	0.2950
	0.5	93.70	2.4219	0.4130	0.3575	0.5570	95.75	2.9694	0.5618	0.2265	0.3022
	1	94.00	2.4810	0.4445	0.1779	0.2668	95.25	3.0414	0.5857	0.0737	0.0962
	2	94.25	2.9422	0.5918	-0.0636	-0.0826	96.10	3.6141	0.8493	-0.0254	-0.0276
	4	95.00	3.7837	0.9163	-0.3190	-0.3333	95.05	4.6084	1.3587	0.1742	0.1494
	10	95.30	5.3763	1.8061	-0.6188	-0.4609	93.70	6.7834	2.9951	-0.0569	-0.0329

Table 4.5: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	95.15	2.8366	6.7273	18.3102	22.8963	98.10	3.3373	8.3850	19.6158	21.9253
	0.25	95.05	3.5095	8.4569	12.0403	13.2032	97.05	4.0896	10.6256	7.2380	7.0373
	0.5	95.00	2.9528	5.8684	5.4194	7.0904	96.30	3.4984	7.9917	1.9654	2.1986
	1	95.20	2.0279	2.8168	2.7903	5.2633	96.50	2.4246	3.7998	0.6968	1.1301
	2	94.75	1.2451	1.0789	1.5511	4.7264	97.00	1.4867	1.3809	1.1493	3.0935
	4	94.20	0.7106	0.3616	0.8469	4.4573	96.90	0.8511	0.4397	0.6917	3.2993
	10	93.55	0.3111	0.0763	0.7672	8.8168	94.90	0.3669	0.0847	0.4352	4.7341
(30,50)	0.1	94.70	2.3896	4.5540	11.9848	18.0419	97.65	2.7685	5.7755	11.7559	15.6534
	0.25	96.05	2.9343	5.6044	4.3845	5.8653	97.55	3.4637	7.3604	1.0250	1.1945
	0.5	94.90	2.4583	4.0759	2.6210	4.1078	96.75	2.9373	5.5231	-0.9507	-1.2791
	1	94.60	1.7008	1.9737	1.6449	3.7041	96.30	2.0466	2.6377	-0.8808	-1.7149
	2	94.55	1.0595	0.7821	0.9988	3.5729	96.45	1.2672	1.0341	-0.4097	-1.2737
	4	94.25	0.6144	0.2704	0.6039	3.6743	96.30	0.7319	0.3350	0.0446	0.2436
	10	94.10	0.2735	0.0558	0.3166	4.2424	94.20	0.3229	0.0668	0.0597	0.7301
(50,50)	0.1	95.35	2.2250	3.8054	10.2115	16.7809	97.55	2.6097	4.9339	10.6745	15.3716
	0.25	94.75	2.7357	5.2725	5.9884	8.2732	97.10	3.2308	6.9176	3.5600	4.2832
	0.5	95.55	2.2821	3.3454	3.5221	6.0992	97.25	2.7277	4.5262	1.3306	1.9776
	1	94.60	1.5626	1.6143	2.0810	5.1851	97.10	1.8820	2.1434	-0.2492	-0.5382
	2	94.65	0.9586	0.6341	1.1527	4.5814	96.40	1.1552	0.8531	-0.1301	-0.4454
	4	93.40	0.5476	0.2164	0.6268	4.2642	95.40	0.6620	0.2730	0.2019	1.2220
	10	92.60	0.2408	0.0443	0.3014	4.5316	95.40	0.2931	0.0543	0.0044	0.0594
(100,50)	0.1	94.95	2.1094	3.4542	11.5685	20.0713	97.10	2.5086	4.3505	12.9150	19.9621
	0.25	95.25	2.5849	4.4257	5.7665	8.6987	96.95	3.0588	5.8028	2.7755	3.6450
	0.5	95.85	2.1423	2.9218	4.3429	8.0586	96.30	2.5604	4.0294	2.2629	3.5662
	1	95.65	1.4530	1.3377	2.3596	6.4632	96.30	1.7510	1.8447	1.1331	2.6385
	2	95.05	0.8782	0.5006	1.1383	5.0927	95.50	1.0607	0.6883	0.4664	1.7778
	4	94.85	0.4940	0.1648	0.5048	3.9343	95.75	0.5993	0.2252	-0.0569	-0.3790
	10	93.95	0.2138	0.0328	0.1760	3.0721	95.90	0.2615	0.0425	-0.0747	-1.1450
(100,100)	0.1	94.75	1.6044	1.8729	5.5390	12.9021	97.35	1.9094	2.5438	6.0556	12.0908
	0.25	95.50	1.9563	2.4702	2.4380	4.9100	97.05	2.3490	3.4398	1.9658	3.3528
	0.5	94.95	1.6127	1.6452	1.5587	3.8448	97.30	1.9487	2.4074	0.5008	1.0205
	1	95.50	1.1009	0.7792	1.0599	3.7990	96.55	1.3411	1.1178	0.5809	1.7373
	2	95.75	0.6746	0.3017	0.6719	3.8703	96.80	0.8222	0.4110	0.2667	1.3154
	4	95.55	0.3858	0.1020	0.4036	3.9976	96.85	0.4708	0.1339	-0.0537	-0.4642
	10	94.60	0.1703	0.0208	0.2032	4.4578	95.55	0.2099	0.0268	0.0448	0.8642

Table 4.5: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	95.15	0.2837	0.0673	1.8310	22.8963	98.10	0.3337	0.0839	1.9616	21.9253
	0.25	95.05	0.8774	0.5286	3.0101	13.2032	97.05	1.0224	0.6641	1.8095	7.0373
	0.5	95.00	1.4764	1.4671	2.7097	7.0904	96.30	1.7492	1.9979	0.9827	2.1986
	1	95.20	2.0279	2.8168	2.7903	5.2633	96.50	2.4246	3.7998	0.6968	1.1301
	2	94.75	2.4901	4.3157	3.1023	4.7264	97.00	2.9733	5.5237	2.2986	3.0935
	4	94.20	2.8422	5.7850	3.3877	4.4573	96.90	3.4043	7.0359	2.7667	3.2993
	10	93.55	3.1105	7.6267	7.6720	8.8168	94.90	3.6688	8.4658	4.3521	4.7341
(30,50)	0.1	94.70	0.2390	0.0455	1.1985	18.0419	97.65	0.2768	0.0578	1.1756	15.6534
	0.25	96.05	0.7336	0.3503	1.0961	5.8653	97.55	0.8659	0.4600	0.2562	1.1945
	0.5	94.90	1.2292	1.0190	1.3105	4.1078	96.75	1.4687	1.3808	-0.4754	-1.2791
	1	94.60	1.7008	1.9737	1.6449	3.7041	96.30	2.0466	2.6377	-0.8808	-1.7149
	2	94.55	2.1190	3.1283	1.9976	3.5729	96.45	2.5343	4.1365	-0.8193	-1.2737
	4	94.25	2.4577	4.3259	2.4156	3.6743	96.30	2.9276	5.3597	0.1784	0.2436
	10	94.10	2.7345	5.5780	3.1664	4.2424	94.20	3.2287	6.6766	0.5967	0.7301
(50,50)	0.1	95.35	0.2225	0.0381	1.0212	16.7809	97.55	0.2610	0.0493	1.0675	15.3716
	0.25	94.75	0.6839	0.3295	1.4971	8.2732	97.10	0.8077	0.4324	0.8900	4.2832
	0.5	95.55	1.1410	0.8363	1.7610	6.0992	97.25	1.3639	1.1315	0.6653	1.9776
	1	94.60	1.5626	1.6143	2.0810	5.1851	97.10	1.8820	2.1434	-0.2492	-0.5382
	2	94.65	1.9172	2.5364	2.3055	4.5814	96.40	2.3104	3.4126	-0.2602	-0.4454
	4	93.40	2.1905	3.4620	2.5074	4.2642	95.40	2.6478	4.3675	0.8077	1.2220
	10	92.60	2.4084	4.4319	3.0145	4.5316	95.40	2.9311	5.4311	0.0438	0.0594
(100,50)	0.1	94.95	0.2109	0.0345	1.1569	20.0713	97.10	0.2509	0.0435	1.2915	19.9621
	0.25	95.25	0.6462	0.2766	1.4416	8.6987	96.95	0.7647	0.3627	0.6939	3.6450
	0.5	95.85	1.0712	0.7304	2.1715	8.0586	96.30	1.2802	1.0074	1.1314	3.5662
	1	95.65	1.4530	1.3377	2.3596	6.4632	96.30	1.7510	1.8447	1.1331	2.6385
	2	95.05	1.7564	2.0026	2.2766	5.0927	95.50	2.1213	2.7530	0.9329	1.7778
	4	94.85	1.9759	2.6373	2.0194	3.9343	95.75	2.3970	3.6036	-0.2276	-0.3790
	10	93.95	2.1378	3.2842	1.7602	3.0721	95.90	2.6152	4.2537	-0.7469	-1.1450
(100,100)	0.1	94.75	0.1604	0.0187	0.5539	12.9021	97.35	0.1909	0.0254	0.6056	12.0908
	0.25	95.50	0.4891	0.1544	0.6095	4.9100	97.05	0.5872	0.2150	0.4914	3.3528
	0.5	94.95	0.8063	0.4113	0.7794	3.8448	97.30	0.9744	0.6018	0.2504	1.0205
	1	95.50	1.1009	0.7792	1.0599	3.7990	96.55	1.3411	1.1178	0.5809	1.7373
	2	95.75	1.3493	1.2067	1.3437	3.8703	96.80	1.6445	1.6442	0.5334	1.3154
	4	95.55	1.5431	1.6324	1.6143	3.9976	96.85	1.8830	2.1422	-0.2149	-0.4642
	10	94.6	1.7026	2.0814	2.0322	4.4578	95.55	2.0992	2.6842	0.4478	0.8642

Table 4.6: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model b) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	94.05	6.4219	2.6268	1.5784	0.9782	96.45	7.9288	3.4561	1.2219	0.6585
	0.25	94.80	5.4942	1.9333	2.4002	1.7521	96.75	6.6712	2.4460	1.2347	0.7917
	0.5	95.20	5.1083	1.6147	1.5603	1.2370	96.90	6.1138	2.1683	0.9703	0.6602
	1	95.00	5.3076	1.7921	0.9708	0.7269	95.90	6.3199	2.4772	0.8979	0.5712
	2	94.65	6.2027	2.4026	0.2626	0.1694	95.70	7.3083	3.3980	0.6033	0.3274
	4	94.75	7.8120	3.6512	-0.4233	-0.2215	93.25	9.1703	5.5135	0.1470	0.0626
	10	94.85	11.0027	6.9449	-0.9989	-0.3792	80.60	11.4870	12.1310	1.1825	0.3396
(30,50)	0.1	94.85	4.8713	1.5175	1.4324	1.1704	97.70	6.2576	2.0032	0.9904	0.7013
	0.25	95.45	4.3233	1.1664	1.3843	1.2921	97.05	5.3210	1.6757	0.8360	0.6470
	0.5	94.30	4.2223	1.1869	0.7405	0.6811	95.75	5.0861	1.6305	0.4393	0.3442
	1	94.40	4.6977	1.4986	0.2247	0.1835	95.15	5.5825	2.0081	-0.1021	-0.0721
	2	94.05	5.8195	2.2284	-0.3388	-0.2270	94.10	6.8042	3.0448	-0.3499	-0.2005
	4	94.05	7.5817	3.5844	-0.8454	-0.4468	92.55	8.7948	5.1928	-0.4672	-0.2050
	10	94.50	10.8764	6.9363	-1.2714	-0.4832	80.25	11.3732	11.6129	-0.0865	-0.0254
(50,50)	0.1	93.10	4.8112	1.6370	1.1297	0.8862	96.10	6.2364	2.2470	0.9606	0.6420
	0.25	94.80	4.1588	1.0993	1.0197	0.9770	96.70	5.1233	1.5100	0.3085	0.2510
	0.5	94.65	3.8873	0.9478	0.7808	0.8044	96.35	4.7230	1.3355	0.5576	0.4830
	1	94.90	4.0516	1.0610	0.3920	0.3807	96.25	4.8990	1.4863	0.2914	0.2390
	2	95.15	4.7320	1.4368	-0.0416	-0.0347	95.60	5.6643	1.9779	0.1654	0.1176
	4	95.15	5.9284	2.1999	-0.4358	-0.2938	94.05	7.1590	3.3583	-0.2257	-0.1232
	10	94.75	8.2332	4.1941	-0.7692	-0.3758	90.80	10.3637	7.3846	0.4151	0.1527
(100,50)	0.1	93.80	4.7884	1.5509	0.8632	0.6946	97.05	6.2003	2.0595	0.2285	0.1592
	0.25	94.35	4.0343	1.0706	1.3831	1.3485	96.55	4.9752	1.4432	0.9459	0.7896
	0.5	94.30	3.6310	0.8551	1.2019	1.3106	96.00	4.3997	1.1598	0.8301	0.7729
	1	94.45	3.5135	0.8285	0.9630	1.0637	95.80	4.2734	1.1534	0.7776	0.7258
	2	94.60	3.7759	0.9440	0.6724	0.6935	95.55	4.6127	1.3866	0.7269	0.6184
	4	94.75	4.4332	1.2505	0.3931	0.3516	95.40	5.5176	1.9853	0.7108	0.5050
	10	94.70	5.8490	2.1486	0.1320	0.0900	94.35	7.7079	3.9170	1.2674	0.6416
(100,100)	0.1	94.70	3.3454	0.7101	0.2994	0.3554	96.75	4.4343	1.0044	-0.1268	-0.1265
	0.25	94.60	2.9050	0.5433	0.4610	0.6265	96.85	3.5941	0.7730	0.0523	0.0595
	0.5	94.10	2.7236	0.5079	0.3836	0.5389	96.15	3.3283	0.6924	0.2305	0.2770
	1	93.95	2.8379	0.5571	0.2034	0.2725	96.30	3.4830	0.7924	0.0755	0.0848
	2	94.60	3.3081	0.7312	-0.0356	-0.0416	95.65	4.0621	1.0775	0.3416	0.3292
	4	95.00	4.1192	1.0872	-0.2947	-0.2827	94.65	5.0598	1.7002	0.1184	0.0908
	10	95.10	5.6491	2.0105	-0.6139	-0.4333	92.85	7.3073	3.4663	-0.2940	-0.1579

Table 4.6: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	94.15	3.8521	13.0461	14.7597	13.0283	97.00	4.5659	15.6860	13.5710	10.8971
	0.25	94.55	3.7269	9.7550	8.1268	8.2542	96.30	4.4463	12.1673	0.5935	0.5379
	0.5	94.50	2.8321	5.4486	6.7594	9.1936	96.50	3.3731	7.1480	2.4290	2.8734
	1	95.15	1.8905	2.4761	4.5143	9.1073	96.80	2.2498	3.2487	1.4532	2.5498
	2	94.75	1.1715	0.9831	2.7453	8.7872	96.35	1.3915	1.2852	0.8912	2.4860
	4	93.80	0.6883	0.3540	1.5574	8.3045	95.90	0.8199	0.4486	0.5999	2.8329
	10	92.55	0.3182	0.0812	0.7109	7.9118	94.05	0.3674	0.0991	-0.0103	-0.1030
(30,50)	0.1	94.40	3.2812	8.5288	14.0024	15.3356	97.40	3.9268	10.6792	11.0802	10.7815
	0.25	95.00	3.0836	6.4780	6.5158	8.1201	97.25	3.7138	8.4792	1.2880	1.3985
	0.5	94.80	2.3467	3.6873	3.6380	6.0003	96.55	2.8184	4.9467	-0.9487	-1.3486
	1	95.00	1.5902	1.7044	2.4676	5.9861	96.25	1.8967	2.2683	-0.1422	-0.2986
	2	94.45	1.0038	0.6934	1.5090	5.7387	96.70	1.1933	0.8951	0.0279	0.0933
	4	93.75	0.6002	0.2558	0.8642	5.4100	95.90	0.7143	0.3278	-0.1117	-0.6168
	10	92.85	0.2828	0.0601	0.4119	5.3195	93.80	0.3217	0.0776	-0.1579	-1.7928
(50,50)	0.1	93.50	3.0805	7.7642	9.2059	10.5024	96.40	3.7002	9.6987	7.7811	7.9239
	0.25	94.70	2.8999	5.8142	8.2545	10.8867	96.75	3.4997	7.6866	2.7062	3.0874
	0.5	95.30	2.1885	3.0684	4.0166	7.2685	97.50	2.6228	4.1222	1.2681	1.9750
	1	95.10	1.4604	1.3939	2.5934	6.9613	97.35	1.7477	1.8651	0.4580	1.0604
	2	94.60	0.9043	0.5566	1.5190	6.4501	96.70	1.0833	0.7204	0.1194	0.4449
	4	93.95	0.5308	0.2019	0.8312	5.8585	96.20	0.6410	0.2560	0.1428	0.8924
	10	92.75	0.2460	0.0466	0.3668	5.3801	95.75	0.2981	0.0572	-0.1075	-1.4202
(100,50)	0.1	93.95	2.9517	7.1209	8.2759	9.8523	96.75	3.5847	8.9514	5.0699	5.3650
	0.25	94.60	2.7604	5.2812	7.2443	10.0159	97.00	3.3407	6.8911	3.9473	4.7592
	0.5	95.75	2.0644	2.6969	4.3042	8.3149	96.80	2.4790	3.6237	1.4694	2.4411
	1	96.00	1.3581	1.1573	2.6444	7.7950	96.75	1.6255	1.5487	1.0468	2.6604
	2	95.10	0.8254	0.4372	1.4649	7.0209	96.30	0.9905	0.5831	0.5029	2.0824
	4	94.45	0.4753	0.1518	0.7424	6.0355	96.65	0.5760	0.2006	0.2480	1.7507
	10	93.75	0.2157	0.0338	0.2823	4.8591	96.20	0.2654	0.0433	-0.0936	-1.4212
(100,100)	0.1	94.20	2.2503	3.5932	3.0381	5.0735	96.35	2.7593	4.8047	0.7155	1.0321
	0.25	95.45	2.0648	2.7605	3.0601	5.8328	96.85	2.5190	3.8444	0.5363	0.8648
	0.5	95.10	1.5474	1.5086	1.7051	4.3932	97.10	1.8753	2.1237	-0.1271	-0.2756
	1	95.60	1.0311	0.6743	1.2725	4.9052	97.20	1.2432	0.9452	0.5256	1.7094
	2	95.50	0.6372	0.2654	0.8479	5.211	97.10	0.7717	0.3604	0.0362	0.1906
	4	95.15	0.3735	0.0952	0.5192	5.3278	96.95	0.4557	0.1272	0.0947	0.8399
	10	94.30	0.1731	0.0218	0.2554	5.4762	96.00	0.2161	0.0290	0.0474	0.8811

Table 4.6: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	94.15	0.3852	0.1305	1.4760	13.0283	97.00	0.4566	0.1569	1.3571	10.8971
	0.25	94.55	0.9317	0.6097	2.0317	8.2542	96.30	1.1116	0.7605	0.1484	0.5379
	0.5	94.50	1.4161	1.3622	3.3797	9.1936	96.50	1.6866	1.7870	1.2145	2.8734
	1	95.15	1.8905	2.4761	4.5143	9.1073	96.80	2.2498	3.2487	1.4532	2.5498
	2	94.75	2.3430	3.9325	5.4906	8.7872	96.35	2.7830	5.1408	1.7823	2.4860
	4	93.80	2.7534	5.6633	6.2297	8.3045	95.90	3.2797	7.1777	2.3997	2.8329
	10	92.55	3.1817	8.1200	7.1089	7.9118	94.05	3.6737	9.9072	-0.1025	-0.1030
(30,50)	0.1	94.40	0.3281	0.0853	1.4002	15.3356	97.40	0.3927	0.1068	1.1080	10.7815
	0.25	95.00	0.7709	0.4049	1.6289	8.1201	97.25	0.9285	0.5300	0.3220	1.3985
	0.5	94.80	1.1733	0.9218	1.8190	6.0003	96.55	1.4092	1.2367	-0.4743	-1.3486
	1	95.00	1.5902	1.7044	2.4676	5.9861	96.25	1.8967	2.2683	-0.1422	-0.2986
	2	94.45	2.0077	2.7735	3.0180	5.7387	96.70	2.3867	3.5803	0.0558	0.0933
	4	93.75	2.4009	4.0925	3.4567	5.4100	95.90	2.8571	5.2452	-0.4468	-0.6168
	10	92.85	2.8280	6.0089	4.1187	5.3195	93.80	3.2165	7.7560	-1.5790	-1.7928
(50,50)	0.1	93.50	0.3081	0.0776	0.9206	10.5024	96.40	0.3700	0.0970	0.7781	7.9239
	0.25	94.70	0.7250	0.3634	2.0636	10.8867	96.75	0.8749	0.4804	0.6765	3.0874
	0.5	95.30	1.0943	0.7671	2.0083	7.2685	97.50	1.3114	1.0306	0.6341	1.9750
	1	95.10	1.4604	1.3939	2.5934	6.9613	97.35	1.7477	1.8651	0.4580	1.0604
	2	94.60	1.8086	2.2264	3.0379	6.4501	96.70	2.1665	2.8818	0.2389	0.4449
	4	93.95	2.1233	3.2299	3.3247	5.8585	96.20	2.5640	4.0964	0.5713	0.8924
	10	92.75	2.4596	4.6605	3.6685	5.3801	95.75	2.9814	5.7232	-1.0746	-1.4202
(100,50)	0.1	93.95	0.2952	0.0712	0.8276	9.8523	96.75	0.3585	0.0895	0.5070	5.3650
	0.25	94.60	0.6901	0.3301	1.8111	10.0159	97.00	0.8352	0.4307	0.9868	4.7592
	0.5	95.75	1.0322	0.6742	2.1521	8.3149	96.80	1.2395	0.9059	0.7347	2.4411
	1	96.00	1.3581	1.1573	2.6444	7.795	96.75	1.6255	1.5487	1.0468	2.6604
	2	95.10	1.6507	1.749	2.9297	7.0209	96.30	1.9811	2.3323	1.0057	2.0824
	4	94.45	1.9014	2.4284	2.9695	6.0355	96.65	2.3041	3.2092	0.9919	1.7507
	10	93.75	2.1567	3.3808	2.8227	4.8591	96.20	2.6544	4.3322	-0.9356	-1.4212
(100,100)	0.1	94.20	0.2250	0.0359	0.3038	5.0735	96.35	0.2759	0.0480	0.0716	1.0321
	0.25	95.45	0.5162	0.1725	0.7650	5.8328	96.85	0.6298	0.2403	0.1341	0.8648
	0.5	95.10	0.7737	0.3771	0.8526	4.3932	97.10	0.9376	0.5309	-0.0635	-0.2756
	1	95.60	1.0311	0.6743	1.2725	4.9052	97.20	1.2432	0.9452	0.5256	1.7094
	2	95.50	1.2744	1.0615	1.6958	5.2110	97.10	1.5434	1.4417	0.0724	0.1906
	4	95.15	1.4942	1.5230	2.0768	5.3278	96.95	1.8229	2.0347	0.3790	0.8399
	10	94.30	1.7314	2.1799	2.5536	5.4762	96.00	2.1609	2.8958	0.4742	0.8811

Table 4.7: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model b) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-3}$	MSE $\times 10^{-6}$	$Bias$ $\times 10^{-4}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	94.00	7.1393	3.3158	1.9100	1.0545	96.20	9.6602	4.4213	0.6900	0.3283
	0.25	93.95	6.4364	2.7934	2.3390	1.4131	96.45	8.1135	3.5741	1.1776	0.6239
	0.5	95.00	6.2378	2.5279	1.6038	1.0137	96.30	7.5313	3.3989	0.6898	0.3744
	1	94.70	6.5791	2.7491	1.0157	0.6136	95.85	7.8113	3.7682	0.7260	0.3742
	2	94.90	7.5198	3.4511	0.3246	0.1747	94.45	8.8582	4.9652	0.5824	0.2614
	4	95.10	9.0786	4.7981	-0.3792	-0.1731	89.85	10.5390	7.3392	-0.0108	-0.0040
	10	95.02	5.6321	3.6832	-0.2135	-0.1520	91.20	13.4260	4.4632	-0.0106	-0.0023
(30,50)	0.1	93.65	5.4790	2.1632	1.1953	0.8152	95.85	7.6417	3.0356	0.0234	0.0135
	0.25	95.50	5.1115	1.6117	0.9487	0.7492	97.05	6.4286	2.1917	0.2163	0.1461
	0.5	94.95	5.2549	1.7759	1.0725	0.8072	96.50	6.3651	2.4570	0.4535	0.2893
	1	94.70	5.8814	2.1737	0.8435	0.5729	95.50	6.9867	2.9385	0.2697	0.1573
	2	94.60	7.0597	3.1059	0.4717	0.2677	94.25	8.3132	4.4792	0.2760	0.1304
	4	94.75	8.8010	4.7251	0.1169	0.0538	89.00	10.4017	7.3560	-0.3423	-0.1262
	10	94.50	10.5420	6.9372	0.1333	0.0322	87.50	27.3570	75.6489	-0.0265	-0.1035
(50,50)	0.1	93.25	5.3625	2.1218	1.5277	1.0543	95.70	7.5113	2.9221	0.8365	0.4898
	0.25	95.10	4.8523	1.4637	0.7387	0.6116	96.45	6.1842	2.1044	-0.1605	-0.1106
	0.5	94.95	4.7665	1.4979	1.4519	1.1944	96.55	5.8520	2.0064	0.9495	0.6717
	1	95.15	5.0435	1.5898	0.9873	0.7852	96.10	6.1007	2.2032	0.8059	0.5436
	2	95.05	5.7527	2.0424	0.6576	0.4605	95.10	6.9491	2.9780	0.6499	0.3768
	4	95.15	5.9284	2.1999	-0.4358	-0.2938	94.05	7.5642	3.3763	-0.2258	-0.1317
	10	94.75	8.1332	4.1961	-0.7692	-0.3643	90.90	12.3637	7.3846	0.4151	0.1732
(100,50)	0.1	93.45	5.2847	1.9243	1.1974	0.8662	96.35	7.4684	2.7068	0.0852	0.0518
	0.25	94.65	4.6545	1.3838	1.5209	1.3035	96.45	5.9464	1.9447	0.8599	0.6177
	0.5	94.65	4.3848	1.2533	1.5055	1.3568	96.40	5.4076	1.7498	0.9927	0.7524
	1	94.75	4.3525	1.2310	1.2974	1.1728	96.30	5.3209	1.7299	0.8920	0.6796
	2	95.05	4.6328	1.3951	1.0304	0.8755	95.80	5.7331	2.0283	1.0807	0.7609
	4	94.75	4.4332	1.2505	0.3931	0.4573	95.40	6.5176	1.9878	0.7251	0.5483
	10	94.70	7.4910	2.2356	0.1390	0.0754	94.35	6.7079	3.9170	1.2674	0.6416
(100,100)	0.1	93.85	3.7131	0.9860	0.8023	0.8105	94.20	5.2555	1.3834	0.1047	0.0890
	0.25	94.75	3.3801	0.7372	0.5071	0.5915	95.65	4.3101	1.0545	-0.1197	-0.1165
	0.5	94.60	3.3201	0.7422	0.3925	0.4560	96.30	4.1349	1.0613	0.1155	0.1121
	1	94.80	3.5147	0.8243	0.2225	0.2451	95.60	4.3360	1.1849	0.3195	0.2936
	2	94.70	4.0043	1.0517	-0.0031	-0.0030	95.25	4.9386	1.5603	0.2009	0.1608
	4	95.00	4.1568	1.0871	-0.2957	-0.2953	94.75	5.0532	1.7178	0.1184	0.0918
	10	95.10	5.6322	2.0106	-0.6140	-0.4332	92.75	8.3173	3.4534	-0.2968	-0.1779

Table 4.7: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.10	4.5878	16.7575	10.1160	7.8366	96.85	5.7643	20.8798	-1.0305	-0.7130
	0.25	93.95	3.5118	8.4227	3.9966	4.3578	96.75	4.4222	10.5342	-5.8390	-5.6968
	0.5	94.85	2.4836	4.0050	1.8978	2.9994	96.70	3.0313	5.3451	-4.6374	-6.3542
	1	95.65	1.6199	1.6762	3.2841	8.0453	97.05	1.9375	2.2803	-0.5650	-1.1831
	2	95.85	1.0113	0.6595	2.1665	8.4646	97.40	1.1986	0.8941	-0.0152	-0.0507
	4	95.65	0.6091	0.2461	1.2976	8.2981	96.40	0.7171	0.3229	0.0785	0.4367
	10	95.55	0.4021	0.2213	1.1085	6.3452	95.40	0.5232	0.1229	0.0563	0.2526
(30,50)	0.1	93.60	3.8012	11.6344	4.5322	4.2045	95.20	4.9375	14.8852	-6.6495	-5.4569
	0.25	94.55	2.8755	5.5330	2.3275	3.1298	96.50	3.6438	7.4564	-5.5419	-6.4296
	0.5	94.80	2.0422	2.7735	3.3172	6.3098	96.75	2.5060	3.6455	-2.0830	-3.4512
	1	95.20	1.3595	1.2475	2.7740	7.8763	96.45	1.6304	1.6509	-0.2443	-0.6012
	2	94.90	0.8660	0.5063	1.8553	8.2713	96.35	1.0251	0.6860	0.2202	0.8407
	4	94.45	0.5314	0.1935	1.1398	8.2200	94.90	0.6255	0.2597	0.5229	3.2457
	10	94.15	0.4232	0.1563	1.0932	8.1145	93.95	0.4214	0.2232	0.7316	5.6703
(50,50)	0.1	92.15	3.6243	11.2616	8.5566	8.0873	95.00	4.7679	14.0912	0.9518	0.8016
	0.25	94.35	2.7298	5.0552	0.2584	0.3633	96.30	3.4871	7.0331	-7.9062	-9.4673
	0.5	94.15	1.9060	2.5547	3.4018	6.7440	96.55	2.3571	3.3692	-0.7895	-1.3599
	1	94.85	1.2503	1.0150	2.1762	6.8448	96.85	1.5076	1.4097	-0.9153	-2.4378
	2	94.75	0.7801	0.4034	1.4102	7.0368	96.80	0.9341	0.5587	-0.1107	-0.4682
	4	93.95	0.5426	0.2347	0.8562	6.5435	96.20	0.6205	0.2560	-0.1428	-0.8924
	10	92.75	0.3427	0.0364	0.4368	5.3801	95.75	0.3814	0.0572	-0.1075	-1.5342
(100,50)	0.1	92.30	3.5042	9.8591	5.9051	5.9562	95.20	4.6648	12.5041	-3.4928	-3.1243
	0.25	94.40	2.5996	4.4162	4.5947	6.9289	97.15	3.3414	5.9307	-0.9405	-1.2210
	0.5	95.95	1.8050	2.0648	3.2393	7.1451	97.00	2.2332	2.7589	-0.6212	-1.1824
	1	95.85	1.1628	0.8360	2.2345	7.7491	97.10	1.4056	1.1285	-0.1883	-0.5605
	2	95.75	0.7107	0.3148	1.3918	7.8666	96.90	0.8546	0.4298	0.0609	0.2939
	4	94.45	0.4563	0.1237	0.6874	6.0355	96.65	0.6761	0.2003	0.2489	1.9341
	10	93.85	0.2157	0.0342	0.2947	4.8591	96.20	0.2835	0.0432	-0.0938	-1.4212
(100,100)	0.1	92.75	2.6029	5.2323	3.3219	4.5961	93.60	3.5167	6.8235	-3.3176	-4.0185
	0.25	94.65	1.9215	2.5103	1.6577	3.3095	95.85	2.4826	3.5106	-4.3815	-7.4134
	0.5	94.75	1.3495	1.1613	1.3018	3.8217	96.95	1.6740	1.6402	-0.7758	-1.9154
	1	95.90	0.8834	0.4904	1.0844	4.9015	96.80	1.0707	0.6913	-0.4482	-1.7044
	2	95.95	0.5505	0.1927	0.7798	5.6243	97.20	0.6641	0.2681	-0.0558	-0.3407
	4	95.25	0.3734	0.0952	0.5653	5.3289	96.95	0.4587	0.1273	0.0947	0.8467
	10	94.30	0.1831	0.0219	0.2674	5.6222	96.10	0.2187	0.0290	0.0474	0.8824

Table 4.7: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.10	0.4588	0.1676	1.0116	7.8366	96.85	0.5764	0.2088	-0.1030	-0.7130
	0.25	93.95	0.8780	0.5264	0.9991	4.3578	96.75	1.1055	0.6584	-1.4597	-5.6968
	0.5	94.85	1.2418	1.0013	0.9489	2.9994	96.70	1.5157	1.3363	-2.3187	-6.3542
	1	95.65	1.6199	1.6762	3.2841	8.0453	97.05	1.9375	2.2803	-0.5650	-1.1831
	2	95.85	2.0226	2.6379	4.3330	8.4646	97.40	2.3971	3.5764	-0.0303	-0.0507
	4	95.65	2.4365	3.9374	5.1905	8.2981	96.40	2.8682	5.1656	0.3139	0.4367
	10	95.55	1.9846	2.5769	3.2143	7.5621	95.40	3.2158	3.4217	0.2431	0.2543
(30,50)	0.1	93.60	0.3801	0.1163	0.4532	4.2045	95.20	0.4938	0.1489	-0.6650	-5.4569
	0.25	94.55	0.7189	0.3458	0.5819	3.1298	96.50	0.9109	0.4660	-1.3855	-6.4296
	0.5	94.80	1.0211	0.6934	1.6586	6.3098	96.75	1.2530	0.9114	-1.0415	-3.4512
	1	95.20	1.3595	1.2475	2.7740	7.8763	96.45	1.6304	1.6509	-0.2443	-0.6012
	2	94.90	1.7320	2.0253	3.7106	8.2713	96.35	2.0502	2.7438	0.4405	0.8407
	4	94.45	2.1256	3.0958	4.5593	8.2200	94.90	2.5021	4.1551	2.0916	3.2457
	10	94.15	4.3218	4.0322	5.5432	8.1145	93.95	2.7543	6.2432	3.4290	5.6703
(50,50)	0.1	92.15	0.3624	0.1126	0.8557	8.0873	95.00	0.4768	0.1409	0.0952	0.8016
	0.25	94.35	0.6825	0.3160	0.0646	0.3633	96.30	0.8718	0.4396	-1.9766	-9.4673
	0.5	94.15	0.9530	0.6387	1.7009	6.7440	96.55	1.1786	0.8423	-0.3947	-1.3599
	1	94.85	1.2503	1.0150	2.1762	6.8448	96.85	1.5076	1.4097	-0.9153	-2.4378
	2	94.75	1.5602	1.6137	2.8204	7.0368	96.80	1.8682	2.2350	-0.2214	-0.4682
	4	93.95	2.1453	3.2299	3.4210	5.3422	96.20	2.1327	4.4322	0.4753	0.8924
	10	92.75	2.4753	4.3322	3.6532	4.3761	95.75	2.9763	6.4326	-1.0765	-1.5218
(100,50)	0.1	92.30	0.3504	0.0986	0.5905	5.9562	95.20	0.4665	0.1250	-0.3493	-3.1243
	0.25	94.40	0.6499	0.2760	1.1487	6.9289	97.15	0.8353	0.3707	-0.2351	-1.2210
	0.5	95.95	0.9025	0.5162	1.6197	7.1451	97.00	1.1166	0.6897	-0.3106	-1.1824
	1	95.85	1.1628	0.8360	2.2345	7.7491	97.10	1.4056	1.1285	-0.1883	-0.5605
	2	95.75	1.4215	1.2592	2.7835	7.8666	96.90	1.7091	1.7191	0.1219	0.2939
	4	94.45	1.8733	2.2135	3.2157	6.0355	96.65	2.4672	3.5722	0.9920	1.8901
	10	93.85	2.1567	3.4521	2.8227	4.8678	96.40	2.6678	4.3572	-0.9378	-1.4338
(100,100)	0.1	92.75	0.2603	0.0523	0.3322	4.5961	93.60	0.3517	0.0682	-0.3318	-4.0185
	0.25	94.65	0.4804	0.1569	0.4144	3.3095	95.85	0.6206	0.2194	-1.0954	-7.4134
	0.5	94.75	0.6747	0.2903	0.6509	3.8217	96.95	0.8370	0.4101	-0.3879	-1.9154
	1	95.90	0.8834	0.4904	1.0844	4.9015	96.80	1.0707	0.6913	-0.4482	-1.7044
	2	95.95	1.1011	0.7709	1.5595	5.6243	97.20	1.3282	1.0726	-0.1116	-0.3407
	4	95.25	1.4945	1.5283	2.0765	5.3678	96.95	1.8242	2.0356	0.3776	0.8399
	10	94.30	1.7314	3.1799	2.4162	5.4782	96.10	2.1641	2.8556	0.4743	0.8813

Table 4.8: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model c) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^0$	$\times 10^0$	$\times 10^{-2}$	$\times 10^{-2}$	%	$\times 10^0$	$\times 10^0$	$\times 10^{-2}$	$\times 10^{-2}$
(30,30)	0.1	94.00	3.6689	0.8534	11.4142	12.4484	94.95	4.2085	1.0865	11.5181	11.1154
	0.25	95.05	3.1071	0.5841	8.1067	10.6646	95.60	3.6296	0.7775	4.0736	4.6236
	0.5	94.50	2.7509	0.5225	9.0079	12.5569	95.50	3.2743	0.7003	7.4020	8.8777
	1	93.60	2.7613	0.5258	4.6471	6.4203	95.95	3.2769	0.6745	4.7694	5.8159
	2	94.25	3.2180	0.6724	1.2763	1.5563	95.00	3.7630	0.9220	2.5527	2.6587
	4	94.50	4.0931	1.0092	-1.1778	-1.1722	93.40	4.7146	1.4363	1.9269	1.6077
	10	94.95	5.7850	1.8980	-2.8995	-2.1046	86.40	6.3840	3.0014	3.4495	1.9910
(30,50)	0.1	94.30	2.8211	0.5023	6.1004	8.6374	95.95	3.3546	0.6738	7.2606	8.8781
	0.25	94.75	2.4330	0.3917	6.3312	10.1657	96.40	2.8975	0.5234	4.2696	5.9102
	0.5	93.35	2.2728	0.3651	7.7919	13.0002	95.30	2.7415	0.4890	6.0587	8.6943
	1	93.30	2.4540	0.4198	4.0915	6.3256	95.75	2.9482	0.5694	4.0265	5.3422
	2	94.75	3.0357	0.5846	0.1620	0.2118	95.15	3.5478	0.8022	2.1525	2.4034
	4	95.85	3.9736	0.9156	-3.1815	-3.3260	93.25	4.5349	1.3188	0.3034	0.2641
	10	95.00	5.6525	1.7667	-6.6879	-5.0367	86.00	6.2026	2.8576	0.8824	0.5219
(50,50)	0.1	94.20	2.7891	0.5197	6.8014	9.4748	95.45	3.3240	0.6767	6.5699	8.0102
	0.25	94.20	2.3649	0.3600	5.7582	9.6391	95.75	2.8295	0.5109	3.3404	4.6772
	0.5	94.30	2.1075	0.3114	5.3924	9.7064	95.65	2.5569	0.4140	3.7450	5.8290
	1	94.70	2.1136	0.3148	2.9889	5.3338	96.35	2.5752	0.4247	4.2967	6.6061
	2	95.25	2.4540	0.3906	0.3773	0.6036	95.75	2.9351	0.5373	2.7261	3.7206
	4	95.70	3.0927	0.5739	-1.8115	-2.3914	95.10	3.6624	0.8589	0.7425	0.8010
	10	94.80	4.2848	1.0835	-3.4065	-3.2736	92.80	5.1575	1.6864	0.3667	0.2823
(100,50)	0.1	94.15	2.7879	0.5206	4.3016	5.9711	96.50	3.3084	0.7049	5.0775	6.0573
	0.25	94.15	2.3188	0.3491	4.1463	7.0332	95.55	2.7749	0.4799	1.9069	2.7530
	0.5	94.05	1.9728	0.2645	2.9814	5.8056	95.75	2.3855	0.3434	2.0787	3.5487
	1	94.35	1.8269	0.2377	1.1040	2.2646	96.05	2.2364	0.3154	1.7958	3.1983
	2	94.90	1.9364	0.2553	-0.4855	-0.9607	95.00	2.3547	0.3620	1.5779	2.6227
	4	95.05	2.2910	0.3318	-1.3993	-2.4293	95.20	2.7709	0.4969	0.7533	1.0684
	10	94.85	3.0417	0.5814	-1.1891	-1.5592	94.55	3.8706	0.9152	-0.4595	-0.4802
(100,100)	0.1	94.95	1.9462	0.2502	3.8031	7.6236	96.80	2.4062	0.3583	3.1637	5.2912
	0.25	95.05	1.6513	0.1726	3.6583	8.8369	96.35	2.0017	0.2478	2.6369	5.3039
	0.5	94.45	1.4755	0.1477	2.6586	6.932	96.00	1.8021	0.2070	1.8628	4.0967
	1	93.70	1.4829	0.1560	1.3544	3.4300	95.75	1.8263	0.2064	1.2832	2.8251
	2	93.95	1.7207	0.2028	0.1270	0.2821	95.00	2.0914	0.2778	1.4344	2.7219
	4	94.90	2.1606	0.3038	-0.7457	-1.3526	94.25	2.6200	0.4614	0.4950	0.7285
	10	94.30	2.9644	0.5671	-1.0824	-1.4371	94.40	3.7603	0.9612	0.6394	0.6521

Table 4.8: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	95.30	2.8394	6.6260	18.5236	2.3363	97.30	3.4005	8.6094	20.0156	2.2086
	0.25	94.80	3.4835	8.2782	6.8283	0.7524	96.90	4.0518	10.5018	4.1048	0.4008
	0.5	94.45	2.9418	6.1916	3.5008	0.4452	96.45	3.4892	7.9430	0.3155	0.0354
	1	94.50	2.0291	2.8111	1.5633	0.2949	96.35	2.4365	3.7643	-0.7752	-0.1263
	2	94.30	1.2456	1.0970	0.7154	0.2160	96.45	1.4940	1.4329	-0.4359	-0.1151
	4	93.35	0.7105	0.3718	0.3595	0.1865	96.10	0.8497	0.4665	-0.2723	-0.1261
	10	93.45	0.3098	0.0773	0.0795	0.0905	94.40	0.3681	0.0917	-0.0982	-0.1025
(30,50)	0.1	94.80	2.3740	4.3957	10.3546	1.5808	97.20	2.8080	5.9796	11.4808	1.5010
	0.25	95.20	2.9315	5.8748	5.8676	0.7676	96.65	3.4636	7.7271	4.7717	0.5435
	0.5	95.00	2.4572	3.9864	5.6341	0.8957	96.40	2.9539	5.5925	3.1255	0.4182
	1	94.45	1.6987	1.9132	3.7988	0.8716	96.30	2.0591	2.6282	2.0393	0.3980
	2	94.30	1.0589	0.7598	2.3012	0.8375	96.00	1.2758	1.0061	0.8966	0.2827
	4	94.05	0.6148	0.2652	1.3104	0.8071	96.05	0.7342	0.3422	1.0450	0.5656
	10	93.85	0.2741	0.0558	0.5775	0.7752	94.05	0.3220	0.0699	0.2952	0.3533
(50,50)	0.1	94.65	2.2229	4.0212	11.1690	1.7889	96.55	2.6464	5.2360	11.7673	1.6478
	0.25	95.10	2.7440	5.0598	7.9405	1.1230	96.35	3.2486	6.7958	6.7280	0.8187
	0.5	95.30	2.2830	3.4668	3.2907	0.5596	96.35	2.7344	4.7033	1.4686	0.2141
	1	94.80	1.5628	1.6353	2.0028	0.4957	96.30	1.8913	2.3179	0.2836	0.0589
	2	94.45	0.9584	0.6284	1.1544	0.4609	96.40	1.1615	0.8574	0.0794	0.0271
	4	94.30	0.5475	0.2108	0.6452	0.4448	95.95	0.6630	0.2712	0.3171	0.1925
	10	94.50	0.2409	0.0426	0.3114	0.4775	95.55	0.2929	0.0516	0.1505	0.2095
(100,50)	0.1	94.80	2.0803	3.3797	8.4059	1.4609	96.70	2.5036	4.5683	10.4741	1.5682
	0.25	94.50	2.582	4.5950	5.2351	0.7744	96.05	3.0443	6.2989	3.2763	0.4131
	0.5	95.20	2.1389	2.9037	2.5854	0.4802	95.75	2.5592	4.1166	1.0626	0.1656
	1	94.95	1.4524	1.3498	0.8441	0.2298	95.90	1.7519	1.9564	-0.1109	-0.0251
	2	94.75	0.8784	0.5093	0.0770	0.0341	95.60	1.0670	0.7242	-0.6110	-0.2270
	4	94.50	0.4940	0.1678	-0.1440	-0.1111	96.25	0.6006	0.2293	-0.3710	-0.2450
	10	94.05	0.2135	0.0332	-0.1206	-0.2092	96.25	0.2615	0.0423	-0.1653	-0.2541
(100,100)	0.1	94.70	1.6046	1.9039	6.2074	1.4369	96.35	1.9447	2.6639	7.1151	1.3915
	0.25	95.30	1.9536	2.5449	3.2014	0.6357	96.90	2.3436	3.3022	1.8902	0.3290
	0.5	95.25	1.6104	1.6949	1.8376	0.4467	97.20	1.9538	2.3150	0.7110	0.1478
	1	95.30	1.1001	0.8014	1.0190	0.3601	96.40	1.3450	1.1059	0.1509	0.0454
	2	95.25	0.6746	0.3084	0.5235	0.2982	96.65	0.8252	0.4086	0.2375	0.1175
	4	94.50	0.3858	0.1035	0.2614	0.2570	96.95	0.4716	0.1343	0.1810	0.1562
	10	94.20	0.1704	0.0209	0.1146	0.2507	96.05	0.2116	0.0276	0.0022	0.0043

Table 4.8: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	95.30	0.2839	0.0663	1.8524	2.3363	97.30	0.3400	0.0861	2.0016	2.2086
	0.25	94.80	0.8709	0.5174	1.7071	0.7524	96.90	1.0129	0.6564	1.0262	0.4008
	0.5	94.45	1.4709	1.5479	1.7504	0.4452	96.45	1.7446	1.9858	0.1577	0.0354
	1	94.5	2.0291	2.8111	1.5633	0.2949	96.35	2.4365	3.7643	-0.7752	-0.1263
	2	94.30	2.4911	4.3880	1.4308	0.2160	96.45	2.9880	5.7315	-0.8718	-0.1151
	4	93.35	2.8419	5.9487	1.4382	0.1865	96.10	3.3990	7.4636	-1.0892	-0.1261
	10	93.45	3.0982	7.7290	0.7954	0.0905	94.40	3.6815	9.1741	-0.9821	-0.1025
(30,50)	0.1	94.80	0.2374	0.0440	1.0355	1.5808	97.20	0.2808	0.0598	1.1481	1.5010
	0.25	95.20	0.7329	0.3672	1.4669	0.7676	96.65	0.8659	0.4829	1.1929	0.5435
	0.5	95.00	1.2286	0.9966	2.8171	0.8957	96.40	1.4769	1.3981	1.5627	0.4182
	1	94.45	1.6987	1.9132	3.7988	0.8716	96.30	2.0591	2.6282	2.0393	0.3980
	2	94.30	2.1178	3.0393	4.6024	0.8375	96.00	2.5515	4.0245	1.7933	0.2827
	4	94.05	2.4591	4.2435	5.2417	0.8071	96.05	2.9370	5.4758	4.1799	0.5656
	10	93.85	2.7406	5.5805	5.7750	0.7752	94.05	3.2198	6.9897	2.9525	0.3533
(50,50)	0.1	94.65	0.2223	0.0402	1.1169	1.7889	96.55	0.2646	0.0524	1.1767	1.6478
	0.25	95.10	0.6860	0.3162	1.9851	1.1230	96.35	0.8121	0.4247	1.6820	0.8187
	0.5	95.30	1.1415	0.8667	1.6453	0.5596	96.35	1.3672	1.1758	0.7343	0.2141
	1	94.80	1.5628	1.6353	2.0028	0.4957	96.30	1.8913	2.3179	0.2836	0.0589
	2	94.45	1.9169	2.5134	2.3089	0.4609	96.40	2.3229	3.4298	0.1588	0.0271
	4	94.30	2.1902	3.3723	2.5810	0.4448	95.95	2.6520	4.3395	1.2682	0.1925
	10	94.50	2.4093	4.2598	3.1136	0.4775	95.55	2.9293	5.1590	1.5046	0.2095
(100,50)	0.1	94.80	0.2080	0.0338	0.8406	1.4609	96.70	0.2504	0.0457	1.0474	1.5682
	0.25	94.50	0.6455	0.2872	1.3088	0.7744	96.05	0.7611	0.3937	0.8191	0.4131
	0.5	95.20	1.0695	0.7259	1.2927	0.4802	95.75	1.2796	1.0291	0.5313	0.1656
	1	94.95	1.4524	1.3498	0.8441	0.2298	95.90	1.7519	1.9564	-0.1109	-0.0251
	2	94.75	1.7569	2.0373	0.1540	0.0341	95.60	2.1340	2.8969	-1.2219	-0.2270
	4	94.50	1.9760	2.6845	-0.5759	-0.1111	96.25	2.4022	3.6695	-1.4842	-0.2450
	10	94.05	2.1353	3.3197	-1.2056	-0.2092	96.25	2.6149	4.2308	-1.6526	-0.2541
(100,100)	0.1	94.70	0.1605	0.0190	0.6207	1.4369	96.35	0.1945	0.0266	0.7115	1.3915
	0.25	95.30	0.4884	0.1591	0.8004	0.6357	96.90	0.5859	0.2064	0.4725	0.3290
	0.5	95.25	0.8052	0.4237	0.9188	0.4467	97.20	0.9769	0.5787	0.3555	0.1478
	1	95.30	1.1001	0.8014	1.0190	0.3601	96.40	1.3450	1.1059	0.1509	0.0454
	2	95.25	1.3491	1.2337	1.0471	0.2982	96.65	1.6505	1.6344	0.4750	0.1175
	4	94.50	1.5433	1.6552	1.0454	0.2570	96.95	1.8865	2.1483	0.7239	0.1562
	10	94.20	1.7038	2.0884	1.1457	0.2507	96.05	2.1161	2.7560	0.0223	0.0043

Table 4.9: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model c) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^0$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^0$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	94.65	3.9334	0.9876	11.6724	11.8244	94.95	4.6245	1.3794	11.8254	10.1177
	0.25	94.85	3.3495	0.7036	11.1649	13.4269	95.70	3.9692	0.9766	7.1379	7.2400
	0.5	94.10	3.0498	0.6236	6.5908	8.3730	95.35	3.6030	0.8267	4.3896	4.8322
	1	93.90	3.1174	0.6600	3.0169	3.7152	94.85	3.6664	0.8863	2.2836	2.4257
	2	93.75	3.5782	0.8236	-0.1321	-0.1455	94.10	4.1626	1.1712	-0.1081	-0.0999
	4	94.75	4.4169	1.1603	-2.0433	-1.8967	93.30	5.0907	1.7242	-1.9205	-1.4624
	10	95.25	6.0277	2.0288	-1.5816	-1.1102	83.75	6.2420	3.2838	-0.8724	-0.4813
(30,50)	0.1	94.05	3.0220	0.5914	6.3721	8.3123	96.45	3.7307	0.8310	3.4565	3.7935
	0.25	94.25	2.6345	0.4534	9.0337	13.5344	95.85	3.1748	0.6408	7.6281	9.5701
	0.5	94.50	2.5501	0.4282	7.1509	10.9912	95.80	3.0464	0.5811	5.2057	6.8429
	1	93.95	2.7812	0.5072	3.4679	4.8738	95.95	3.3194	0.6871	2.7916	3.3687
	2	94.15	3.3644	0.7062	-0.1401	-0.1667	94.75	3.9462	0.9949	2.0203	2.0254
	4	94.40	4.2680	1.0736	-3.0512	-2.9452	93.05	4.8852	1.6160	2.1487	1.6901
	10	94.45	5.8796	1.9566	-5.0693	-3.6255	81.55	5.9943	3.2192	-4.6682	-2.6021
(50,50)	0.1	92.65	2.9880	0.6329	10.1891	12.9103	95.80	3.6826	0.8605	8.7488	9.4712
	0.25	94.25	2.5438	0.4230	7.4924	11.5942	96.25	3.0894	0.5921	4.9661	6.4656
	0.5	94.55	2.3385	0.3689	5.9081	9.7715	96.90	2.8366	0.5060	4.7462	6.6853
	1	95.25	2.3858	0.3809	3.4598	5.6135	96.30	2.8778	0.5185	4.3056	5.9885
	2	95.60	2.7234	0.4687	0.8358	1.2206	95.70	3.2419	0.6811	3.3493	4.0607
	4	95.60	3.3271	0.6640	-1.4293	-1.7538	95.15	3.9863	1.0315	-0.6226	-0.6129
	10	95.00	4.4502	1.1816	-3.2240	-2.9664	90.80	5.5370	1.9445	-0.5105	-0.3660
(100,50)	0.1	94.20	2.9698	0.5910	8.3287	10.8954	96.25	3.6752	0.8329	8.6919	9.5648
	0.25	95.00	2.4590	0.3858	5.6865	9.1915	96.00	2.9748	0.5332	3.5705	4.8944
	0.5	94.25	2.1712	0.3138	4.0876	7.3151	95.90	2.6314	0.4118	2.7016	4.2130
	1	94.45	2.0611	0.2894	2.2315	4.1510	96.05	2.5088	0.3903	3.0402	4.8712
	2	95.10	2.1694	0.3107	0.5854	1.0501	96.15	2.6247	0.4494	2.1638	3.2288
	4	95.05	2.4917	0.3915	-0.4885	-0.7806	95.70	3.0499	0.5983	0.2648	0.3422
	10	94.60	3.1850	0.6436	-0.5974	-0.7445	94.60	4.1626	1.0240	-0.7147	-0.7061
(100,100)	0.1	92.90	2.0789	0.2970	3.4768	6.3913	96.10	2.6437	0.4251	2.8464	4.3689
	0.25	95.35	1.7697	0.1962	4.1274	9.3554	96.85	2.1670	0.2882	3.0243	5.6414
	0.5	95.05	1.6363	0.1740	3.0482	7.3259	96.60	1.9940	0.2512	2.4323	4.8575
	1	94.70	1.6729	0.1870	1.7334	4.0109	95.20	2.0343	0.2641	1.9098	3.7179
	2	94.70	1.9097	0.2398	0.4945	1.0097	95.20	2.3137	0.3482	0.9489	1.6077
	4	95.30	2.3238	0.3466	-0.4298	-0.7298	94.55	2.8649	0.5384	-0.2021	-0.2753
	10	94.40	3.0830	0.6136	-0.8868	-1.1319	93.70	3.9951	1.0970	3.2899	3.1419

Table 4.9: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	93.85	3.8543	12.8279	18.2027	1.6279	96.60	4.6175	15.9836	20.0020	1.6019
	0.25	94.10	3.7215	9.8807	11.0679	1.1201	96.40	4.3841	13.0709	6.0919	0.5335
	0.5	95.70	2.8245	5.0317	4.7071	0.6649	96.60	3.3643	6.6937	-0.7377	-0.0901
	1	95.60	1.8924	2.2885	2.3024	0.4817	96.60	2.2641	3.1914	-1.6045	-0.2841
	2	95.15	1.1765	0.9126	0.9089	0.3009	96.40	1.4057	1.2637	-0.8271	-0.2327
	4	94.15	0.6926	0.3304	0.2536	0.1395	96.60	0.8265	0.4252	-0.2604	-0.1262
	10	93.60	0.3202	0.0761	0.0029	0.0033	94.80	0.3710	0.0933	-0.2775	-0.2874
(30,50)	0.1	93.90	3.2537	8.4137	8.8517	0.9693	96.70	3.9445	10.5409	5.6064	0.5467
	0.25	94.55	3.0782	6.6743	8.3551	1.0278	96.45	3.6925	9.0276	4.2636	0.4491
	0.5	95.65	2.3366	3.6361	8.4813	1.4203	97.00	2.8129	4.9422	4.8930	0.6975
	1	95.45	1.5848	1.6683	5.4352	1.3423	96.45	1.9079	2.2792	2.7835	0.5839
	2	95.40	1.0023	0.6783	3.1339	1.2118	96.30	1.2016	0.9264	1.4168	0.4659
	4	94.80	0.6007	0.2509	1.6807	1.0668	95.70	0.7159	0.3334	0.9463	0.5189
	10	93.80	0.2838	0.0592	0.7086	0.9248	93.50	0.3223	0.0783	0.4396	0.4973
(50,50)	0.1	93.60	3.1022	7.8248	13.1236	1.4998	96.40	3.7484	10.4006	12.1377	1.1984
	0.25	95.15	2.9024	5.7563	5.5021	0.7269	96.65	3.4737	7.7138	2.7392	0.3120
	0.5	95.65	2.1890	3.1586	3.7991	0.6774	96.15	2.6242	4.3755	0.7059	0.1067
	1	95.25	1.4607	1.4160	2.4992	0.6655	97.15	1.7614	1.9805	0.1243	0.0279
	2	94.85	0.9045	0.5587	1.4906	0.6317	97.05	1.0934	0.7587	0.3642	0.1322
	4	94.25	0.5310	0.2007	0.8273	0.5848	96.20	0.6453	0.2628	0.4444	0.2742
	10	93.65	0.2460	0.0460	0.3688	0.5446	95.65	0.3001	0.0556	0.0726	0.0974
(100,50)	0.1	94.15	2.9661	6.9132	11.8462	1.4391	96.70	3.6065	9.4185	13.5593	1.4107
	0.25	94.80	2.7518	4.9761	5.5656	0.7913	96.30	3.2893	6.9439	2.6169	0.3141
	0.5	95.25	2.0604	2.6825	2.9729	0.5748	95.65	2.4657	3.7384	0.3023	0.0494
	1	94.60	1.3574	1.1721	1.3816	0.4038	96.10	1.6305	1.6428	-0.2985	-0.0736
	2	94.70	0.8259	0.4489	0.4940	0.2332	96.30	0.9985	0.6117	-0.2567	-0.1038
	4	94.40	0.4760	0.1568	0.0910	0.0726	96.40	0.5804	0.2102	-0.1190	-0.0820
	10	93.45	0.2158	0.0349	-0.0523	-0.0884	96.45	0.2667	0.0434	-0.2001	-0.3037
(100,100)	0.1	93.00	2.2491	3.9357	5.4134	0.8659	96.60	2.7734	5.3152	5.0203	0.6901
	0.25	94.65	2.0574	2.8475	3.4559	0.6488	96.60	2.4962	3.7793	1.7933	0.2918
	0.5	95.30	1.5442	1.5640	2.2270	0.5639	97.15	1.8711	2.0948	0.4922	0.1075
	1	95.55	1.0299	0.7013	1.3310	0.5031	96.20	1.2513	0.9655	0.0145	0.0047
	2	95.00	0.6373	0.2758	0.7182	0.4327	96.65	0.7774	0.3680	0.2027	0.1056
	4	94.40	0.3739	0.0986	0.3585	0.3612	96.50	0.4589	0.1304	0.2563	0.2245
	10	93.30	0.1734	0.0224	0.1388	0.2932	95.80	0.2180	0.0288	-0.0098	-0.0183

Table 4.9: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	93.85	0.3854	0.1283	1.8203	1.6279	96.60	0.4617	0.1598	2.0002	1.6019
	0.25	94.10	0.9304	0.6175	2.7670	1.1201	96.40	1.0960	0.8169	1.5230	0.5335
	0.5	95.70	1.4123	1.2579	2.3535	0.6649	96.60	1.6821	1.6734	-0.3689	-0.0901
	1	95.60	1.8924	2.2885	2.3024	0.4817	96.60	2.2641	3.1914	-1.6045	-0.2841
	2	95.15	2.3530	3.6502	1.8177	0.3009	96.40	2.8113	5.0548	-1.6542	-0.2327
	4	94.15	2.7703	5.2860	1.0143	0.1395	96.60	3.3061	6.8028	-1.0415	-0.1262
	10	93.60	3.2016	7.6090	0.0290	0.0033	94.80	3.7100	9.3259	-2.7753	-0.2874
(30,50)	0.1	93.90	0.3254	0.0841	0.8852	0.9693	96.70	0.3944	0.1054	0.5606	0.5467
	0.25	94.55	0.7696	0.4171	2.0888	1.0278	96.45	0.9231	0.5642	1.0659	0.4491
	0.5	95.65	1.1683	0.9090	4.2407	1.4203	97.00	1.4064	1.2355	2.4465	0.6975
	1	95.45	1.5848	1.6683	5.4352	1.3423	96.45	1.9079	2.2792	2.7835	0.5839
	2	95.40	2.0045	2.7131	6.2677	1.2118	96.30	2.4032	3.7056	2.8336	0.4659
	4	94.80	2.4027	4.0143	6.7229	1.0668	95.70	2.8637	5.3337	3.7852	0.5189
	10	93.80	2.8376	5.9182	7.0858	0.9248	93.50	3.2234	7.8315	4.3961	0.4973
(50,50)	0.1	93.60	0.3102	0.0782	1.3124	1.4998	96.40	0.3748	0.1040	1.2138	1.1984
	0.25	95.15	0.7256	0.3598	1.3755	0.7269	96.65	0.8684	0.4821	0.6848	0.3120
	0.5	95.65	1.0945	0.7897	1.8995	0.6774	96.15	1.3121	1.0939	0.3529	0.1067
	1	95.25	1.4607	1.4160	2.4992	0.6655	97.15	1.7614	1.9805	0.1243	0.0279
	2	94.85	0.9045	0.5587	1.4906	0.6317	97.05	1.0934	0.7587	0.3642	0.1322
	4	94.25	2.1241	3.2117	3.3094	0.5848	96.20	2.5811	4.2040	1.7775	0.2742
	10	93.65	2.4598	4.5982	3.6882	0.5446	95.65	3.0013	5.5554	0.7263	0.0974
(100,50)	0.1	94.15	0.2966	0.0691	1.1846	1.4391	96.70	0.3607	0.0942	1.3559	1.4107
	0.25	94.80	0.6879	0.3110	1.3914	0.7913	96.30	0.8223	0.4340	0.6542	0.3141
	0.5	95.25	1.0302	0.6706	1.4864	0.5748	95.65	1.2329	0.9346	0.1511	0.0494
	1	94.60	1.3574	1.1721	1.3816	0.4038	96.10	1.6305	1.6428	-0.2985	-0.0736
	2	94.70	1.6518	1.7955	0.9879	0.2332	96.30	1.9969	2.447	-0.5134	-0.1038
	4	94.40	1.9038	2.5091	0.3639	0.0726	96.40	2.3216	3.3634	-0.4759	-0.0820
	10	93.45	2.1585	3.4885	-0.5225	-0.0884	96.45	2.6672	4.3406	-2.0005	-0.3037
(100,100)	0.1	93.00	0.2249	0.0394	0.5413	0.8659	96.60	0.2773	0.0532	0.5020	0.6901
	0.25	94.65	0.5143	0.1780	0.8640	0.6488	96.60	0.6241	0.2362	0.4483	0.2918
	0.5	95.30	0.7721	0.3910	1.1135	0.5639	97.15	0.9355	0.5237	0.2461	0.1075
	1	95.55	1.0299	0.7013	1.3310	0.5031	96.20	1.2513	0.9655	0.0145	0.0047
	2	95.00	1.2746	1.1033	1.4364	0.4327	96.65	1.5547	1.4719	0.4054	0.1056
	4	94.40	1.4957	1.5772	1.4340	0.3612	96.50	1.8356	2.0860	1.0253	0.2245
	10	93.30	1.7342	2.2411	1.3878	0.2932	95.80	2.1799	2.8770	-0.0982	-0.0183

Table 4.10: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model c) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^0$	$\times 10^0$	$\times 10^{-2}$	$\times 10^{-2}$	%	$\times 10^0$	$\times 10^0$	$\times 10^{-2}$	$\times 10^{-2}$
(30,30)	0.1	93.80	4.3188	1.2252	12.9990	11.8224	94.25	5.4300	1.8188	13.6793	10.1931
	0.25	94.40	3.8056	0.9498	13.3598	13.8356	95.50	4.6355	1.3178	8.6063	7.5164
	0.5	94.15	3.6339	0.8639	7.1974	7.7651	95.60	4.3337	1.1675	3.9041	3.6146
	1	94.15	3.7721	0.9361	2.8956	2.9934	95.75	4.4499	1.2504	1.1516	1.0296
	2	94.30	4.2299	1.1566	-1.1753	-1.0927	94.40	4.9893	1.6082	-0.9513	-0.7500
	4	94.50	5.0018	1.5706	-4.3565	-3.4775	88.15	5.7858	2.4284	-0.9712	-0.6231
	10	94.30	6.4678	2.5545	-6.2082	-3.8862	72.70	5.9746	4.3632	5.6886	2.7237
(30,50)	0.1	92.60	3.3395	0.8494	13.9283	15.2847	95.70	4.3298	1.2347	15.1964	13.8022
	0.25	95.10	3.0460	0.5789	8.5750	11.3396	96.55	3.7549	0.8615	7.8688	8.5063
	0.5	94.75	3.0660	0.5914	5.6703	7.3915	95.50	3.6521	0.8148	6.0823	6.7518
	1	94.35	3.3616	0.7162	2.1173	2.5021	94.80	3.9401	0.9673	1.6610	1.6886
	2	94.60	3.9440	0.9787	-1.4855	-1.5014	93.05	4.6110	1.3737	-3.2147	-2.7432
	4	95.00	4.7999	1.4274	-4.6282	-3.8757	87.60	5.5794	2.2040	-4.9711	-3.3495
	10	95.10	6.3047	2.4408	-7.3285	-4.6948	71.05	5.9130	4.1627	-1.7342	-0.8498
(50,50)	0.1	93.60	3.2602	0.7444	8.9997	10.4853	96.05	4.2413	1.0699	8.3815	8.1278
	0.25	93.05	2.8989	0.5923	9.9302	13.0083	95.15	3.5801	0.8532	8.9596	9.7432
	0.5	93.70	2.7917	0.5488	8.1657	11.0878	94.30	3.3620	0.7885	5.8441	6.5941
	1	94.15	2.8996	0.5844	6.1167	8.0248	94.55	3.4687	0.8345	3.8197	4.1839
	2	94.60	3.2414	0.7076	4.2695	5.0810	94.20	3.9038	1.0494	3.4708	3.3892
	4	94.70	3.8092	0.9440	3.0413	3.1310	93.80	4.6919	1.5319	3.1382	2.5357
	10	94.25	4.8643	1.5120	2.9734	2.4182	86.00	6.0489	2.7524	5.1511	3.1056
(100,50)	0.1	93.60	3.2195	0.7291	9.4626	11.1476	95.10	4.2120	1.1437	11.6429	10.9493
	0.25	94.35	2.7743	0.5057	6.1306	8.6507	95.85	3.4627	0.7563	3.9862	4.5875
	0.5	94.50	2.5586	0.4312	4.5338	6.9191	95.75	3.1339	0.6236	4.3737	5.5455
	1	94.25	2.4906	0.4096	2.6039	4.0708	95.65	3.0174	0.5768	2.4548	3.2332
	2	94.65	2.5975	0.4418	0.7155	1.0763	96.05	3.1695	0.6386	1.2700	1.5889
	4	94.95	2.879	0.5412	-0.8101	-1.1010	95.15	3.6167	0.8027	-0.4284	-0.4780
	10	94.05	3.4885	0.8250	-1.8135	-1.9966	91.65	4.7163	1.4027	-1.4758	-1.2458
(100,100)	0.1	94.45	2.2648	0.3452	5.6507	9.6605	97.50	3.0085	0.5295	5.3714	7.4003
	0.25	95.40	2.0223	0.2571	4.8585	9.6234	97.55	2.5324	0.3846	4.1425	6.6929
	0.5	95.35	1.9524	0.2384	3.7687	7.7404	96.60	2.3865	0.3499	3.6572	6.1933
	1	95.45	2.0278	0.2588	2.5131	4.9447	95.85	2.4566	0.3749	2.1158	3.4566
	2	95.30	2.2630	0.3233	1.3210	2.3231	94.90	2.7762	0.5001	0.5626	0.7954
	4	95.35	2.6472	0.4449	0.3867	0.5796	94.95	3.3686	0.7384	0.3110	0.3618
	10	94.60	3.3461	0.7313	-0.1823	-0.2131	93.05	4.5246	1.4173	2.4230	2.0351

Table 4.10: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	92.55	4.5499	18.0359	10.5568	0.7883	95.10	5.5371	22.8315	9.2624	0.6140
	0.25	94.15	3.4874	8.1048	9.4301	1.0530	96.85	4.1926	11.0684	1.6069	0.1527
	0.5	94.70	2.4610	3.8333	5.5606	0.9015	96.90	2.9468	5.1851	-0.3778	-0.0525
	1	94.90	1.6168	1.6489	3.3901	0.8376	96.65	1.9315	2.2650	-0.4394	-0.0923
	2	95.20	1.0132	0.6625	1.8221	0.7095	96.40	1.2137	0.9209	-0.4628	-0.1525
	4	94.75	0.6125	0.2525	0.8562	0.5395	95.10	0.7323	0.3430	-0.2935	-0.1585
	10	93.0	0.2993	0.0653	0.2531	0.3132	92.90	0.3400	0.0745	-0.5027	-0.5832
(30,50)	0.1	92.45	3.7775	11.9921	11.0728	1.0161	96.35	4.7901	15.6138	9.1788	0.7364
	0.25	94.10	2.8549	5.5788	7.3209	0.9846	96.40	3.5200	7.7585	2.0941	0.2377
	0.5	94.90	2.0331	2.7049	5.5342	1.0699	96.90	2.4797	3.7425	2.0879	0.3414
	1	95.50	1.3569	1.1924	3.7688	1.0977	96.70	1.6384	1.5874	1.9073	0.4791
	2	95.40	0.8661	0.4922	2.3181	1.0504	96.40	1.0419	0.6528	1.3146	0.5151
	4	94.20	0.5324	0.1922	1.3154	0.9529	94.50	0.6379	0.2517	0.6523	0.4114
	10	93.30	0.2650	0.0510	0.5763	0.8097	90.40	0.2922	0.0594	0.0128	0.0165
(50,50)	0.1	92.70	3.6186	10.4845	7.7415	0.7580	96.45	4.5678	13.6697	6.0749	0.5202
	0.25	93.10	2.6976	5.2564	5.0505	0.6981	96.05	3.3286	7.3576	1.4880	0.1735
	0.5	94.30	1.9015	2.4663	3.2274	0.6511	96.40	2.3097	3.4652	0.6349	0.1078
	1	94.95	1.2490	1.0379	1.8642	0.5795	96.25	1.5074	1.4669	0.4703	0.1228
	2	95.15	0.7820	0.4071	0.9274	0.4600	96.50	0.9477	0.5564	0.1489	0.0631
	4	95.20	0.4713	0.1518	0.3773	0.3063	96.75	0.5794	0.2012	-0.1750	-0.1234
	10	94.15	0.2299	0.0385	0.0611	0.0984	95.35	0.2844	0.0505	-0.3512	-0.4948
(100,50)	0.1	92.70	3.4878	9.7003	9.2318	0.9412	95.90	4.4141	13.7193	9.4413	0.8085
	0.25	94.45	2.5902	4.4604	3.3858	0.5075	96.35	3.1798	6.4181	-0.5202	-0.0649
	0.5	95.00	1.8034	2.0599	2.1233	0.4682	97.00	2.1817	2.9071	-1.0647	-0.1975
	1	95.25	1.1638	0.8487	1.2196	0.4189	96.50	1.4046	1.1703	-0.5527	-0.1615
	2	94.55	0.7122	0.3252	0.5987	0.3321	96.00	0.8608	0.4457	-0.3399	-0.1610
	4	94.45	0.4194	0.1185	0.2328	0.2139	96.15	0.5127	0.1597	-0.0625	-0.0494
	10	93.65	0.1990	0.0293	0.0250	0.0461	96.45	0.2497	0.0381	-0.1457	-0.2361
(100,100)	0.1	92.95	2.5991	5.2319	5.0145	0.6948	96.40	3.3799	7.3031	3.8075	0.4459
	0.25	94.25	1.9119	2.5362	3.0939	0.6154	96.50	2.3830	3.5531	0.9457	0.1586
	0.5	95.05	1.3445	1.2022	2.1020	0.6072	96.45	1.6471	1.6541	0.5571	0.1370
	1	95.10	0.8819	0.5120	1.3094	0.5795	96.15	1.0719	0.6999	0.2702	0.1021
	2	95.00	0.5504	0.2029	0.7359	0.5172	95.65	0.6714	0.2851	0.2147	0.1271
	4	94.45	0.3306	0.0761	0.3751	0.4303	96.25	0.4109	0.1070	0.0943	0.0912
	10	93.45	0.1607	0.0193	0.1359	0.3096	95.85	0.2058	0.0274	-0.0885	-0.1691

Table 4.10: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	92.55	0.4550	0.1804	1.0557	0.7883	95.10	0.5537	0.2283	0.9262	0.6140
	0.25	94.15	0.8719	0.5066	2.3575	1.0530	96.85	1.0481	0.6918	0.4017	0.1527
	0.5	94.70	1.2305	0.9583	2.7803	0.9015	96.90	1.4734	1.2963	-0.1889	-0.0525
	1	94.90	1.6168	1.6489	3.3901	0.8376	96.65	1.9315	2.2650	-0.4394	-0.0923
	2	95.20	2.0265	2.6502	3.6442	0.7095	96.40	2.4274	3.6835	-0.9256	-0.1525
	4	94.75	2.4500	4.0399	3.4247	0.5395	95.10	2.9292	5.4874	-1.1739	-0.1585
	10	93.40	2.9929	6.5344	2.5313	0.3132	92.90	3.400	7.4526	-5.0272	-0.5832
(30,50)	0.1	92.45	0.3777	0.1199	1.1073	1.0161	96.35	0.4790	0.1561	0.9179	0.7364
	0.25	94.10	0.7137	0.3487	1.8302	0.9846	96.40	0.8800	0.4849	0.5235	0.2377
	0.5	94.90	1.0166	0.6762	2.7671	1.0699	96.90	1.2399	0.9356	1.0440	0.3414
	1	95.50	1.3569	1.1924	3.7688	1.0977	96.70	1.6384	1.5874	1.9073	0.4791
	2	95.40	1.7322	1.9687	4.6362	1.0504	96.40	2.0838	2.6113	2.6291	0.5150
	4	94.20	2.1296	3.0752	5.2617	0.9529	94.50	2.5517	4.0269	2.6094	0.4114
	10	93.30	2.6502	5.0966	5.7633	0.8097	90.40	2.9225	5.9399	0.1275	0.1654
(50,50)	0.1	92.70	0.3619	0.1048	0.7742	0.7580	96.45	0.4568	0.1367	0.6075	0.5202
	0.25	93.10	0.6744	0.3285	1.2626	0.6981	96.05	0.8322	0.4599	0.372	0.1735
	0.5	94.30	0.9508	0.6166	1.6137	0.6511	96.40	1.1549	0.8663	0.3174	0.1078
	1	94.95	1.2490	1.0379	1.8642	0.5795	96.25	1.5074	1.4669	0.4703	0.1228
	2	95.15	1.5640	1.6286	1.8548	0.4600	96.50	1.8954	2.2255	0.2977	0.0631
	4	95.20	1.8853	2.4291	1.5092	0.3063	96.75	2.3175	3.2188	-0.7000	-0.1234
	10	94.15	2.2988	3.8507	0.6110	0.0984	95.35	2.8435	5.0487	-3.5121	-0.4948
(100,50)	0.1	92.7	0.3488	0.097	0.9232	0.9412	95.90	0.4414	0.1372	0.9441	0.8085
	0.25	94.45	0.6475	0.2788	0.8464	0.5075	96.35	0.7949	0.4011	-0.13	-0.0649
	0.5	95.00	0.9017	0.5150	1.0617	0.4682	97.00	1.0909	0.7268	-0.5324	-0.1975
	1	95.25	1.1638	0.8487	1.2196	0.4189	96.50	1.4046	1.1703	-0.5527	-0.1615
	2	94.55	1.4245	1.3007	1.1973	0.3321	96.00	1.7216	1.7827	-0.6799	-0.1610
	4	94.45	1.6775	1.8958	0.9313	0.2139	96.15	2.0507	2.5548	-0.2499	-0.0494
	10	93.65	1.9896	2.9270	0.2497	0.0461	96.45	2.4975	3.8097	-1.4572	-0.2361
(100,100)	0.1	92.95	0.2599	0.0523	0.5015	0.6948	96.40	0.3380	0.0730	0.3808	0.4459
	0.25	94.25	0.478	0.1585	0.7735	0.6154	96.5	0.5957	0.2221	0.2364	0.1586
	0.5	95.05	0.6722	0.3006	1.0510	0.6072	96.45	0.8235	0.4135	0.2786	0.1370
	1	95.10	0.8819	0.5120	1.3094	0.5795	96.15	1.0719	0.6999	0.2702	0.1021
	2	95.00	1.1008	0.8115	1.4719	0.5172	95.65	1.3428	1.1404	0.4294	0.1271
	4	94.45	1.3224	1.2175	1.5004	0.4303	96.25	1.6438	1.7118	0.3772	0.0912
	10	93.45	1.6065	1.9284	1.3592	0.3096	95.85	2.0584	2.7390	-0.8853	-0.1691

Table 4.11: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model d) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	94.05	1.5372	1.6058	9.4764	24.3322	94.80	1.9060	2.1193	7.7193	17.0047
	0.25	93.85	1.5745	1.7702	7.8194	18.9099	95.65	1.8779	2.3056	4.8791	10.2115
	0.5	93.70	1.5937	1.8231	4.3511	10.2413	96.05	1.9269	2.3354	4.2192	8.7619
	1	93.35	1.7544	2.2073	-0.6307	-1.3422	95.90	2.1247	2.9450	3.0740	5.6722
	2	94.35	2.2077	3.2107	-5.3166	-9.4221	95.50	2.5880	4.5203	2.3169	3.4472
	4	95.10	3.0339	5.3955	-7.6201	-10.4277	93.70	3.4061	7.6095	-0.9223	-1.0571
	10	95.00	4.7316	12.4858	-3.0962	-2.7712	87.65	5.0456	17.4423	-4.2780	-3.2401
(30,50)	0.1	93.15	1.1963	1.0291	7.9172	25.4610	96.45	1.5147	1.3938	4.9608	13.4032
	0.25	94.30	1.2430	1.0449	5.7578	18.0973	96.30	1.5101	1.3714	3.0649	8.3024
	0.5	93.00	1.3110	1.2509	1.4199	4.0167	94.85	1.6043	1.6996	0.5676	1.3767
	1	92.80	1.5544	1.7560	-3.0489	-7.2934	94.85	1.8894	2.4282	0.2743	0.5566
	2	94.30	2.0845	2.8680	-7.0056	-13.1917	94.10	2.4581	3.9264	-1.4736	-2.3517
	4	94.70	2.9598	5.1330	-8.5875	-12.0703	92.35	3.3108	7.4612	-1.1471	-1.3278
	10	95.55	4.6984	11.9102	1.9610	1.7967	89.05	5.0690	16.5076	0.8760	0.6816
(50,50)	0.1	94.50	1.1916	0.9775	7.3452	24.1636	96.30	1.4954	1.3028	4.3179	12.0462
	0.25	94.40	1.2057	0.9750	5.7495	18.7288	95.55	1.4663	1.3310	2.9962	8.2384
	0.5	94.00	1.2183	1.0381	2.8922	9.0107	96.30	1.5004	1.3924	2.7886	7.4924
	1	93.80	1.3412	1.2620	-0.8728	-2.4570	96.15	1.6591	1.7406	2.2715	5.4513
	2	94.60	1.6813	1.8362	-4.1339	-9.6901	95.30	2.0251	2.6513	2.6570	5.1657
	4	95.10	2.2884	3.1108	-5.2671	-9.4835	94.95	2.6655	4.5800	0.9693	1.4320
	10	95.70	3.4804	6.5509	-1.3603	-1.6805	94.00	3.9721	9.8031	-1.8936	-1.9124
(100,50)	0.1	94.00	1.1779	0.9819	6.8236	22.3062	95.65	1.4804	1.3439	4.3761	12.0203
	0.25	94.70	1.1745	0.9340	4.1717	13.7757	96.00	1.4263	1.2767	1.8068	5.0617
	0.5	93.75	1.1411	0.9345	1.3387	4.3820	95.50	1.4027	1.3019	1.3051	3.6185
	1	93.55	1.1606	0.9987	-2.2321	-7.0790	94.55	1.4283	1.3783	0.9105	2.4528
	2	93.95	1.3215	1.2197	-5.2600	-15.2310	94.85	1.5999	1.7524	-0.8485	-2.0268
	4	94.95	1.6763	1.7609	-6.3194	-15.2293	95.50	2.0029	2.6188	-2.4708	-4.8326
	10	94.90	2.4248	3.5355	-2.8147	-4.7379	95.20	2.9581	5.2254	-1.5909	-2.2009
(100,100)	0.1	93.85	0.8345	0.5016	4.4952	20.4821	96.30	1.0618	0.7070	1.4643	5.5138
	0.25	93.80	0.8482	0.5224	3.5284	15.6211	95.85	1.0439	0.7236	1.2765	4.7496
	0.5	93.05	0.8577	0.5610	1.5581	6.5913	95.65	1.0636	0.7456	0.8444	3.0932
	1	93.05	0.9442	0.6702	-1.1326	-4.3779	95.25	1.1793	0.9093	0.6520	2.1621
	2	94.55	1.1803	0.9405	-3.4012	-11.1564	95.80	1.4478	1.3502	0.7211	1.9624
	4	95.60	1.5921	1.5432	-3.9670	-10.1480	95.35	1.9083	2.4055	1.0655	2.1723
	10	95.65	2.3795	3.3977	-0.3943	-0.6763	93.55	2.8692	5.2700	0.6821	0.9393

Table 4.11: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	94.30	3.1420	8.0165	13.0043	1.4676	96.30	3.6988	9.5089	11.9286	1.2322
	0.25	94.65	3.2349	7.4276	10.4461	1.2208	96.75	3.7731	8.9335	4.8493	0.5136
	0.5	94.50	2.7238	5.1535	9.8189	1.3804	96.50	3.2476	6.4820	3.7288	0.4635
	1	94.50	1.9797	2.6810	6.4191	1.2490	96.10	2.3834	3.5082	2.2190	0.3748
	2	95.00	1.2874	1.1325	3.3870	1.0114	96.10	1.5424	1.5051	1.7571	0.4533
	4	94.40	0.7561	0.4005	1.4933	0.7481	95.60	0.8958	0.5206	1.7076	0.7503
	10	94.55	0.3201	0.0733	-0.1862	-0.2175	94.40	0.3827	0.0929	0.3876	0.4023
(30,50)	0.1	93.95	2.7463	5.9836	12.9363	1.6958	96.75	3.2437	7.1503	9.8106	1.1678
	0.25	94.95	2.7936	5.3346	9.2663	1.2787	97.15	3.3118	6.8679	3.7333	0.4508
	0.5	94.70	2.3558	3.8189	5.8843	0.9563	96.60	2.8357	5.0614	-1.1058	-0.1554
	1	95.00	1.7331	2.0363	3.7602	0.8360	97.05	2.0951	2.7134	-0.6731	-0.1292
	2	94.70	1.1420	0.8764	1.8297	0.6191	96.30	1.3596	1.1723	0.4873	0.1423
	4	94.60	0.6795	0.3124	0.6870	0.3889	95.55	0.7966	0.4064	1.0358	0.5144
	10	94.15	0.2921	0.0635	-0.1853	-0.2325	94.05	0.3460	0.0771	0.5369	0.6126
(50,50)	0.1	94.95	2.4952	4.6276	9.7733	1.4514	97.40	2.9662	5.7962	5.9058	0.7779
	0.25	94.85	2.5161	4.3519	7.1925	1.0965	97.20	2.9872	5.6651	3.0396	0.4041
	0.5	95.15	2.1005	2.9434	5.3004	0.9814	96.50	2.5301	4.0666	0.2820	0.0442
	1	95.50	1.5253	1.5328	3.0061	0.7699	96.20	1.8603	2.1621	-0.1757	-0.0378
	2	95.45	0.9920	0.6456	1.1241	0.4427	96.90	1.2045	0.8603	0.0469	0.0160
	4	95.45	0.5840	0.2264	0.1382	0.0919	96.30	0.7018	0.3006	0.0756	0.0436
	10	95.15	0.2510	0.0435	-0.1152	-0.1747	96.30	0.3036	0.0567	0.3425	0.4552
(100,50)	0.1	94.20	2.3002	3.9695	11.1366	1.7954	96.05	2.7707	5.2242	10.1366	1.4161
	0.25	95.40	2.2967	3.4936	7.3041	1.2450	96.60	2.7097	4.7031	2.7741	0.4047
	0.5	95.65	1.8952	2.3579	5.2585	1.0891	96.60	2.2722	3.1843	-0.0068	-0.0012
	1	95.65	1.3558	1.2165	2.8077	0.8074	96.90	1.6531	1.7074	-0.0997	-0.0241
	2	95.35	0.8659	0.5043	0.8860	0.3947	96.70	1.0576	0.7037	-0.3652	-0.1377
	4	94.0	0.5017	0.1738	-0.0706	-0.0535	96.30	0.6096	0.2355	-0.1516	-0.0987
	10	94.40	0.2134	0.0329	-0.3035	-0.5295	95.85	0.2619	0.0431	0.1209	0.1841
(100,100)	0.1	94.55	1.7883	2.3587	4.9924	1.0332	96.60	2.1600	3.0342	2.5204	0.4579
	0.25	93.95	1.7855	2.2416	4.4353	0.9407	96.10	2.1399	3.0002	-0.2271	-0.0414
	0.5	94.15	1.4826	1.5316	3.5981	0.9231	95.70	1.8070	2.2049	-0.7541	-0.1606
	1	94.65	1.0751	0.8044	2.0776	0.7343	95.05	1.3221	1.1647	-0.4755	-0.1393
	2	94.65	0.6985	0.3398	0.6991	0.3794	95.60	0.8548	0.4819	-0.1853	-0.0844
	4	94.50	0.4117	0.1189	-0.0423	-0.0388	95.10	0.4999	0.1624	-0.0039	-0.0030
	10	94.45	0.1785	0.0228	-0.2469	-0.5171	95.95	0.2197	0.0303	0.1217	0.2213

Table 4.11: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	94.30	0.3142	0.0802	1.3004	1.4676	96.30	0.3699	0.0951	1.1929	1.2322
	0.25	94.65	0.8087	0.4642	2.6115	1.2208	96.75	0.9433	0.5583	1.2123	0.5136
	0.5	94.50	1.3619	1.2884	4.9094	1.3804	96.50	1.6238	1.6205	1.8644	0.4635
	1	94.50	1.9797	2.6810	6.4191	1.2490	96.10	2.3834	3.5082	2.2190	0.3748
	2	95.00	2.5748	4.5299	6.7740	1.0114	96.10	3.0848	6.0205	3.5141	0.4533
	4	94.40	3.0244	6.4078	5.9733	0.7481	95.60	3.5833	8.3289	6.8302	0.7503
	10	94.55	3.2010	7.3275	-1.8616	-0.2175	94.40	3.8269	9.2919	3.8761	0.4023
(30,50)	0.1	93.95	0.2746	0.0598	1.2936	1.6958	96.75	0.3244	0.0715	0.9811	1.1678
	0.25	94.95	0.6984	0.3334	2.3166	1.2787	97.15	0.8280	0.4292	0.9333	0.4508
	0.5	94.70	1.1779	0.9547	2.9422	0.9563	96.60	1.4178	1.2654	-0.5529	-0.1554
	1	95.00	1.7331	2.0363	3.7602	0.8360	97.05	2.0951	2.7134	-0.6731	-0.1292
	2	94.70	2.2840	3.5055	3.6594	0.6191	96.30	2.7192	4.6893	0.9746	0.1423
	4	94.60	2.7182	4.9984	2.7478	0.3889	95.55	3.1863	6.5028	4.1433	0.5144
	10	94.15	2.9208	6.3538	-1.8535	-0.2325	94.05	3.4602	7.7067	5.3688	0.6126
(50,50)	0.1	94.95	0.2495	0.0463	0.9773	1.4514	97.40	0.2966	0.0580	0.5906	0.7779
	0.25	94.85	0.6290	0.2720	1.7981	1.0965	97.20	0.7468	0.3541	0.7599	0.4041
	0.5	95.15	1.0503	0.7358	2.6502	0.9814	96.50	1.2651	1.0167	0.1410	0.0442
	1	95.50	1.5253	1.5328	3.0061	0.7699	96.20	1.8603	2.1621	-0.1757	-0.0378
	2	95.45	1.9840	2.5825	2.2483	0.4427	96.90	2.4091	3.4411	0.0938	0.0160
	4	95.45	2.3358	3.6225	0.5530	0.0919	96.30	2.8073	4.8093	0.3022	0.0436
	10	95.15	2.5099	4.3450	-1.1516	-0.1747	96.30	3.0363	5.6682	3.4245	0.4552
(100,50)	0.1	94.20	0.2300	0.0397	1.1137	1.7954	96.05	0.2771	0.0522	1.0137	1.4161
	0.25	95.40	0.5742	0.2184	1.8260	1.2450	96.60	0.6774	0.2939	0.6935	0.4047
	0.5	95.65	0.9476	0.5895	2.6293	1.0891	96.60	1.1361	0.7961	-0.0034	-0.0012
	1	95.65	1.3558	1.2165	2.8077	0.8074	96.90	1.6531	1.7074	-0.0997	-0.0241
	2	95.35	1.7317	2.0172	1.7720	0.3947	96.70	2.1152	2.8148	-0.7305	-0.1377
	4	94.80	2.0070	2.7807	-0.2823	-0.0535	96.30	2.4386	3.7684	-0.6062	-0.0987
	10	94.40	2.1342	3.2934	-3.0352	-0.5295	95.85	2.6189	4.3115	1.2089	0.1841
(100,100)	0.1	94.55	0.1788	0.0236	0.4992	1.0332	96.60	0.2160	0.0303	0.2520	0.4579
	0.25	93.95	0.4464	0.1401	1.1088	0.9407	96.10	0.5350	0.1875	-0.0568	-0.0414
	0.5	94.15	0.7413	0.3829	1.7990	0.9231	95.70	0.9035	0.5512	-0.3771	-0.1606
	1	94.65	1.0751	0.8044	2.0776	0.7343	95.05	1.3221	1.1647	-0.4755	-0.1393
	2	94.65	1.3970	1.3592	1.3982	0.3794	95.60	1.7095	1.9277	-0.3707	-0.0844
	4	94.50	1.6469	1.9017	-0.1692	-0.0388	95.10	1.9996	2.5989	-0.0155	-0.0030
	10	94.45	1.7846	2.2849	-2.4690	-0.5171	95.95	2.1970	3.0273	1.2174	0.2213

Table 4.12: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model d) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	93.60	1.9601	2.6490	9.7186	19.2237	95.70	2.5000	3.6708	7.7392	12.8759
	0.25	94.65	1.9586	2.5960	10.5375	21.1333	95.15	2.3634	3.6421	8.1978	13.7074
	0.5	94.10	1.9765	2.7345	6.0257	11.5975	95.65	2.3686	3.5692	5.5932	9.4010
	1	94.20	2.1586	3.1783	1.2946	2.2963	95.35	2.5541	4.1374	4.6809	7.2948
	2	95.05	2.6172	4.2687	-2.3834	-3.6495	95.15	3.0223	6.0920	2.9121	3.7327
	4	95.80	3.4117	6.5378	-3.2700	-4.0465	93.40	3.8475	9.9239	0.9303	0.9337
	10	94.95	4.9989	13.7329	3.2492	2.7730	82.20	5.0798	21.5444	3.7751	2.5721
(30,50)	0.1	94.40	1.5276	1.6226	9.5351	24.3574	96.85	1.9627	2.3074	5.9647	12.5108
	0.25	94.45	1.5524	1.6262	5.7782	14.4745	96.25	1.8793	2.2944	2.9334	6.1340
	0.5	93.60	1.6416	1.8711	2.0109	4.6527	94.65	1.9797	2.6081	1.4758	2.8902
	1	93.50	1.9119	2.5010	-2.4601	-4.9240	95.10	2.2781	3.4379	-0.6698	-1.1421
	2	94.65	2.4482	3.8076	-6.1161	-9.9583	93.90	2.8507	5.2970	-1.5904	-2.1851
	4	95.00	3.2863	6.2903	-7.4517	-9.4348	92.80	3.6823	9.4612	-2.4626	-2.5320
	10	95.70	4.8758	12.9872	0.0674	0.0591	82.65	4.9440	20.3726	-4.8858	-3.4242
(50,50)	0.1	93.25	1.5055	1.6711	9.7994	24.6857	95.85	1.9453	2.3908	6.5922	13.6029
	0.25	94.85	1.4989	1.4836	6.9251	18.2726	95.90	1.8412	2.1361	4.1830	9.0857
	0.5	94.40	1.5130	1.5380	3.7784	9.6769	95.80	1.8492	2.0938	3.4324	7.5205
	1	94.65	1.6495	1.8045	0.1012	0.2382	96.15	2.0059	2.5231	3.4091	6.8009
	2	95.00	1.9885	2.4736	-2.7971	-5.6315	95.75	2.3688	3.6090	2.4113	4.0161
	4	95.05	2.5622	3.8722	-3.6006	-5.7945	94.65	2.9859	5.8878	1.1797	1.5372
	10	95.45	3.6605	7.8405	0.3660	0.4132	91.90	4.2741	11.9501	-1.3522	-1.2368
(100,50)	0.1	94.05	1.4906	1.5712	8.3796	21.6238	96.00	1.9216	2.2922	5.9518	12.5255
	0.25	95.15	1.4512	1.4151	5.0013	13.4109	95.85	1.7801	2.0201	2.6410	5.8846
	0.5	94.40	1.4072	1.3841	2.0086	5.4056	95.75	1.7145	1.9064	2.0028	4.5907
	1	94.35	1.4263	1.4401	-1.3690	-3.6089	95.55	1.7287	2.0815	0.8154	1.7871
	2	94.50	1.5797	1.6766	-3.9672	-9.7322	95.00	1.9026	2.4848	-0.9504	-1.9065
	4	94.90	1.9028	2.2516	-4.6411	-9.8255	95.70	2.3010	3.3290	-2.1545	-3.7357
	10	94.70	2.5815	4.0513	-1.0389	-1.6321	94.45	3.2083	6.2107	-1.1979	-1.5198
(100,100)	0.1	94.30	1.0578	0.7878	4.9852	18.0444	96.20	1.3712	1.1747	1.4694	4.2901
	0.25	94.05	1.0529	0.7861	4.0763	14.6910	95.95	1.3026	1.1035	1.4648	4.4126
	0.5	93.15	1.0637	0.8199	1.9787	6.9252	95.20	1.3110	1.1330	1.4464	4.2998
	1	93.75	1.1596	0.9474	-0.5317	-1.7272	95.90	1.4289	1.3211	0.6421	1.7665
	2	94.95	1.3945	1.2606	-2.3942	-6.7570	95.65	1.6988	1.8238	0.8059	1.8869
	4	95.20	1.7832	1.9220	-2.5675	-5.8651	95.20	2.1476	3.0406	0.9602	1.7411
	10	95.30	2.5119	3.8273	1.1913	1.9254	93.45	3.1078	6.3254	0.7202	0.9054

Table 4.12: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	92.75	3.7865	12.0784	12.4457	1.1395	95.90	4.5675	13.9779	6.7618	0.5727
	0.25	94.25	3.4211	8.1618	13.6709	1.5305	95.95	4.0531	10.4540	3.7906	0.3709
	0.5	95.05	2.6620	4.7110	9.8865	1.4552	96.50	3.1761	6.1636	2.1877	0.2787
	1	95.15	1.8629	2.2930	5.5363	1.1637	96.75	2.2378	2.9999	0.7183	0.1311
	2	95.50	1.2093	0.9740	2.2751	0.7308	96.20	1.4484	1.2932	0.1612	0.0448
	4	95.30	0.7338	0.3697	0.4807	0.2501	96.05	0.8716	0.4970	0.4672	0.2095
	10	93.40	0.3419	0.0906	-0.2893	-0.3040	94.20	0.4006	0.1156	0.1380	0.1283
(30,50)	0.1	93.75	3.2581	7.9123	13.4788	1.5326	97.40	3.9480	9.7508	7.8644	0.7988
	0.25	94.40	2.8942	5.9458	10.3123	1.3492	96.35	3.4761	7.8438	-0.1980	-0.0224
	0.5	94.40	2.2656	3.6197	8.6820	1.4580	96.70	2.7427	4.8532	-1.3213	-0.1896
	1	94.55	1.6094	1.8067	5.3059	1.2578	96.05	1.9385	2.4777	-0.2442	-0.0490
	2	94.25	1.0619	0.7824	2.5264	0.9067	95.70	1.2631	1.0373	0.2237	0.0694
	4	93.80	0.6544	0.3000	0.8615	0.4979	94.75	0.7672	0.3918	0.4000	0.2021
	10	94.05	0.3107	0.0692	-0.2073	-0.2492	92.40	0.3525	0.0888	0.3792	0.4027
(50,50)	0.1	93.05	3.0167	7.4917	15.9326	1.8723	97.05	3.6943	9.3763	9.7337	1.0101
	0.25	95.05	2.6586	4.8505	10.5032	1.5252	97.00	3.1910	6.3965	1.9389	0.2424
	0.5	95.25	2.0570	2.8274	7.6406	1.4516	96.55	2.4857	3.9376	-0.0070	-0.0011
	1	95.10	1.4366	1.3720	4.3866	1.1924	96.40	1.7426	1.9126	0.0121	0.0028
	2	95.15	0.9296	0.5790	1.8298	0.7624	96.50	1.1255	0.7592	0.0371	0.0135
	4	94.70	0.5635	0.2178	0.3809	0.2581	96.85	0.6793	0.2844	0.3341	0.1981
	10	93.90	0.2644	0.0508	-0.2541	-0.3567	96.15	0.3199	0.0661	0.3181	0.3917
(100,50)	0.1	94.20	2.8462	6.1218	14.8191	1.9284	96.55	3.4935	8.3155	11.0687	1.2226
	0.25	95.25	2.4768	4.0482	9.8166	1.5612	97.05	2.9586	5.4647	1.6219	0.2194
	0.5	95.35	1.8915	2.3409	6.9548	1.4522	96.35	2.2693	3.2024	0.5725	0.1012
	1	95.50	1.2966	1.1117	3.8240	1.1542	96.55	1.5753	1.4893	-0.1222	-0.0317
	2	95.05	0.8201	0.4551	1.4406	0.6767	96.45	1.0002	0.6272	0.1756	0.0701
	4	94.30	0.4862	0.1659	0.1345	0.1044	96.40	0.5952	0.2287	-0.0370	-0.0244
	10	93.55	0.2237	0.0375	-0.3878	-0.6347	96.00	0.2774	0.0483	-0.1972	-0.2839
(100,100)	0.1	93.65	2.1687	3.5202	6.5040	1.1026	96.65	2.6805	4.6449	0.2944	0.0432
	0.25	94.20	1.8854	2.5256	6.7726	1.3597	96.10	2.2927	3.4440	-0.8618	-0.1468
	0.5	94.10	1.4542	1.4821	5.2350	1.3722	95.45	1.7729	2.0595	-0.3156	-0.0695
	1	94.15	1.0140	0.7218	3.0224	1.1319	95.50	1.2410	1.0287	-0.5170	-0.1612
	2	94.10	0.6547	0.3047	1.1722	0.6729	95.65	0.8001	0.4219	-0.3278	-0.1596
	4	93.95	0.3961	0.1144	0.1121	0.1048	95.55	0.4835	0.1547	-0.0871	-0.0700
	10	93.05	0.1865	0.0266	-0.3280	-0.6365	95.20	0.2310	0.0340	0.0548	0.0940

Table 4.12: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	92.75	0.3786	0.1208	1.2446	1.1395	95.90	0.4567	0.1398	0.6762	0.5727
	0.25	94.25	0.8553	0.5101	3.4177	1.5305	95.95	1.0133	0.6534	0.9477	0.3709
	0.5	95.05	1.3310	1.1778	4.9432	1.4552	96.50	1.5880	1.5409	1.0939	0.2787
	1	95.15	1.8629	2.2930	5.5363	1.1637	96.75	2.2378	2.9999	0.7183	0.1311
	2	95.50	2.4187	3.8960	4.5503	0.7308	96.20	2.8969	5.1730	0.3225	0.0448
	4	95.30	2.9353	5.9147	1.9230	0.2501	96.05	3.4863	7.9525	1.8686	0.2095
	10	93.40	3.4190	9.0627	-2.8933	-0.3040	94.20	4.0056	11.5628	1.3797	0.1283
(30,50)	0.1	93.75	0.3258	0.0791	1.3479	1.5326	97.40	0.3948	0.0975	0.7864	0.7988
	0.25	94.40	0.7235	0.3716	2.5781	1.3492	96.35	0.8690	0.4902	-0.0495	-0.0224
	0.5	94.40	1.1328	0.9049	4.3410	1.4580	96.70	1.3714	1.2133	-0.6606	-0.1896
	1	94.55	1.6094	1.8067	5.3059	1.2578	96.05	1.9385	2.4777	-0.2442	-0.0490
	2	94.25	2.1238	3.1296	5.0528	0.9067	95.70	2.5262	4.1493	0.4473	0.0694
	4	93.80	2.6177	4.8000	3.4459	0.4979	94.75	3.0690	6.2692	1.6001	0.2021
	10	94.05	3.1070	6.9198	-2.0728	-0.2492	92.40	3.5250	8.8775	3.7921	0.4027
(50,50)	0.1	93.05	0.3017	0.0749	1.5933	1.8723	97.05	0.3694	0.0938	0.9734	1.0101
	0.25	95.05	0.6647	0.3032	2.6258	1.5252	97.00	0.7978	0.3998	0.4847	0.2424
	0.5	95.25	1.0285	0.7068	3.8203	1.4516	96.55	1.2429	0.9844	-0.0035	-0.0011
	1	95.10	1.4366	1.3720	4.3866	1.1924	96.40	1.7426	1.9126	0.0121	0.0028
	2	95.15	1.8592	2.3162	3.6597	0.7624	96.50	2.2510	3.0367	0.0742	0.0135
	4	94.70	2.2541	3.4846	1.5236	0.2581	96.85	2.7171	4.5509	1.3364	0.1981
	10	93.90	2.6445	5.0801	-2.5411	-0.3567	96.15	3.1993	6.6051	3.1815	0.3917
(100,50)	0.1	94.20	0.2846	0.0612	1.4819	1.9284	96.55	0.3493	0.0832	1.1069	1.2226
	0.25	95.25	0.6192	0.2530	2.4541	1.5612	97.05	0.7396	0.3415	0.4055	0.2194
	0.5	95.35	0.9457	0.5852	3.4774	1.4522	96.35	1.1347	0.8006	0.2863	0.1012
	1	95.50	1.2966	1.1117	3.8240	1.1542	96.55	1.5753	1.4893	-0.1222	-0.0317
	2	95.05	1.6401	1.8202	2.8812	0.6767	96.45	2.0004	2.5088	0.3512	0.0701
	4	94.30	1.9450	2.6541	0.5381	0.1044	96.40	2.3807	3.6593	-0.1478	-0.0244
	10	93.55	2.2372	3.7471	-3.8782	-0.6347	96.00	2.7745	4.8273	-1.9723	-0.2839
(100,100)	0.1	93.65	0.2169	0.0352	0.6504	1.1026	96.65	0.2680	0.0464	0.0294	0.0432
	0.25	94.20	0.4714	0.1578	1.6932	1.3597	96.10	0.5732	0.2152	-0.2155	-0.1468
	0.5	94.10	0.7271	0.3705	2.6175	1.3722	95.45	0.8865	0.5149	-0.1578	-0.0695
	1	94.15	1.0140	0.7218	3.0224	1.1319	95.50	1.2410	1.0287	-0.5170	-0.1612
	2	94.10	1.3095	1.2188	2.3444	0.6729	95.65	1.6002	1.6876	-0.6556	-0.1596
	4	93.95	1.5844	1.8305	0.4483	0.1048	95.55	1.9339	2.4752	-0.3485	-0.0700
	10	93.05	1.8646	2.6646	-3.2800	-0.6365	95.20	2.3101	3.3990	0.5483	0.0940

Table 4.13: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model d) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-2}$
(30,30)	0.1	94.35	2.6896	5.0627	16.2442	23.4436	95.50	3.6273	7.7935	14.9769	17.2103
	0.25	95.80	2.6491	4.5398	11.7420	17.6935	96.45	3.3014	6.3967	8.3923	10.5486
	0.5	95.65	2.7054	4.6808	7.7902	11.4581	96.40	3.2483	6.4732	7.2755	9.0777
	1	95.25	2.9418	5.3911	3.8524	5.2527	95.50	3.4306	7.5465	5.9100	6.8172
	2	95.35	3.4253	6.9862	1.2514	1.4970	94.25	3.9213	10.4341	3.0348	2.9716
	4	95.30	4.1899	9.9897	1.2426	1.2430	89.15	4.6477	15.8376	-1.3644	-1.0839
	10	95.10	5.6404	17.7230	7.0838	5.3273	72.90	4.8890	28.1245	0.8998	0.5364
(30,50)	0.1	94.25	2.0933	2.9117	9.8055	18.4748	96.90	2.8047	4.2897	7.2452	11.1275
	0.25	94.70	2.1353	3.0084	8.3753	15.4471	96.60	2.6570	4.3162	6.6133	10.1151
	0.5	94.85	2.2801	3.3475	5.3083	9.2114	96.30	2.7304	4.5947	5.5184	8.1663
	1	95.15	2.6159	4.1975	2.2543	3.4808	95.50	3.0396	5.8844	4.6433	6.0626
	2	95.45	3.1890	5.8594	0.3766	0.4919	94.40	3.6321	8.4553	0.2857	0.3107
	4	95.70	4.0172	8.8176	0.7461	0.7944	89.50	4.5205	13.7603	-3.3631	-2.8675
	10	95.80	5.4936	16.0914	6.2129	4.9024	72.80	4.8539	25.5092	-1.7691	-1.1074
(50,50)	0.1	94.40	2.0579	3.0098	11.4008	21.2397	96.70	2.7939	4.5062	9.4807	14.2627
	0.25	94.60	2.0324	2.9101	8.2232	15.4199	96.35	2.5506	4.3539	6.1886	9.4181
	0.5	94.15	2.0777	3.0212	5.6603	10.3503	96.25	2.5216	4.4204	5.9298	8.9524
	1	94.40	2.2568	3.4460	3.3207	5.6645	94.95	2.6724	4.8701	5.2052	7.4777
	2	94.95	2.6181	4.3932	2.2882	3.4534	94.95	3.0630	6.4907	3.5163	4.3676
	4	94.85	3.1800	6.2145	3.5448	4.5000	93.65	3.7205	9.6176	0.5702	0.5813
	10	94.90	4.2136	10.9894	10.0922	9.6697	85.00	4.9150	17.5364	-2.1020	-1.5871
(100,50)	0.1	95.05	2.0170	2.6928	8.5258	16.6519	97.10	2.7492	4.2807	7.1490	10.9898
	0.25	95.25	1.9463	2.4671	6.2644	12.7103	96.65	2.4493	3.8093	5.5450	9.0184
	0.5	94.90	1.9068	2.4020	3.8251	7.8266	95.75	2.3193	3.4328	3.7963	6.4915
	1	94.95	1.9413	2.4853	1.4842	2.9778	95.50	2.3262	3.4691	2.0827	3.5374
	2	95.30	2.0957	2.8204	0.1504	0.2831	95.10	2.5141	4.0960	0.7676	1.1991
	4	95.30	2.3917	3.5742	0.6303	1.0540	95.30	2.9322	5.3938	-0.9435	-1.2844
	10	94.50	2.9950	5.6983	5.1385	6.8212	93.00	3.8264	9.4423	-1.6955	-1.7447
(100,100)	0.1	94.20	1.4372	1.4411	5.6883	15.1515	97.25	1.9330	2.3067	2.2324	4.6519
	0.25	94.20	1.4218	1.3805	4.7326	12.8388	96.25	1.7906	2.0584	1.8166	4.0062
	0.5	94.25	1.4525	1.4239	3.0589	8.1311	96.00	1.7839	2.0657	1.9342	4.2584
	1	94.65	1.5754	1.6189	1.4517	3.6095	95.60	1.9088	2.3890	1.7673	3.6173
	2	94.85	1.8211	2.0693	0.7606	1.6719	95.20	2.1952	3.1530	1.1239	2.0015
	4	94.60	2.1960	2.9445	1.7344	3.1971	94.50	2.6994	4.6853	0.0399	0.0582
	10	95.00	2.8706	5.2234	6.5785	9.1379	92.70	3.7206	9.3420	-0.1140	-0.1180

Table 4.13: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	93.95	4.3994	15.0889	20.4425	1.6873	96.50	5.5525	19.3471	10.5497	0.7605
	0.25	94.00	3.3502	7.8756	14.4040	1.6445	96.30	4.0824	10.3677	1.7219	0.1691
	0.5	94.90	2.4018	3.8835	9.7541	1.5844	96.15	2.9008	5.3117	0.1697	0.0233
	1	95.20	1.6067	1.7120	5.5311	1.3486	96.10	1.9343	2.3641	-0.2272	-0.0467
	2	94.95	1.0252	0.6979	2.4857	0.9449	96.25	1.2269	0.9484	0.4300	0.1396
	4	94.75	0.6293	0.2684	0.7183	0.4388	95.35	0.7442	0.3534	0.5587	0.2972
	10	93.85	0.3114	0.0701	-0.2113	-0.2523	93.15	0.3475	0.0789	-0.0626	-0.0704
(30,50)	0.1	94.20	3.6720	9.9344	16.0622	1.6325	97.80	4.7150	13.0701	3.8952	0.3408
	0.25	94.30	2.7732	5.2027	11.5037	1.6151	97.50	3.4297	6.9401	-0.2878	-0.0345
	0.5	95.15	2.0109	2.6317	7.9974	1.5779	96.85	2.4551	3.5850	-0.3301	-0.0551
	1	95.30	1.3703	1.2018	4.5469	1.3227	96.65	1.6524	1.6791	-0.0006	-0.0002
	2	95.70	0.8913	0.5091	2.0068	0.8927	96.50	1.0619	0.6886	0.5868	0.2236
	4	95.30	0.5563	0.2028	0.5350	0.3759	95.30	0.6563	0.2616	0.5152	0.3186
	10	94.10	0.2804	0.0548	-0.2233	-0.3017	91.05	0.3017	0.0621	-0.0085	-0.0108
(50,50)	0.1	92.70	3.4730	9.2484	16.8943	1.7840	97.15	4.4944	12.1365	7.2965	0.6636
	0.25	94.10	2.5987	4.5710	11.5302	1.7303	97.20	3.2211	6.3406	1.1119	0.1396
	0.5	94.55	1.8578	2.2640	7.7783	1.6566	96.25	2.2710	3.1891	-0.2370	-0.0420
	1	95.20	1.2421	1.0065	4.2761	1.3600	96.10	1.5090	1.4550	-0.0225	-0.0059
	2	94.85	0.7906	0.4151	1.7773	0.8754	95.70	0.9581	0.5898	0.1460	0.0601
	4	94.40	0.4835	0.1618	0.3681	0.2894	95.95	0.5853	0.2207	0.1222	0.0822
	10	93.00	0.2387	0.0430	-0.3232	-0.4935	94.55	0.2872	0.0532	0.0693	0.0950
(100,50)	0.1	94.20	3.3468	7.8692	13.6731	1.5596	97.35	4.3456	11.2635	6.3770	0.6018
	0.25	95.20	2.4774	3.9292	10.1680	1.6435	97.05	3.0594	5.6296	1.2058	0.1607
	0.5	95.60	1.7469	1.9273	6.9598	1.6053	97.10	2.1298	2.7399	0.5256	0.1004
	1	95.45	1.1442	0.8332	3.9078	1.3661	96.90	1.3894	1.1777	0.6755	0.1968
	2	95.25	0.7106	0.3291	1.6961	0.9388	97.10	0.8630	0.4501	0.2503	0.1179
	4	94.75	0.4237	0.1218	0.4213	0.3819	97.20	0.5208	0.1608	-0.1270	-0.1001
	10	93.20	0.2032	0.0303	-0.2366	-0.4302	96.20	0.2544	0.0398	-0.1507	-0.2390
(100,100)	0.1	93.00	2.4954	4.7385	8.3240	1.2179	97.20	3.2635	6.5386	-1.3955	-0.1726
	0.25	94.00	1.8451	2.4233	7.7210	1.5877	96.40	2.3101	3.4196	-1.5155	-0.2592
	0.5	94.05	1.3164	1.2077	5.5348	1.6128	95.85	1.6228	1.6992	-0.8971	-0.2176
	1	94.25	0.8780	0.5373	3.1456	1.3694	96.30	1.0750	0.7433	-0.5996	-0.2200
	2	94.40	0.5567	0.2205	1.3342	0.9020	96.25	0.6825	0.3053	-0.3034	-0.1736
	4	94.10	0.3390	0.0850	0.2782	0.3018	96.00	0.4189	0.1134	-0.1146	-0.1076
	10	92.35	0.1667	0.0222	-0.2565	-0.5445	95.40	0.2107	0.0297	-0.1223	-0.2244

Table 4.13: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-3}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	93.95	0.4399	0.1509	2.0443	1.6873	96.50	0.5553	0.1935	1.0550	0.7605
	0.25	94.00	0.8375	0.4922	3.6010	1.6445	96.30	1.0206	0.6480	0.4305	0.1691
	0.5	94.90	1.2009	0.9709	4.8771	1.5844	96.15	1.4504	1.3279	0.0848	0.0233
	1	95.20	1.6067	1.7120	5.5311	1.3486	96.10	1.9343	2.3641	-0.2272	-0.0467
	2	94.95	2.0505	2.7916	4.9714	0.9449	96.25	2.4539	3.7934	0.8599	0.1396
	4	94.75	2.5173	4.2946	2.8732	0.4388	95.35	2.9767	5.6541	2.2347	0.2972
	10	93.85	3.1143	7.0127	-2.1130	-0.2523	93.15	3.4749	7.8873	-0.6258	-0.0704
(30,50)	0.1	94.20	0.3672	0.0993	1.6062	1.6325	97.80	0.4715	0.1307	0.3895	0.3408
	0.25	94.30	0.6933	0.3252	2.8759	1.6151	97.50	0.8574	0.4338	-0.0719	-0.0345
	0.5	95.15	1.0055	0.6579	3.9987	1.5779	96.85	1.2276	0.8963	-0.1651	-0.0551
	1	95.30	1.3703	1.2018	4.5469	1.3227	96.65	1.6524	1.6791	-0.0006	-0.0002
	2	95.70	1.7826	2.0365	4.0136	0.8927	96.50	2.1237	2.7546	1.1737	0.2236
	4	95.30	2.2253	3.2441	2.1399	0.3759	95.30	2.6251	4.1852	2.0606	0.3186
	10	94.10	2.8036	5.4773	-2.2327	-0.3017	91.05	3.0166	6.2063	-0.0848	-0.0108
(50,50)	0.1	92.70	0.3473	0.0925	1.6894	1.7840	97.15	0.4494	0.1214	0.7297	0.6636
	0.25	94.10	0.6497	0.2857	2.8826	1.7303	97.20	0.8053	0.3963	0.2780	0.1396
	0.5	94.55	0.9289	0.5660	3.8892	1.6566	96.25	1.1355	0.7973	-0.1185	-0.0420
	1	95.20	1.2421	1.0065	4.2761	1.3600	96.10	1.5090	1.4550	-0.0225	-0.0059
	2	94.85	1.5811	1.6605	3.5547	0.8754	95.70	1.9161	2.3591	0.2920	0.0601
	4	94.40	1.9339	2.5887	1.4722	0.2894	95.95	2.3412	3.5306	0.4888	0.0822
	10	93.00	2.3866	4.2981	-3.2324	-0.4935	94.55	2.8724	5.3196	0.6934	0.0950
(100,50)	0.1	94.20	0.3347	0.0787	1.3673	1.5596	97.35	0.4346	0.1126	0.6377	0.6018
	0.25	95.20	0.6193	0.2456	2.5420	1.6435	97.05	0.7648	0.3519	0.3015	0.1607
	0.5	95.60	0.8735	0.4818	3.4799	1.6053	97.10	1.0649	0.6850	0.2628	0.1004
	1	95.45	1.1442	0.8332	3.9078	1.3661	96.90	1.3894	1.1777	0.6755	0.1968
	2	95.25	1.4212	1.3165	3.3922	0.9388	97.10	1.7260	1.8005	0.5006	0.1179
	4	94.75	1.6948	1.9491	1.6851	0.3819	97.20	2.0831	2.5723	-0.5079	-0.1001
	10	93.20	2.0323	3.0298	-2.3663	-0.4302	96.20	2.5441	3.9768	-1.5070	-0.2390
(100,100)	0.1	93.00	0.2495	0.0474	0.8324	1.2179	97.20	0.3263	0.0654	-0.1396	-0.1726
	0.25	94.00	0.4613	0.1515	1.9302	1.5877	96.40	0.5775	0.2137	-0.3789	-0.2592
	0.5	94.05	0.6582	0.3019	2.7674	1.6128	95.85	0.8114	0.4248	-0.4486	-0.2176
	1	94.25	0.878	0.5373	3.1456	1.3694	96.30	1.0750	0.7433	-0.5996	-0.2200
	2	94.40	1.1135	0.8818	2.6683	0.9020	96.25	1.3651	1.2213	-0.6068	-0.1736
	4	94.10	1.3559	1.3607	1.1129	0.3018	96.00	1.6757	1.8137	-0.4584	-0.1076
	10	92.35	1.6674	2.2237	-2.5646	-0.5445	95.40	2.1067	2.9709	-1.2229	-0.2244

Table 4.14: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model e) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^0$	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-1}$	%	$\times 10^0$	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-1}$
(30,30)	0.1	97.05	1.6222	1.2067	-17.8176	-5.9738	92.90	1.3507	1.2660	2.0736	0.5836
	0.25	97.65	1.3189	0.8219	3.0397	1.0660	95.25	1.2495	0.9646	1.0994	0.3541
	0.5	95.35	1.1647	0.9294	10.5170	3.6745	96.10	1.2821	0.9667	1.1191	0.3601
	1	94.15	1.1468	0.9295	0.6745	0.2212	96.30	1.4018	1.2022	0.0493	0.0142
	2	95.25	1.3042	1.1536	-12.263	-3.8707	95.45	1.4029	1.2325	-2.0336	-0.5801
	4	97.95	1.6351	1.0972	-4.1774	-1.2710	95.00	1.4002	1.1858	-2.2975	-0.6685
	10	97.75	2.3186	2.2077	30.4395	8.5021	91.65	1.4804	1.5308	-1.6405	-0.4196
(30,50)	0.1	96.30	1.1785	0.7788	-16.3031	-7.1963	96.30	1.0961	0.7226	1.0045	0.3738
	0.25	97.15	1.0058	0.5258	2.7845	1.2231	96.30	0.9938	0.5930	1.3139	0.5402
	0.5	93.80	0.9483	0.6919	8.5693	3.4449	95.80	1.0756	0.6878	1.4841	0.5666
	1	93.50	1.0152	0.7600	-1.6893	-0.6138	95.90	1.2477	1.0480	0.3321	0.1026
	2	94.90	1.2445	1.0731	-13.8982	-4.6840	94.90	1.3260	1.1209	-2.1768	-0.6514
	4	98.45	1.6421	1.0417	-4.1466	-1.2952	94.25	1.3549	1.1342	-1.9931	-0.5927
	10	97.25	2.4147	2.6566	34.4233	8.9712	89.75	1.4515	1.4714	-2.4571	-0.6417
(50,50)	0.1	95.70	1.2263	0.8535	-19.2017	-8.7190	95.30	1.0796	0.6931	0.6725	0.2555
	0.25	98.20	1.0054	0.4645	2.2549	1.0517	95.75	0.9677	0.5722	0.9195	0.3846
	0.5	93.60	0.8910	0.6228	10.0656	4.4066	95.80	1.0044	0.6359	1.0621	0.4214
	1	92.80	0.8763	0.5963	0.8773	0.3594	95.70	1.0999	0.8069	0.7608	0.2679
	2	92.90	0.9911	0.7730	-11.5701	-4.5755	95.65	1.0902	0.7965	-1.0622	-0.3765
	4	97.80	1.2313	0.6810	-3.6115	-1.3970	94.60	1.0788	0.7428	-0.9248	-0.3394
	10	95.00	1.6976	1.7483	30.3927	10.5817	93.75	1.1962	0.9254	-0.8610	-0.2831
(100,50)	0.1	94.20	1.3063	1.1546	-25.3848	-11.2351	96.80	1.0620	0.6635	0.8439	0.3277
	0.25	98.95	1.0246	0.4180	1.5740	0.7719	96.50	0.9357	0.5218	0.3679	0.1610
	0.5	95.00	0.8536	0.5755	12.0925	5.8347	96.05	0.9402	0.5272	0.9218	0.4017
	1	92.85	0.7607	0.4596	4.0877	1.9419	95.35	0.9567	0.6065	0.4934	0.2003
	2	90.95	0.7683	0.5023	-8.9127	-4.3332	95.00	0.8544	0.4791	-0.7625	-0.3485
	4	97.00	0.8727	0.3919	-3.3287	-1.7052	96.05	0.8055	0.3943	-1.1353	-0.5725
	10	91.55	1.1127	0.9821	24.5741	12.6325	95.85	0.8748	0.4617	-1.0296	-0.4796
(100,100)	0.1	89.45	0.8454	0.6746	-20.5634	-12.9580	97.15	0.7672	0.3590	0.2149	0.1134
	0.25	97.85	0.7009	0.2277	0.9231	0.6127	96.30	0.6782	0.2849	-0.2423	-0.1435
	0.5	91.25	0.6249	0.3393	8.9053	5.5215	96.05	0.7057	0.3195	-0.4733	-0.2648
	1	92.25	0.6156	0.2982	0.0891	0.0516	95.30	0.7914	0.4165	-0.5442	-0.2667
	2	89.60	0.6956	0.4622	-11.9154	-6.6568	95.80	0.7741	0.3792	-1.1852	-0.6096
	4	97.75	0.8602	0.3485	-3.5896	-1.9590	95.45	0.7594	0.3629	-1.1800	-0.6205
	10	88.80	1.1732	1.3475	30.3893	14.7545	94.65	0.8508	0.4803	-1.9335	-0.8855

Table 4.14: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	89.80	3.3255	11.6184	-6.8651	-0.8259	95.95	3.5783	8.2966	-0.1974	-0.0217
	0.25	97.00	3.2662	5.9229	-0.7780	-0.1016	97.15	3.3723	6.7187	0.1613	0.0197
	0.5	92.05	2.6961	6.4741	5.0478	0.8054	97.10	3.0621	5.6151	0.3145	0.0420
	1	68.50	1.9591	8.4664	7.9677	1.7309	95.50	2.4818	3.9225	0.8687	0.1400
	2	77.00	1.2852	2.8487	4.3532	1.4093	95.15	1.5442	1.5185	0.5268	0.1364
	4	94.35	0.7639	0.4177	0.8132	0.4336	96.05	0.8188	0.4187	0.0680	0.0332
	10	91.30	0.3297	0.0991	-0.5018	-0.5834	95.00	0.3322	0.0707	-0.0344	-0.0409
(30,50)	0.1	88.90	2.8843	9.0609	-6.0911	-0.8325	96.85	3.3136	6.4904	-0.0921	-0.0114
	0.25	96.70	2.7937	4.5267	-0.6266	-0.0935	96.45	3.1315	5.8384	-0.1916	-0.0251
	0.5	87.75	2.3272	5.8486	5.1241	0.9024	95.80	2.8707	5.3532	0.1688	0.0231
	1	59.60	1.7274	7.8498	7.8279	1.8858	94.40	2.2345	3.3032	0.6988	0.1225
	2	74.85	1.1581	2.4759	4.1385	1.4977	95.25	1.3043	1.0842	0.3826	0.1170
	4	95.30	0.7003	0.3276	0.6779	0.4039	95.25	0.6697	0.2872	0.0948	0.0560
	10	91.60	0.3063	0.0864	-0.5860	-0.8122	96.15	0.2719	0.0447	0.0042	0.0062
(50,50)	0.1	84.55	2.6247	9.1754	-7.3046	-1.1786	97.05	2.8638	5.0645	-0.3465	-0.0487
	0.25	96.05	2.5321	3.7715	-1.2538	-0.2085	95.80	2.6128	4.2379	-0.0601	-0.0092
	0.5	87.55	2.0799	4.6771	4.7388	0.9608	96.20	2.3738	3.4762	0.2170	0.0368
	1	50.55	1.5093	7.3347	7.7618	2.1439	96.25	1.9251	2.3632	0.5502	0.1139
	2	64.50	0.9894	2.3766	4.2350	1.7534	95.90	1.1967	0.9192	0.2287	0.0756
	4	92.65	0.5895	0.2816	0.8072	0.5486	95.75	0.6354	0.2541	0.0392	0.0246
	10	88.45	0.2580	0.0695	-0.5023	-0.7551	95.55	0.2601	0.0433	-0.0161	-0.0245
(100,50)	0.1	76.70	2.4338	10.3178	-8.6909	-1.6525	97.20	2.4489	3.6994	-0.0701	-0.0115
	0.25	96.70	2.3640	2.9882	-2.1033	-0.4167	97.20	2.1295	2.6290	-0.1312	-0.0256
	0.5	89.70	1.9103	3.7381	4.4319	1.0520	96.35	1.8989	2.1373	0.1407	0.0304
	1	39.70	1.3494	7.1408	7.8384	2.4821	95.65	1.6329	1.6540	0.4269	0.1055
	2	48.45	0.8591	2.4599	4.4707	2.0813	95.30	1.1175	0.8226	0.1583	0.0553
	4	86.65	0.5008	0.2847	1.0412	0.7839	95.15	0.6099	0.2512	-0.0029	-0.0018
	10	87.85	0.2176	0.0499	-0.3532	-0.5771	94.95	0.2498	0.0418	-0.0271	-0.0420
(100,100)	0.1	68.20	1.8821	7.5737	-7.4200	-1.6312	95.70	2.0699	2.7094	0.0275	0.0053
	0.25	95.95	1.7929	1.9396	-1.2993	-0.3087	95.35	1.8551	2.1664	0.0361	0.0077
	0.5	79.20	1.4656	3.4161	4.7456	1.3906	95.60	1.6872	1.7950	0.0983	0.0232
	1	18.45	1.0619	6.6823	7.7878	3.1338	95.80	1.3656	1.1503	0.3002	0.0889
	2	35.15	0.6959	2.0883	4.2605	2.5773	95.85	0.8463	0.4414	0.1214	0.0579
	4	89.80	0.4157	0.1693	0.8223	0.8150	95.85	0.4506	0.1245	0.0279	0.0250
	10	82.05	0.1838	0.0458	-0.4956	-1.0763	96.05	0.1853	0.0204	0.0064	0.0141

Table 4.14: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$	%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$
(30,30)	0.1	89.80	0.3325	0.1162	-0.6865	-0.8259	95.95	0.3578	0.0830	-0.0197	-0.0217
	0.25	97.00	0.8166	0.3702	-0.1945	-0.1016	97.15	0.8431	0.4199	0.0403	0.0197
	0.5	92.05	1.3481	1.6185	2.5239	0.8054	97.10	1.5311	1.4038	0.1572	0.0420
	1	68.50	1.9591	8.4664	7.9677	1.7309	95.50	2.4818	3.9225	0.8687	0.1400
	2	77.00	2.5704	11.3946	8.7064	1.4093	95.15	3.0885	6.0740	1.0536	0.1364
	4	94.35	3.0557	6.6839	3.2528	0.4336	96.05	3.2752	6.6985	0.2720	0.0332
	10	91.30	3.2965	9.9122	-5.0177	-0.5834	95.00	3.3216	7.0688	-0.3437	-0.0409
(30,50)	0.1	88.90	0.2884	0.0906	-0.6091	-0.8325	96.85	0.3314	0.0649	-0.0092	-0.0114
	0.25	96.70	0.6984	0.2829	-0.1566	-0.0935	96.45	0.7829	0.3649	-0.0479	-0.0251
	0.5	87.75	1.1636	1.4621	2.5620	0.9024	95.80	1.4353	1.3383	0.0844	0.0231
	1	59.60	1.7274	7.8498	7.8279	1.8858	94.40	2.2345	3.3032	0.6988	0.1225
	2	74.85	2.3162	9.9036	8.2770	1.4977	95.25	2.6087	4.3366	0.7652	0.1170
	4	95.30	2.8011	5.2412	2.7116	0.4039	95.25	2.6790	4.5955	0.3791	0.0560
	10	91.60	3.0628	8.6376	-5.8603	-0.8122	96.15	2.7187	4.4683	0.0415	0.0062
(50,50)	0.1	84.55	0.2625	0.0918	-0.7305	-1.1786	97.05	0.2864	0.0506	-0.0347	-0.0487
	0.25	96.05	0.6330	0.2357	-0.3134	-0.2085	95.80	0.6532	0.2649	-0.0150	-0.0092
	0.5	87.55	1.0399	1.1693	2.3694	0.9608	96.20	1.1869	0.8691	0.1085	0.0368
	1	50.55	1.5093	7.3347	7.7618	2.1439	96.25	1.9251	2.3632	0.5502	0.1139
	2	64.50	1.9788	9.5062	8.4700	1.7534	95.90	2.3933	3.6770	0.4574	0.0756
	4	92.65	2.3582	4.5050	3.2288	0.5486	95.75	2.5416	4.0659	0.1567	0.0246
	10	88.45	2.5801	6.9466	-5.0234	-0.7551	95.55	2.6013	4.3321	-0.1612	-0.0245
(100,50)	0.1	76.70	0.2434	0.1032	-0.8691	-1.6525	97.20	0.2449	0.0370	-0.0070	-0.0115
	0.25	96.70	0.5910	0.1868	-0.5258	-0.4167	97.20	0.5324	0.1643	-0.0328	-0.0256
	0.5	89.70	0.9552	0.9345	2.2160	1.0520	96.35	0.9494	0.5343	0.0704	0.0304
	1	39.70	1.3494	7.1408	7.8384	2.4821	95.65	1.6329	1.6540	0.4269	0.1055
	2	48.45	1.7182	9.8397	8.9414	2.0813	95.30	2.2351	3.2905	0.3166	0.0553
	4	86.65	2.0032	4.5558	4.1647	0.7839	95.15	2.4396	4.0193	-0.0116	-0.0018
	10	87.85	2.1758	4.9907	-3.5320	-0.5771	94.95	2.4982	4.1769	-0.2712	-0.0420
(100,100)	0.1	68.20	0.1882	0.0757	-0.7420	-1.6312	95.70	0.2070	0.0271	0.0028	0.0053
	0.25	95.95	0.4482	0.1212	-0.3248	-0.3087	95.35	0.4638	0.1354	0.0090	0.0077
	0.5	79.20	0.7328	0.8540	2.3728	1.3906	95.60	0.8436	0.4487	0.0491	0.0232
	1	18.45	1.0619	6.6823	7.7878	3.1338	95.80	1.3656	1.1503	0.3002	0.0889
	2	35.15	1.3919	8.3531	8.5209	2.5773	95.85	1.6926	1.7655	0.2428	0.0579
	4	89.80	1.6627	2.7096	3.2891	0.8150	95.85	1.8026	1.9914	0.1115	0.0250
	10	82.05	1.8378	4.5753	-4.9558	-1.0763	96.05	1.8535	2.0406	0.0638	0.0141

Table 4.15: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model e) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	94.20	1.5778	1.5516	-27.3130	-9.6205	92.90	1.4067	1.4084	2.4183	0.6456
	0.25	96.15	1.3315	0.8858	-12.9823	-4.8461	94.95	1.2723	1.0644	1.0832	0.3321
	0.5	96.95	1.2160	0.7596	-7.2342	-2.7195	96.10	1.1998	0.8529	-0.0295	-0.0101
	1	96.45	1.2144	0.8175	-4.6396	-1.6441	96.55	1.1842	0.8069	-0.8960	-0.3155
	2	97.55	1.3582	0.8790	0.0915	0.0309	95.90	1.2229	0.8828	-1.0667	-0.3592
	4	98.35	1.6595	1.1402	12.7480	4.0759	94.90	1.3220	1.0951	-1.7111	-0.5176
	10	93.95	2.2916	3.2145	43.5148	11.9696	89.45	1.4335	1.5830	-2.3784	-0.5987
(30,50)	0.1	91.65	1.1522	1.0564	-24.0518	-10.9993	95.65	1.1666	0.8898	1.1201	0.3757
	0.25	95.90	1.0277	0.6218	-12.7791	-5.9664	96.05	1.0419	0.6510	0.8428	0.3304
	0.5	95.75	1.0001	0.6083	-8.2275	-3.5378	95.60	1.0177	0.6076	0.2532	0.1027
	1	96.00	1.0764	0.6925	-6.0367	-2.3563	95.70	1.0544	0.6672	-0.2080	-0.0805
	2	97.60	1.2890	0.7778	-0.6474	-0.2321	95.15	1.1392	0.7831	-0.7845	-0.2804
	4	97.65	1.6543	1.1219	13.7949	4.5185	93.75	1.2598	0.9868	-1.7392	-0.5544
	10	92.90	2.3895	3.8747	49.4611	13.0842	88.60	1.3925	1.5207	-3.5958	-0.9258
(50,50)	0.1	89.50	1.1950	1.3036	-28.7893	-13.2099	95.30	1.1619	0.8427	0.3916	0.1349
	0.25	94.95	1.0149	0.6101	-14.1290	-6.9717	95.90	1.0072	0.6346	0.4743	0.1883
	0.5	95.20	0.9301	0.5116	-7.6931	-3.6161	95.70	0.9432	0.5551	0.2375	0.1008
	1	95.15	0.9281	0.5345	-4.5304	-1.9979	95.25	0.9174	0.5448	0.4149	0.1777
	2	96.20	1.0324	0.5636	0.6378	0.2687	95.00	0.9375	0.5653	-0.0504	-0.0212
	4	96.80	1.2467	0.7860	13.3633	5.4209	94.10	1.0148	0.6598	-0.5944	-0.2314
	10	87.65	1.6793	2.7032	43.9206	15.7810	93.50	1.1955	0.9571	-1.5327	-0.4959
(100,50)	0.1	85.60	1.2638	1.7059	-35.3015	-16.4603	96.05	1.1321	0.7967	0.7102	0.2516
	0.25	95.25	1.0216	0.6160	-15.9067	-8.3472	96.20	0.9660	0.5723	0.3710	0.1551
	0.5	96.60	0.8814	0.4129	-7.0731	-3.7122	95.65	0.8717	0.4604	0.1800	0.0839
	1	95.60	0.8047	0.3797	-2.8271	-1.4661	95.65	0.7892	0.3790	0.2115	0.1086
	2	96.00	0.8087	0.3671	1.8916	0.9919	96.05	0.7443	0.3420	-0.5400	-0.2920
	4	94.95	0.8946	0.4978	12.4619	6.7323	95.80	0.7721	0.3608	-0.8321	-0.4384
	10	75.65	1.1058	1.7532	37.3435	19.7133	95.70	0.8869	0.4678	-1.1479	-0.5313
(100,100)	0.1	72.65	0.8258	1.1331	-30.0538	-19.8175	97.15	0.8328	0.4308	0.1660	0.0799
	0.25	90.50	0.7083	0.4436	-15.5163	-10.8907	97.00	0.7092	0.3054	-0.3012	-0.1723
	0.5	93.75	0.6527	0.3045	-8.8881	-5.9177	95.75	0.6630	0.2769	-0.5878	-0.3534
	1	94.40	0.6527	0.2857	-5.3476	-3.3341	95.80	0.6444	0.2580	-0.7023	-0.4376
	2	96.7	0.7252	0.2823	0.2800	0.1666	96.40	0.6644	0.2766	-0.6673	-0.4014
	4	94.35	0.8722	0.4858	13.4150	7.6684	95.45	0.7182	0.3356	-0.8940	-0.4885
	10	65.90	1.1601	2.3512	44.1978	22.1553	94.75	0.8599	0.4993	-1.9630	-0.8817

Table 4.15: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$	%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$
(30,30)	0.1	93.30	3.7625	11.8639	-5.3198	-0.5596	96.25	3.6970	8.4007	0.1605	0.0175
	0.25	96.60	3.1981	5.8124	1.8921	0.2561	96.45	3.3426	6.6453	0.2420	0.0297
	0.5	87.80	2.5097	7.2363	6.3619	1.1263	96.35	2.9671	5.4392	0.6174	0.0840
	1	76.00	1.8214	6.3098	6.8296	1.6832	96.35	2.3253	3.5156	0.5781	0.0979
	2	79.90	1.2329	2.5067	4.1567	1.4890	95.60	1.5016	1.4495	0.4478	0.1184
	4	91.85	0.7746	0.5367	1.4251	0.7800	95.75	0.8383	0.4529	0.2388	0.1129
	10	95.10	0.3667	0.0901	-0.1801	-0.1932	95.35	0.3499	0.0809	0.0384	0.0427
(30,50)	0.1	91.40	3.1698	9.1458	-4.0945	-0.4736	96.15	3.4310	7.4674	-0.0844	-0.0098
	0.25	93.85	2.7039	5.2064	2.7592	0.4137	95.75	3.1389	6.2056	0.0900	0.0114
	0.5	83.75	2.1764	6.8965	6.5615	1.2887	95.60	2.7659	4.9367	0.4047	0.0577
	1	69.75	1.6297	5.8986	6.7422	1.8325	95.70	2.0983	2.8468	0.5241	0.0987
	2	79.50	1.1363	2.1994	3.9450	1.5552	95.25	1.3028	1.0950	0.3688	0.1121
	4	93.80	0.7298	0.4193	1.2062	0.7288	95.10	0.7056	0.3298	0.1822	0.1008
	10	94.65	0.3497	0.0806	-0.3184	-0.3792	94.85	0.2903	0.0570	0.0589	0.0782
(50,50)	0.1	88.60	2.9379	9.0047	-5.9697	-0.8091	96.00	2.9028	5.1062	-0.2183	-0.0306
	0.25	95.95	2.4658	3.6637	1.7265	0.2975	96.25	2.5884	4.1952	0.2402	0.0371
	0.5	81.45	1.9362	5.6236	6.0630	1.3735	96.25	2.3089	3.4123	0.3184	0.0546
	1	59.75	1.4045	5.4056	6.6353	2.0948	96.40	1.8090	2.0076	0.3862	0.0865
	2	67.15	0.9484	2.1110	4.0516	1.8695	95.80	1.1664	0.8673	0.2695	0.0919
	4	87.75	0.5955	0.3876	1.3777	0.9794	95.70	0.6504	0.2688	0.1085	0.0663
	10	95.00	0.2845	0.0547	-0.1836	-0.2562	95.00	0.2727	0.0470	0.0238	0.0348
(100,50)	0.1	82.95	2.8100	10.5964	-8.1556	-1.2981	97.05	2.3810	3.2148	-0.1561	-0.0275
	0.25	98.25	2.3327	2.4890	0.5355	0.1079	96.65	2.0635	2.4993	0.1446	0.0289
	0.5	82.70	1.7817	4.5267	5.5676	1.4735	96.30	1.8708	2.0712	0.2697	0.0593
	1	49.05	1.2435	5.1202	6.6184	2.4325	96.10	1.5446	1.4735	0.2751	0.0718
	2	49.15	0.8066	2.1706	4.2674	2.2819	95.65	1.0505	0.7183	0.1954	0.0731
	4	75.45	0.4898	0.4178	1.6413	1.3471	95.60	0.6039	0.2407	0.0749	0.0483
	10	94.20	0.2295	0.0390	0.0091	0.0145	94.75	0.2580	0.0453	0.0056	0.0083
(100,100)	0.1	80.70	2.0914	6.4181	-6.0694	-1.1604	95.95	2.0525	2.5777	-0.0174	-0.0034
	0.25	95.00	1.7385	1.9378	1.6869	0.4148	95.60	1.8359	2.1002	0.1593	0.0348
	0.5	63.65	1.3629	4.5897	6.0410	1.9696	95.95	1.6391	1.7182	0.1830	0.0442
	1	26.15	0.9883	4.8692	6.6267	3.0306	96.65	1.2841	1.0207	0.1653	0.0518
	2	36.75	0.6670	1.8656	4.0534	2.7160	96.00	0.8273	0.4218	0.1241	0.0605
	4	78.65	0.4189	0.2855	1.3833	1.4256	96.15	0.4587	0.1334	0.0674	0.0584
	10	94.25	0.2015	0.0280	-0.1815	-0.3650	95.65	0.1929	0.0230	0.0245	0.0511

Table 4.15: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	93.30	0.3763	0.1186	-0.5320	-0.5596	96.25	0.3697	0.0840	0.0161	0.0175
	0.25	96.60	0.7995	0.3633	0.4730	0.2561	96.45	0.8357	0.4153	0.0605	0.0297
	0.5	87.80	1.2548	1.8091	3.1809	1.1263	96.35	1.4835	1.3598	0.3087	0.0840
	1	76.00	1.8214	6.3098	6.8296	1.6832	96.35	2.3253	3.5156	0.5781	0.0979
	2	79.90	2.4658	10.0267	8.3134	1.4890	95.60	3.0032	5.7982	0.8955	0.1184
	4	91.85	3.0986	8.5879	5.7003	0.7800	95.75	3.3533	7.2464	0.9553	0.1129
	10	95.10	3.6668	9.0123	-1.8012	-0.1932	95.35	3.4990	8.0919	0.3836	0.0427
(30,50)	0.1	91.40	0.3170	0.0915	-0.4095	-0.4736	96.15	0.3431	0.0747	-0.0084	-0.0098
	0.25	93.85	0.6760	0.3254	0.6898	0.4137	95.75	0.7847	0.3879	0.0225	0.0114
	0.5	83.75	1.0882	1.7241	3.2808	1.2887	95.60	1.3830	1.2342	0.2024	0.0577
	1	69.75	1.6297	5.8986	6.7422	1.8325	95.70	2.0983	2.8468	0.5241	0.0987
	2	79.50	2.2725	8.7977	7.8899	1.5552	95.25	2.6056	4.3798	0.7376	0.1121
	4	93.80	2.9191	6.7092	4.8249	0.7288	95.10	2.8226	5.2764	0.7287	0.1008
	10	94.65	3.4965	8.0573	-3.1837	-0.3792	94.85	2.9026	5.6973	0.5887	0.0782
(50,50)	0.1	88.60	0.2938	0.0900	-0.5970	-0.8091	96.00	0.2903	0.0511	-0.0218	-0.0306
	0.25	95.95	0.6165	0.2290	0.4316	0.2975	96.25	0.6471	0.2622	0.0601	0.0371
	0.5	81.45	0.9681	1.4059	3.0315	1.3735	96.25	1.1544	0.8531	0.1592	0.0546
	1	59.75	1.4045	5.4056	6.6353	2.0948	96.40	1.8090	2.0076	0.3862	0.0865
	2	67.15	1.8969	8.4439	8.1032	1.8695	95.80	2.3327	3.4692	0.5391	0.0919
	4	87.75	2.3820	6.2017	5.5109	0.9794	95.70	2.6016	4.3004	0.4342	0.0663
	10	95.00	2.8453	5.4703	-1.8363	-0.2562	95.00	2.7270	4.6968	0.2383	0.0348
(100,50)	0.1	82.95	0.2810	0.1060	-0.8156	-1.2981	97.05	0.2381	0.0321	-0.0156	-0.0275
	0.25	98.25	0.5832	0.1556	0.1339	0.1079	96.65	0.5159	0.1562	0.0361	0.0289
	0.5	82.70	0.8908	1.1317	2.7838	1.4735	96.30	0.9354	0.5178	0.1348	0.0593
	1	49.05	1.2435	5.1202	6.6184	2.4325	96.10	1.5446	1.4735	0.2751	0.0718
	2	49.15	1.6132	8.6824	8.5348	2.2819	95.65	2.1009	2.8732	0.3907	0.0731
	4	75.45	1.9590	6.6846	6.5654	1.3471	95.60	2.4155	3.8512	0.2996	0.0483
	10	94.20	2.2946	3.9046	0.0906	0.0145	94.75	2.5803	4.5255	0.0562	0.0083
(100,100)	0.1	80.70	0.2091	0.0642	-0.6069	-1.1604	95.95	0.2052	0.0258	-0.0017	-0.0034
	0.25	95.00	0.4346	0.1211	0.4217	0.4148	95.60	0.4590	0.1313	0.0398	0.0348
	0.5	63.65	0.6815	1.1474	3.0205	1.9696	95.95	0.8196	0.4295	0.0915	0.0442
	1	26.15	0.9883	4.8692	6.6267	3.0306	96.65	1.2841	1.0207	0.1653	0.0518
	2	36.75	1.3340	7.4624	8.1067	2.7160	96.00	1.6545	1.6873	0.2483	0.0605
	4	78.65	1.6754	4.5677	5.5334	1.4256	96.15	1.8350	2.1351	0.2697	0.0584
	10	94.25	2.0153	2.8019	-1.8153	-0.3650	95.65	1.9287	2.2973	0.2447	0.0511

Table 4.16: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model e) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-1}$	$cover$ %	$width$ $\times 10^0$	MSE $\times 10^{-1}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{-1}$
(30,30)	0.1	84.65	1.5794	2.7340	-43.8277	-15.366	92.35	1.4042	1.5142	1.3317	0.3424
	0.25	91.45	1.3834	1.5452	-29.833	-11.6521	95.15	1.2491	0.9346	0.4613	0.1509
	0.5	95.75	1.3033	1.0081	-18.6488	-7.2556	95.70	1.1163	0.7273	0.6250	0.2318
	1	98.30	1.3129	0.7450	-5.9903	-2.2490	95.95	1.0577	0.6603	0.7084	0.2757
	2	98.45	1.4373	0.8565	9.2589	3.3341	95.55	1.0932	0.7041	0.3392	0.1278
	4	95.30	1.6936	1.6561	28.2601	9.6485	94.40	1.2114	0.8763	-0.0966	-0.0326
	10	84.40	2.2563	4.8837	60.8624	17.7176	87.25	1.3119	1.4001	-1.8922	-0.5062
(30,50)	0.1	74.25	1.1730	2.1648	-40.6494	-17.9532	94.35	1.2008	0.9756	0.3760	0.1204
	0.25	87.10	1.0905	1.2921	-29.1715	-13.8861	95.70	1.0145	0.6523	-0.4524	-0.1771
	0.5	94.55	1.0900	0.8332	-18.9543	-8.7045	95.85	0.9362	0.5354	-1.1311	-0.4893
	1	98.30	1.1735	0.5901	-6.3474	-2.7062	95.35	0.9392	0.5309	-1.2613	-0.5481
	2	98.35	1.3675	0.7513	9.8684	3.8581	95.20	1.0224	0.6250	-1.5402	-0.6171
	4	94.85	1.6934	1.7636	30.9662	10.9136	93.00	1.1734	0.8599	-1.6098	-0.5497
	10	81.10	2.3685	5.9124	68.2874	19.3158	87.40	1.2992	1.3335	-4.2094	-1.1602
(50,50)	0.1	70.75	1.1965	2.4860	-44.8975	-20.7008	93.75	1.198	0.9390	1.1166	0.3645
	0.25	84.65	1.0590	1.3310	-30.7685	-15.6908	95.90	0.9714	0.5941	1.4166	0.5820
	0.5	92.80	1.0018	0.7621	-19.2334	-9.7102	96.45	0.8547	0.4543	0.7163	0.3362
	1	97.80	1.0098	0.4627	-6.1032	-2.9582	96.40	0.8119	0.3932	0.6644	0.3352
	2	97.40	1.1021	0.5574	9.6181	4.4598	96.00	0.8417	0.4306	0.3866	0.1863
	4	91.95	1.2861	1.3542	28.9271	12.7142	95.25	0.9404	0.5510	0.1489	0.0634
	10	67.70	1.6720	4.4610	61.2933	23.0937	93.20	1.1527	0.8576	-1.0707	-0.3658
(100,50)	0.1	59.15	1.2486	3.3086	-53.5909	-25.6411	94.35	1.1643	0.8555	0.6691	0.2288
	0.25	78.15	1.0498	1.5762	-35.3537	-19.5661	96.60	0.9232	0.5033	0.3884	0.1731
	0.5	90.35	0.9373	0.7769	-21.7475	-12.4710	96.80	0.7806	0.3617	-0.0392	-0.0206
	1	97.80	0.8741	0.3579	-7.6614	-4.4284	96.10	0.6980	0.2806	-0.3143	-0.1876
	2	97.70	0.8727	0.3543	7.6917	4.4761	96.55	0.6799	0.2652	-0.4216	-0.2589
	4	87.05	0.9374	0.9232	25.0883	14.6326	96.40	0.7210	0.3028	-0.3433	-0.1972
	10	47.15	1.1116	3.0616	52.1628	28.2552	95.55	0.8657	0.4346	-1.1966	-0.5748
(100,100)	0.1	29.20	0.8307	2.4479	-47.1674	-31.5716	96.95	0.8705	0.4623	0.2025	0.0942
	0.25	58.65	0.7414	1.2712	-32.9559	-24.2160	96.90	0.6846	0.28580	-0.0793	-0.0469
	0.5	83.55	0.7040	0.6384	-21.1282	-15.2450	96.45	0.5994	0.2118	-0.4908	-0.3374
	1	96.40	0.7110	0.2694	-7.5639	-5.1909	95.60	0.5679	0.1934	-0.6329	-0.4555
	2	96.90	0.7749	0.3114	8.6459	5.6186	96.00	0.5892	0.2165	-0.7619	-0.5184
	4	80.95	0.9002	1.0749	28.3779	17.2786	95.70	0.6628	0.2869	-1.2978	-0.7683
	10	31.35	1.1541	4.0884	60.9375	31.4597	94.75	0.8353	0.4690	-1.8956	-0.8785

Table 4.16: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	94.45	3.6676	9.3283	-1.7654	-0.1859	94.65	3.6734	8.6588	0.5762	0.0620
	0.25	95.05	2.8618	4.9513	2.2101	0.3307	95.80	3.2087	6.7383	0.5530	0.0675
	0.5	95.05	2.2033	3.3906	3.2368	0.6685	96.05	2.6582	4.7096	0.4055	0.0592
	1	94.95	1.6109	1.9496	2.7951	0.8175	95.85	1.9936	2.6576	0.2345	0.0455
	2	95.60	1.1228	0.8447	1.6621	0.6969	95.70	1.3389	1.1694	0.1894	0.0555
	4	96.20	0.7410	0.3022	0.5906	0.3612	95.30	0.8100	0.4453	0.1367	0.0649
	10	95.95	0.3835	0.0901	-0.1759	-0.1886	93.45	0.3638	0.0948	0.1084	0.112
(30,50)	0.1	93.65	3.0275	6.6064	-0.2358	-0.0290	95.60	3.4161	7.3158	0.1784	0.0209
	0.25	93.60	2.4182	4.2276	2.9441	0.5077	96.15	2.9372	5.5206	0.1945	0.0262
	0.5	93.40	1.9251	3.0425	3.5009	0.8211	96.15	2.4167	3.7875	0.1726	0.0280
	1	94.60	1.4621	1.6900	2.7579	0.9044	95.70	1.7957	2.1103	0.1548	0.0337
	2	96.50	1.0559	0.6789	1.4800	0.6900	95.25	1.1932	0.9354	0.1639	0.0537
	4	98.10	0.7163	0.2323	0.3761	0.2546	95.65	0.7163	0.3402	0.1621	0.0882
	10	96.65	0.3781	0.0830	-0.3452	-0.4093	92.50	0.3157	0.0697	0.1404	0.1706
(50,50)	0.1	94.65	2.8388	5.3480	-1.7334	-0.2439	95.40	2.9183	5.0071	0.3741	0.0529
	0.25	95.45	2.1982	2.9444	2.2360	0.4521	96.50	2.5054	3.7497	0.4524	0.0741
	0.5	93.45	1.6926	2.3233	3.2530	0.9143	96.90	2.0735	2.5674	0.3062	0.0605
	1	93.45	1.2376	1.4178	2.8126	1.1232	96.85	1.5568	1.4224	0.1854	0.0492
	2	94.35	0.8609	0.5913	1.6868	0.9628	96.50	1.0454	0.6379	0.1664	0.0660
	4	95.70	0.5671	0.1849	0.6194	0.5115	96.25	0.6345	0.2389	0.1285	0.0834
	10	96.40	0.2958	0.0513	-0.1532	-0.2190	95.40	0.2938	0.0538	0.0780	0.1069
(100,50)	0.1	93.60	2.7958	5.9456	-4.3009	-0.6719	96.30	2.4286	3.4120	0.3943	0.0676
	0.25	97.70	2.1016	2.0590	0.9127	0.2053	97.10	2.1002	2.7080	0.3151	0.0606
	0.5	94.60	1.5565	1.7234	2.6724	0.8410	96.45	1.7644	1.9434	0.1577	0.0358
	1	90.05	1.0839	1.2343	2.7297	1.2340	96.55	1.3428	1.1016	0.1542	0.0465
	2	88.55	0.7162	0.5801	1.8673	1.2272	96.35	0.9121	0.5140	0.1225	0.0541
	4	91.40	0.4508	0.1844	0.8859	0.8608	95.95	0.5593	0.1983	0.0711	0.0506
	10	95.00	0.2265	0.0342	0.0758	0.1306	95.80	0.2615	0.0442	0.0434	0.0654
(100,100)	0.1	93.25	2.0006	2.9564	-1.8302	-0.3574	96.45	2.0600	2.5771	0.2707	0.0534
	0.25	93.50	1.5446	1.7403	2.1481	0.6005	96.50	1.7740	2.0174	0.2506	0.0559
	0.5	87.20	1.1902	1.6752	3.1782	1.2320	96.40	1.4725	1.3711	0.0972	0.0262
	1	83.55	0.8701	1.0919	2.7619	1.5221	96.25	1.1079	0.7822	0.0484	0.0173
	2	87.65	0.6044	0.4361	1.6630	1.3164	96.30	0.7445	0.3548	0.0478	0.0254
	4	93.85	0.3974	0.1129	0.6154	0.7104	96.00	0.4514	0.1313	0.0666	0.0582
	10	95.70	0.2078	0.0267	-0.1474	-0.2974	95.35	0.2090	0.0277	0.0488	0.0931

Table 4.16: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	94.45	0.3668	0.0933	-0.1765	-0.1859	94.65	0.3673	0.0866	0.0576	0.0620
	0.25	95.05	0.7155	0.3095	0.5525	0.3307	95.80	0.8022	0.4211	0.1382	0.0675
	0.5	95.05	1.1017	0.8476	1.6184	0.6685	96.05	1.3291	1.1774	0.2027	0.0592
	1	94.95	1.6109	1.9496	2.7951	0.8175	95.85	1.9936	2.6576	0.2345	0.0455
	2	95.60	2.2456	3.3788	3.3241	0.6969	95.70	2.6779	4.6774	0.3789	0.0555
	4	96.20	2.964	4.8350	2.3625	0.3612	95.30	3.2399	7.1255	0.5469	0.0649
	10	95.95	3.8347	9.0101	-1.7593	-0.1886	93.45	3.6381	9.4756	1.0841	0.1120
(30,50)	0.1	93.65	0.3028	0.0661	-0.0236	-0.0290	95.60	0.3416	0.0732	0.0178	0.0209
	0.25	93.60	0.6046	0.2642	0.7360	0.5077	96.15	0.7343	0.3450	0.0486	0.0262
	0.5	93.40	0.9625	0.7606	1.7504	0.8211	96.15	1.2084	0.9469	0.0863	0.0280
	1	94.60	1.4621	1.6900	2.7579	0.9044	95.70	1.7957	2.1103	0.1548	0.0337
	2	96.50	2.1117	2.7156	2.9601	0.6900	95.25	2.3864	3.7414	0.3278	0.0537
	4	98.10	2.8652	3.7168	1.5045	0.2546	95.65	2.8650	5.4432	0.6483	0.0882
	10	96.65	3.7813	8.2994	-3.4515	-0.4093	92.50	3.1574	6.9718	1.4045	0.1706
(50,50)	0.1	94.65	0.2839	0.0535	-0.1733	-0.2439	95.40	0.2918	0.0501	0.0374	0.0529
	0.25	95.45	0.5496	0.1840	0.5590	0.4521	96.50	0.6264	0.2344	0.1131	0.0741
	0.5	93.45	0.8463	0.5808	1.6265	0.9143	96.90	1.0367	0.6419	0.1531	0.0605
	1	93.45	1.2376	1.4178	2.8126	1.1232	96.85	1.5568	1.4224	0.1854	0.0492
	2	94.35	1.7219	2.3652	3.3736	0.9628	96.50	2.0908	2.5516	0.3328	0.0660
	4	95.70	2.2685	2.9582	2.4775	0.5115	96.25	2.5378	3.8219	0.5142	0.0834
	10	96.40	2.9575	5.1261	-1.5323	-0.2190	95.40	2.9381	5.3798	0.7796	0.1069
(100,50)	0.1	93.60	0.2796	0.0595	-0.4301	-0.6719	96.30	0.2429	0.0341	0.0394	0.0676
	0.25	97.70	0.5254	0.1287	0.2282	0.2053	97.10	0.5251	0.1693	0.0788	0.0606
	0.5	94.60	0.7782	0.4308	1.3362	0.8410	96.45	0.8822	0.4859	0.0789	0.0358
	1	90.05	1.0839	1.2343	2.7297	1.2340	96.55	1.3428	1.1016	0.1542	0.0465
	2	88.55	1.4324	2.3204	3.7347	1.2272	96.35	1.8243	2.0561	0.2449	0.0541
	4	91.40	1.8032	2.9496	3.5437	0.8608	95.95	2.2370	3.1721	0.2845	0.0506
	10	95.00	2.2646	3.4231	0.7577	0.1306	95.80	2.6148	4.4207	0.4343	0.0654
(100,100)	0.1	93.25	0.2001	0.0296	-0.1830	-0.3574	96.45	0.2060	0.0258	0.0271	0.0534
	0.25	93.50	0.3862	0.1088	0.5370	0.6005	96.50	0.4435	0.1261	0.0626	0.0559
	0.5	87.20	0.5951	0.4188	1.5891	1.2320	96.40	0.7363	0.3428	0.0486	0.0262
	1	83.55	0.8701	1.0919	2.7619	1.5221	96.25	1.1079	0.7822	0.0484	0.0173
	2	87.65	1.2087	1.7443	3.326	1.3164	96.30	1.4890	1.4192	0.0956	0.0254
	4	93.85	1.5895	1.8061	2.4617	0.7104	96.00	1.8058	2.1010	0.2664	0.0582
	10	95.70	2.0780	2.6719	-1.4736	-0.2974	95.35	2.0899	2.7668	0.4879	0.0931

Table 4.17: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model f) with AUC = 0.70.

		c_{GS}									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	77.80	1.5492	4.7099	4.7379	0.9540	88.40	2.0022	3.2733	0.9824	0.1743
	0.25	82.70	1.4495	2.8807	2.7331	0.5915	93.60	2.1500	3.2084	0.5124	0.0908
	0.5	87.05	1.3913	2.0461	0.1611	0.0356	94.85	2.1826	3.3764	0.2331	0.0401
	1	82.15	1.4555	3.0051	-3.0734	-0.6769	94.30	2.2407	3.5751	-0.0444	-0.0074
	2	75.60	1.7642	6.0258	-6.1135	-1.2777	94.70	2.3776	3.9254	-0.5488	-0.0879
	4	86.65	2.4062	8.0778	-7.2198	-1.3484	94.50	2.4558	4.2159	-0.8270	-0.1284
	10	97.35	3.8348	5.7594	-2.7474	-0.3883	92.60	2.3305	3.7766	-1.1373	-0.1883
(30,50)	0.1	74.15	1.2688	3.4165	4.4820	1.1943	92.60	1.6757	1.9106	0.5168	0.1190
	0.25	78.10	1.1829	2.4265	3.1879	0.8487	95.00	1.7739	2.1509	0.5569	0.1209
	0.5	85.80	1.1665	1.5632	0.7098	0.1824	95.45	1.8709	2.3590	0.1819	0.0375
	1	83.05	1.2827	2.2456	-2.4193	-0.5936	94.70	2.0473	2.7107	0.0074	0.0014
	2	75.70	1.6237	5.0682	-5.5420	-1.2399	94.70	2.2731	3.4536	-0.2942	-0.0501
	4	82.75	2.2506	7.9491	-7.1332	-1.3333	93.10	2.4114	4.1221	-0.7016	-0.1099
	10	96.25	3.4845	6.1990	-3.8310	-0.5568	91.80	2.3072	3.7648	-1.1892	-0.1975
(50,50)	0.1	69.60	1.1968	3.3214	4.4598	1.2215	93.85	1.6437	1.8854	0.4728	0.1095
	0.25	81.85	1.1125	1.9085	2.6135	0.7464	94.55	1.7066	1.9197	0.2775	0.0635
	0.5	87.45	1.0672	1.1728	0.1080	0.0315	95.50	1.7430	1.9396	0.1145	0.0260
	1	77.35	1.1145	2.1214	-3.0563	-0.8868	94.85	1.7845	2.0896	0.0386	0.0084
	2	62.40	1.3432	4.9795	-6.0353	-1.6502	95.05	1.8926	2.4437	-0.2927	-0.0593
	4	75.85	1.8210	6.6876	-7.0668	-1.7168	95.00	1.9544	2.6088	-0.3912	-0.0768
	10	97.65	2.7904	3.5889	-2.8632	-0.5440	94.20	1.8117	2.1339	-0.6740	-0.1475
(100,50)	0.1	66.05	1.1307	3.4061	4.4628	1.1863	92.65	1.6204	1.8461	0.6401	0.1506
	0.25	80.70	1.0525	1.7560	2.3072	0.6594	94.80	1.6564	1.8983	0.4299	0.0991
	0.5	86.80	0.9881	1.0620	-0.3519	-0.1086	94.95	1.6223	1.7812	0.2880	0.0684
	1	68.05	0.9714	2.1512	-3.5409	-1.1818	95.20	1.5616	1.6946	0.1720	0.0418
	2	39.85	1.0815	4.8522	-6.3433	-2.2032	95.20	1.5030	1.4702	-0.0170	-0.0044
	4	56.25	1.3897	5.8414	-7.0139	-2.3093	94.85	1.4476	1.3870	-0.1957	-0.0526
	10	98.20	2.1024	1.8009	-1.8408	-0.4813	95.50	1.3076	1.0823	-0.3756	-0.1149
(100,100)	0.1	50.45	0.8412	2.6151	4.4112	1.7047	95.75	1.2063	0.9362	0.2455	0.0805
	0.25	71.50	0.7803	1.2987	2.6181	1.0569	95.50	1.2320	0.9721	0.2808	0.0904
	0.5	87.20	0.7484	0.5783	0.1521	0.0634	95.45	1.2511	1.0143	0.1444	0.0454
	1	66.15	0.7802	1.4536	-2.9791	-1.2518	95.55	1.2931	1.0483	0.0867	0.0268
	2	33.35	0.9362	4.1519	-5.9447	-2.3909	96.15	1.3564	1.1477	-0.0906	-0.0267
	4	45.85	1.2542	5.8412	-7.1079	-2.5299	94.95	1.3925	1.2507	-0.2593	-0.0735
	10	95.30	1.9111	2.2768	-2.9324	-0.7788	94.35	1.2868	1.1189	-0.3401	-0.1022

Table 4.17: *Continued*.

		$p(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	88.00	3.1619	11.3943	2.8517	0.2772	94.15	3.8785	10.2387	0.4685	0.0463
	0.25	89.80	2.9053	7.9723	1.0879	0.1227	96.20	3.5756	8.0206	0.1841	0.0206
	0.5	91.60	2.5044	5.3553	0.3026	0.0414	95.80	3.0569	5.7981	-0.0146	-0.0019
	1	93.90	1.9848	2.9553	0.4225	0.0779	96.25	2.4344	3.6680	0.0811	0.0134
	2	95.45	1.4131	1.3373	0.9628	0.2728	96.40	1.7499	1.9785	0.2364	0.0532
	4	96.00	0.8583	0.4821	1.0000	0.5114	95.65	1.0228	0.7081	0.2547	0.0961
	10	97.10	0.3169	0.0676	0.3866	0.5326	95.40	0.3551	0.0925	0.1814	0.1920
(30,50)	0.1	90.70	2.8480	7.7725	1.3652	0.1567	96.10	3.5179	8.2284	0.2436	0.0269
	0.25	91.60	2.6454	5.7814	0.1014	0.0133	96.60	3.2369	6.6597	0.1602	0.0196
	0.5	92.45	2.3026	4.2511	-0.5261	-0.0809	96.10	2.8077	4.8971	-0.2063	-0.0295
	1	94.10	1.8391	2.4444	-0.0474	-0.0096	95.90	2.2778	3.2513	-0.1593	-0.0279
	2	95.70	1.3129	1.1573	0.7786	0.2350	95.65	1.6491	1.7474	0.1002	0.0240
	4	94.50	0.7979	0.4837	1.0061	0.5143	94.30	0.9595	0.6523	0.2521	0.0992
	10	95.40	0.3004	0.0725	0.4627	0.6472	94.30	0.3374	0.0879	0.2221	0.2437
(50,50)	0.1	89.05	2.4772	6.4877	2.3579	0.3061	96.00	3.1219	6.2066	0.3822	0.0486
	0.25	90.65	2.2487	4.6268	0.7459	0.1103	96.60	2.8052	4.8865	0.0941	0.0135
	0.5	91.80	1.9306	3.1074	0.0733	0.0131	95.85	2.3759	3.6149	0.0546	0.0091
	1	93.60	1.5282	1.7203	0.2741	0.0662	96.20	1.8970	2.3198	0.1293	0.0269
	2	95.45	1.0889	0.8076	0.8713	0.3220	95.45	1.3745	1.2466	0.1899	0.0539
	4	93.80	0.6662	0.3302	0.9416	0.6056	95.50	0.8133	0.4592	0.1945	0.0911
	10	96.25	0.2503	0.0441	0.3729	0.6782	95.55	0.2784	0.0554	0.1226	0.1670
(100,50)	0.1	84.75	2.1361	6.0423	3.3503	0.4775	94.40	2.7648	5.0795	0.6478	0.0912
	0.25	88.25	1.8933	3.7991	1.4312	0.2387	95.80	2.4052	3.7830	0.3133	0.0510
	0.5	90.80	1.5971	2.3429	0.4689	0.0973	95.55	1.9754	2.5625	0.1571	0.0310
	1	92.80	1.2451	1.2130	0.3948	0.1141	95.50	1.5276	1.4923	0.0684	0.0177
	2	94.05	0.8788	0.5349	0.7992	0.3682	95.85	1.0972	0.7455	0.0033	0.0012
	4	95.15	0.5366	0.2024	0.8110	0.6938	96.15	0.6507	0.2622	0.0510	0.0315
	10	96.60	0.2023	0.0254	0.2985	0.7347	96.20	0.2238	0.0335	0.0549	0.0952
(100,100)	0.1	88.40	1.7642	3.2450	1.9332	0.3607	96.30	2.2908	3.1494	0.0784	0.0140
	0.25	91.00	1.5914	2.2465	0.4757	0.1008	96.35	2.0033	2.4070	0.1561	0.0318
	0.5	92.00	1.3632	1.5173	-0.1015	-0.0261	96.00	1.6797	1.6875	0.1436	0.0350
	1	93.85	1.0775	0.8408	0.1711	0.0591	96.10	1.3418	1.0700	0.1038	0.0317
	2	94.50	0.7668	0.4251	0.8171	0.4316	96.00	0.9795	0.6051	0.1518	0.0618
	4	92.10	0.4706	0.1999	0.9304	0.8737	95.75	0.5873	0.2276	0.1286	0.0856
	10	93.50	0.1800	0.0274	0.3587	0.9400	95.25	0.1982	0.0278	0.0615	0.1175

Table 4.17: *Continued*.

		$q(c_{GS})$									
$AUC = 0.70$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	88.00	0.3162	0.1139	0.2852	0.2772	94.15	0.3878	0.1024	0.0468	0.0463
	0.25	89.80	0.7263	0.4983	0.2720	0.1227	96.20	0.8939	0.5013	0.0460	0.0206
	0.5	91.60	1.2522	1.3388	0.1513	0.0414	95.80	1.5285	1.4495	-0.0073	-0.0019
	1	93.90	1.9848	2.9553	0.4225	0.0779	96.25	2.4344	3.6680	0.0811	0.0134
	2	95.45	2.8262	5.3492	1.9255	0.2728	96.40	3.4999	7.9141	0.4729	0.0532
	4	96.00	3.4334	7.7141	3.9999	0.5114	95.65	4.0910	11.3291	1.0189	0.0961
	10	97.10	3.1690	6.7601	3.8659	0.5326	95.40	3.5512	9.2517	1.8143	0.1920
(30,50)	0.1	90.70	0.2848	0.0777	0.1365	0.1567	96.10	0.3518	0.0823	0.0244	0.0269
	0.25	91.60	0.6614	0.3613	0.0254	0.0133	96.60	0.8092	0.4162	0.0401	0.0196
	0.5	92.45	1.1513	1.0628	-0.2631	-0.0809	96.10	1.4039	1.2243	-0.1032	-0.0295
	1	94.10	1.8391	2.4444	-0.0474	-0.0096	95.90	2.2778	3.2513	-0.1593	-0.0279
	2	95.70	2.6257	4.6293	1.5571	0.2350	95.65	3.2982	6.9896	0.2004	0.0240
	4	94.50	3.1914	7.7391	4.0243	0.5143	94.30	3.8380	10.4370	1.0084	0.0992
	10	95.40	3.0038	7.2503	4.6274	0.6472	94.30	3.3742	8.7944	2.2205	0.2437
(50,50)	0.1	89.05	0.2477	0.0649	0.2358	0.3061	96.00	0.3122	0.0621	0.0382	0.0486
	0.25	90.65	0.5622	0.2892	0.1865	0.1103	96.60	0.7013	0.3054	0.0235	0.0135
	0.5	91.80	0.9653	0.7769	0.0366	0.0131	95.85	1.1879	0.9037	0.0273	0.0091
	1	93.60	1.5282	1.7203	0.2741	0.0662	96.20	1.8970	2.3198	0.1293	0.0269
	2	95.45	2.1777	3.2304	1.7426	0.3220	95.45	2.7490	4.9863	0.3798	0.0539
	4	93.80	2.6650	5.2839	3.7663	0.6056	95.50	3.2531	7.3466	0.7781	0.0911
	10	96.25	2.5033	4.4124	3.7292	0.6782	95.55	2.7838	5.5387	1.2265	0.1670
(100,50)	0.1	84.75	0.2136	0.0604	0.3350	0.4775	94.40	0.2765	0.0508	0.0648	0.0912
	0.25	88.25	0.4733	0.2374	0.3578	0.2387	95.80	0.6013	0.2364	0.0783	0.0510
	0.5	90.80	0.7985	0.5857	0.2344	0.0973	95.55	0.9877	0.6406	0.0786	0.0310
	1	92.80	1.2451	1.2130	0.3948	0.1141	95.50	1.5276	1.4923	0.0684	0.0177
	2	94.05	1.7576	2.1396	1.5985	0.3682	95.85	2.1945	2.9821	0.0065	0.0012
	4	95.15	2.1464	3.2378	3.2440	0.6938	96.15	2.6028	4.1955	0.2038	0.0315
	10	96.60	2.0231	2.5402	2.9845	0.7347	96.20	2.2378	3.3490	0.5487	0.0952
(100,100)	0.1	88.40	0.1764	0.0325	0.1933	0.3607	96.30	0.2291	0.0315	0.0078	0.0140
	0.25	91.00	0.3979	0.1404	0.1189	0.1008	96.35	0.5008	0.1504	0.0390	0.0318
	0.5	92.00	0.6816	0.3793	-0.0508	-0.0261	96.00	0.8399	0.4219	0.0718	0.0350
	1	93.85	1.0775	0.8408	0.1711	0.0591	96.10	1.3418	1.0700	0.1038	0.0317
	2	94.50	1.5337	1.7003	1.6342	0.4316	96.00	1.9590	2.4205	0.3035	0.0618
	4	92.10	1.8825	3.1984	3.7215	0.8737	95.75	2.3493	3.6416	0.5145	0.0856
	10	93.50	1.7997	2.7413	3.5865	0.9400	95.25	1.9821	2.7784	0.6154	0.1175

Table 4.18: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model f) with AUC = 0.80.

		c_{GS}									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$	%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$
(30,30)	0.1	61.95	1.6906	9.8413	7.7604	1.2555	85.65	2.5305	5.2137	1.3987	0.1974
	0.25	73.05	1.5353	5.0481	4.2308	0.7410	92.25	2.6348	5.2607	0.8887	0.1234
	0.5	83.35	1.4617	2.8562	-0.1151	-0.0215	94.15	2.5953	5.4378	0.2881	0.0391
	1	69.90	1.4933	4.9828	-4.9236	-0.9731	94.45	2.4630	4.7675	-0.6338	-0.0922
	2	61.10	1.7158	8.4090	-7.7123	-1.5542	94.25	2.3063	3.8510	-0.8372	-0.1361
	4	82.50	2.2219	8.1889	-7.6057	-1.5507	93.75	2.1731	3.0528	-0.9922	-0.1825
	10	97.15	3.3778	4.5953	-2.6143	-0.4179	91.00	2.0254	2.7089	-1.0496	-0.2058
(30,50)	0.1	49.30	1.3904	8.1771	7.7021	1.6252	92.65	2.0860	2.9824	0.6698	0.1236
	0.25	67.15	1.2614	4.2145	4.7723	1.0841	94.50	2.2180	3.3089	0.4529	0.0790
	0.5	84.90	1.2262	1.8128	0.6075	0.1441	96.10	2.2958	3.5734	0.2156	0.0361
	1	72.55	1.3034	3.4463	-4.1756	-1.0116	95.25	2.2655	3.4155	-0.2663	-0.0456
	2	60.05	1.5487	6.8912	-7.1543	-1.6988	95.70	2.1766	2.8335	-0.5852	-0.1106
	4	77.15	2.0028	7.6981	-7.4818	-1.6321	94.55	2.0969	2.7152	-0.7213	-0.1397
	10	94.75	2.9511	5.4671	-4.3724	-0.7331	91.05	2.0246	2.6992	-1.0871	-0.2139
(50,50)	0.1	49.00	1.2961	7.2513	7.1207	1.5244	92.95	2.0632	2.8648	0.5660	0.1063
	0.25	68.80	1.1783	3.3495	3.9340	0.9265	94.05	2.1246	3.1774	0.4304	0.0766
	0.5	84.10	1.1209	1.5692	-0.3154	-0.0798	94.05	2.1055	3.1156	0.1339	0.0240
	1	59.10	1.1431	3.9078	-5.0269	-1.3525	95.25	1.9715	2.6803	-0.3655	-0.0707
	2	41.00	1.3074	7.2995	-7.7318	-2.1264	95.10	1.7848	2.1614	-0.5817	-0.1261
	4	66.35	1.6656	7.1438	-7.5289	-1.9596	95.15	1.6652	1.7829	-0.6321	-0.1514
	10	96.75	2.4849	3.1691	-3.0291	-0.6382	93.70	1.5726	1.6083	-0.6380	-0.1611
(100,50)	0.1	47.55	1.2226	7.1451	7.0093	1.4832	92.50	2.0447	2.8597	0.7339	0.1385
	0.25	69.65	1.1154	3.0255	3.4835	0.8181	94.00	2.0683	3.0585	0.5928	0.1078
	0.5	81.55	1.0408	1.5546	-0.8958	-0.2332	94.20	1.9896	2.9531	0.3736	0.0689
	1	43.65	1.0099	4.2925	-5.5971	-1.6431	94.20	1.7277	2.1135	-0.0196	-0.0043
	2	18.45	1.0796	7.4584	-8.0837	-2.6590	94.00	1.4194	1.3677	-0.2707	-0.0734
	4	47.10	1.3117	6.2706	-7.3642	-2.5290	94.90	1.2332	1.0362	-0.2980	-0.0930
	10	98.35	1.9217	1.4488	-1.5174	-0.4346	95.15	1.1279	0.8022	-0.3328	-0.1183
(100,100)	0.1	25.90	0.9054	6.1571	7.0795	2.0915	93.90	1.4814	1.4968	0.3246	0.0842
	0.25	56.65	0.8248	2.4436	3.8595	1.2492	94.25	1.5445	1.6347	0.2164	0.0536
	0.5	81.65	0.7863	0.8366	-0.3891	-0.1357	95.15	1.5528	1.7147	0.0393	0.0095
	1	33.95	0.8027	3.3183	-5.0994	-1.9027	94.50	1.4415	1.4708	-0.2827	-0.0739
	2	10.40	0.9156	6.7640	-7.8059	-3.0132	95.30	1.2841	1.0568	-0.3315	-0.1025
	4	29.85	1.1569	6.5284	-7.6165	-2.8237	95.35	1.1734	0.8427	-0.3283	-0.1138
	10	94.10	1.6846	2.1072	-3.1555	-0.9462	95.10	1.1099	0.7740	-0.3197	-0.1156

Table 4.18: *Continued*.

		$p(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	80.15	3.1851	16.4073	6.6504	0.6073	93.95	4.0304	10.7976	0.5654	0.0545
	0.25	83.60	2.7451	10.0225	4.3849	0.4871	95.55	3.5318	8.1307	0.3299	0.0366
	0.5	88.05	2.3212	5.7789	2.4814	0.3452	95.40	2.9200	5.5000	0.0470	0.0063
	1	92.05	1.8577	2.9366	1.0470	0.1969	95.15	2.2563	3.2268	0.0053	0.0009
	2	94.90	1.3834	1.3144	0.4663	0.1297	95.45	1.6368	1.7897	0.0773	0.0183
	4	97.00	0.9230	0.4467	0.2618	0.1248	95.40	1.0601	0.7533	0.1750	0.0639
	10	98.45	0.4062	0.0863	0.0538	0.0581	94.00	0.4672	0.1619	0.1932	0.1536
(30,50)	0.1	85.65	2.9119	10.7978	4.4860	0.4785	96.20	3.6715	8.1900	-0.2107	-0.0233
	0.25	88.65	2.5344	6.8913	3.0462	0.3944	96.20	3.1965	6.0831	-0.1744	-0.0224
	0.5	91.35	2.1573	4.1536	1.6925	0.2721	96.55	2.6597	4.1245	-0.1053	-0.0164
	1	93.30	1.7322	2.2176	0.7159	0.1538	96.80	2.0845	2.5204	-0.0764	-0.0152
	2	95.40	1.2889	1.0432	0.4527	0.1415	95.90	1.5360	1.4207	0.0492	0.0131
	4	96.55	0.8568	0.4142	0.4221	0.2120	95.15	1.0066	0.6561	0.1444	0.0564
	10	97.55	0.3858	0.0887	0.2674	0.2960	92.80	0.4463	0.1528	0.1976	0.1619
(50,50)	0.1	79.50	2.4861	10.1163	5.5523	0.6619	96.10	3.2340	6.4651	-0.1744	-0.0217
	0.25	83.75	2.1328	6.0683	3.6591	0.5319	95.95	2.7786	4.9236	-0.2989	-0.0426
	0.5	88.60	1.7994	3.3655	1.9534	0.3575	96.15	2.2855	3.3302	-0.3051	-0.0529
	1	92.85	1.4366	1.6637	0.6930	0.1724	96.35	1.7577	1.9446	-0.2356	-0.0535
	2	95.60	1.0659	0.7384	0.2549	0.0942	96.35	1.2783	1.0233	-0.0267	-0.0083
	4	97.30	0.7098	0.2732	0.1699	0.1033	95.65	0.8383	0.4676	0.0867	0.0401
	10	98.20	0.3239	0.0495	0.1029	0.1477	95.15	0.3812	0.1020	0.1385	0.1384
(100,50)	0.1	67.40	2.0838	11.2649	7.5449	1.0105	95.30	2.8457	5.1814	0.5806	0.0809
	0.25	73.80	1.7491	5.9851	4.9687	0.8377	95.00	2.3866	3.6827	0.4380	0.0724
	0.5	83.20	1.4531	2.8926	2.7525	0.5955	95.10	1.9234	2.3647	0.2142	0.0441
	1	90.45	1.1461	1.2063	1.0401	0.3138	95.10	1.4453	1.3161	0.0242	0.0067
	2	94.90	0.8446	0.4688	0.2716	0.1264	96.05	1.0299	0.6773	0.0642	0.0247
	4	97.75	0.5627	0.1538	0.0130	0.0105	96.65	0.6747	0.3020	0.0298	0.0171
	10	99.10	0.2541	0.0263	-0.1103	-0.2201	96.40	0.3041	0.0608	0.0600	0.0772
(100,100)	0.1	70.65	1.7605	6.9854	5.7664	0.9529	95.65	2.3476	3.4284	0.0648	0.0111
	0.25	77.60	1.5031	3.8730	3.8671	0.7929	95.35	1.9768	2.4364	0.0517	0.0105
	0.5	86.05	1.2667	1.9445	2.1380	0.5542	96.25	1.6134	1.6304	0.0388	0.0096
	1	91.95	1.0113	0.8700	0.8332	0.2944	95.95	1.2414	0.9661	0.0239	0.0077
	2	94.90	0.7507	0.3730	0.3458	0.1820	95.95	0.9126	0.5294	0.0181	0.0079
	4	97.20	0.5020	0.1374	0.2185	0.1897	95.55	0.6067	0.2422	0.0861	0.0554
	10	98.45	0.2315	0.0250	0.0862	0.1751	95.90	0.2806	0.0523	0.0649	0.0900

Table 4.18: *Continued*.

		$q(c_{GS})$									
$AUC = 0.80$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$	$width$	MSE	$Bias$	$SBias$	$cover$	$width$	MSE	$Bias$	$SBias$
		%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$	%	$\times 10^{-1}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{0}$
(30,30)	0.1	80.15	0.3185	0.1641	0.6650	0.6073	93.95	0.4030	0.1080	0.0565	0.0545
	0.25	83.60	0.6863	0.6264	1.0962	0.4871	95.55	0.8830	0.5082	0.0825	0.0366
	0.5	88.05	1.1606	1.4447	1.2407	0.3452	95.40	1.4600	1.3750	0.0235	0.0063
	1	92.05	1.8577	2.9366	1.0470	0.1969	95.15	2.2563	3.2268	0.0053	0.0009
	2	94.90	2.7668	5.2574	0.9325	0.1297	95.45	3.2736	7.1588	0.1545	0.0183
	4	97.00	3.6920	7.1475	1.0472	0.1248	95.40	4.2405	12.0522	0.7000	0.0639
	10	98.45	4.0616	8.6261	0.5384	0.0581	94.00	4.6718	16.1865	1.9320	0.1536
(30,50)	0.1	85.65	0.2912	0.1080	0.4486	0.4785	96.20	0.3672	0.0819	-0.0211	-0.0233
	0.25	88.65	0.6336	0.4307	0.7615	0.3944	96.20	0.7991	0.3802	-0.0436	-0.0224
	0.5	91.35	1.0787	1.0384	0.8462	0.2721	96.55	1.3298	1.0311	-0.0526	-0.0164
	1	93.30	1.7322	2.2176	0.7159	0.1538	96.80	2.0845	2.5204	-0.0764	-0.0152
	2	95.40	2.5778	4.1727	0.9054	0.1415	95.90	3.0721	5.6827	0.0984	0.0131
	4	96.55	3.4270	6.6271	1.6884	0.2120	95.15	4.0264	10.4973	0.5775	0.0564
	10	97.55	3.8582	8.8698	2.6740	0.2960	92.80	4.4635	15.2787	1.9764	0.1619
(50,50)	0.1	79.50	0.2486	0.1012	0.5552	0.6619	96.10	0.3234	0.0647	-0.0174	-0.0217
	0.25	83.75	0.5332	0.3793	0.9148	0.5319	95.95	0.6947	0.3077	-0.0747	-0.0426
	0.5	88.60	0.8997	0.8414	0.9767	0.3575	96.15	1.1428	0.8325	-0.1526	-0.0529
	1	92.85	1.4366	1.6637	0.6930	0.1724	96.35	1.7577	1.9446	-0.2356	-0.0535
	2	95.60	2.1318	2.9536	0.5098	0.0942	96.35	2.5567	4.0930	-0.0533	-0.0083
	4	97.30	2.8392	4.3711	0.6795	0.1033	95.65	3.3533	7.4817	0.3468	0.0401
	10	98.20	3.2391	4.9532	1.0286	0.1477	95.15	3.8120	10.1955	1.3846	0.1384
(100,50)	0.1	67.40	0.2084	0.1126	0.7545	1.0105	95.30	0.2846	0.0518	0.0581	0.0809
	0.25	73.80	0.4373	0.3741	1.2422	0.8377	95.00	0.5966	0.2302	0.1095	0.0724
	0.5	83.20	0.7265	0.7231	1.3762	0.5955	95.10	0.9617	0.5912	0.1071	0.0441
	1	90.45	1.1461	1.2063	1.0401	0.3138	95.10	1.4453	1.3161	0.0242	0.0067
	2	94.90	1.6892	1.8754	0.5432	0.1264	96.05	2.0597	2.7090	0.1283	0.0247
	4	97.75	2.2509	2.4610	0.0519	0.0105	96.65	2.6990	4.8314	0.1192	0.0171
	10	99.10	2.5408	2.6345	-1.1034	-0.2201	96.40	3.0413	6.0814	0.6001	0.0772
(100,100)	0.1	70.65	0.1760	0.0699	0.5766	0.9529	95.65	0.2348	0.0343	0.0065	0.0111
	0.25	77.60	0.3758	0.2421	0.9668	0.7929	95.35	0.4942	0.1523	0.0129	0.0105
	0.5	86.05	0.6333	0.4861	1.0690	0.5542	96.25	0.8067	0.4076	0.0194	0.0096
	1	91.95	1.0113	0.8700	0.8332	0.2944	95.95	1.2414	0.9661	0.0239	0.0077
	2	94.90	1.5014	1.4919	0.6916	0.1820	95.95	1.8251	2.1175	0.0362	0.0079
	4	97.20	2.0079	2.1979	0.8739	0.1897	95.55	2.4269	3.8749	0.3444	0.0554
	10	98.45	2.3146	2.4953	0.8619	0.1751	95.90	2.8065	5.2342	0.6486	0.0900

Table 4.19: Coverage probability ($cover$) and average width ($width$) of 95% CI's for $c_{GS}, p(c_{GS})$ and $q(c_{GS})$, and MSE, $Bias$ and $SBias$ of the point estimators proposed in Section 4.3 under model f) with AUC = 0.90.

		c_{GS}									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	58.05	1.4864	8.9895	7.2072	1.1696	76.30	2.2822	5.3296	0.6552	0.0901
	0.25	71.15	1.3840	4.5435	3.8163	0.6867	89.15	2.4829	4.6893	0.4397	0.0643
	0.5	81.90	1.3400	2.5980	0.3801	0.0748	93.05	2.3938	4.3612	0.0613	0.0093
	1	76.95	1.3600	3.3503	-3.4737	-0.7501	94.70	2.2507	3.9604	-0.4654	-0.0741
	2	70.25	1.5037	5.4779	-6.0382	-1.4104	94.55	2.0392	3.2851	-1.0092	-0.1788
	4	84.80	1.8649	5.2490	-5.8860	-1.3930	92.50	1.8708	2.4638	-1.0872	-0.2244
	10	98.20	3.0107	2.7445	-1.3169	-0.2596	88.40	1.7360	1.9780	-1.0043	-0.2317
(30,50)	0.1	46.75	1.1991	7.2775	7.0808	1.4879	87.00	2.0772	3.5395	0.3912	0.0659
	0.25	67.40	1.1374	3.3398	3.8154	0.8788	93.60	2.1200	3.0945	0.1296	0.0233
	0.5	83.60	1.1331	1.6609	0.4244	0.1047	94.85	2.1126	3.1344	-0.0940	-0.0168
	1	75.10	1.1954	2.656	-3.4492	-0.9006	94.55	2.0730	3.2398	-0.4558	-0.0803
	2	63.05	1.3675	5.1628	-6.1333	-1.6381	93.60	1.9587	2.8069	-0.8880	-0.1700
	4	78.40	1.7155	5.4716	-6.2630	-1.5909	93.00	1.8402	2.3075	-1.0208	-0.2174
	10	95.55	2.6172	3.4822	-2.8986	-0.5638	87.50	1.7241	2.0352	-1.1564	-0.2651
(50,50)	0.1	44.05	1.1385	7.1328	7.0422	1.5101	85.80	2.0518	3.4125	0.5185	0.0891
	0.25	66.55	1.0625	3.1308	3.7018	0.8820	93.10	2.0184	2.8406	0.3966	0.0746
	0.5	81.45	1.0295	1.4764	0.3092	0.0807	94.65	1.9417	2.5960	0.1348	0.0265
	1	68.80	1.0458	2.4325	-3.4966	-1.0050	95.35	1.8137	2.2576	-0.2246	-0.0473
	2	48.35	1.1515	4.6390	-6.0110	-1.8763	95.30	1.6171	1.7823	-0.5174	-0.1235
	4	72.10	1.4115	4.3711	-5.8154	-1.8485	94.25	1.4287	1.3566	-0.5605	-0.1539
	10	97.90	2.1629	1.7221	-1.5621	-0.4062	93.40	1.3116	1.0869	-0.6583	-0.2037
(100,50)	0.1	43.15	1.0975	6.8352	6.8612	1.4871	84.45	2.0111	3.2216	0.4480	0.0792
	0.25	65.95	1.0075	2.9111	3.4919	0.8488	92.40	1.9196	2.5669	0.4067	0.0805
	0.5	81.20	0.9489	1.3631	0.1200	0.0325	94.90	1.7845	2.1223	0.1193	0.0259
	1	61.65	0.9168	2.3512	-3.6088	-1.1140	94.95	1.5836	1.7181	-0.1536	-0.0371
	2	31.05	0.9476	4.3345	-5.9591	-2.1285	95.05	1.3054	1.1523	-0.3824	-0.1133
	4	56.50	1.0988	3.6026	-5.4585	-2.1862	94.85	1.0664	0.7159	-0.3097	-0.1165
	10	99.40	1.6377	0.8106	-0.3420	-0.1210	95.55	0.9189	0.5143	-0.2915	-0.1296
(100,100)	0.1	20.10	0.7969	5.8303	6.9540	2.2045	93.60	1.5689	1.6041	0.3263	0.0817
	0.25	52.70	0.7444	2.1246	3.6351	1.2823	95.15	1.4639	1.3465	0.2438	0.0666
	0.5	83.60	0.7223	0.6757	0.2597	0.1004	95.70	1.4164	1.2414	0.1522	0.0432
	1	51.95	0.7341	1.7997	-3.5291	-1.4987	96.15	1.3287	1.1002	-0.0186	-0.0056
	2	16.15	0.8067	4.1126	-6.0299	-2.7614	96.20	1.1542	0.8438	-0.2067	-0.0713
	4	38.70	0.9800	3.8736	-5.8326	-2.6850	95.95	0.9905	0.5904	-0.1904	-0.0786
	10	98.00	1.4610	0.9741	-1.5722	-0.5830	95.10	0.8922	0.4852	-0.2381	-0.1087

Table 4.19: Continued.

		$p(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	65.70	2.4941	15.8849	9.8024	1.2370	93.35	3.4413	7.1389	0.0038	0.0005
	0.25	70.45	2.1701	10.2628	7.8812	1.2379	95.55	3.0413	5.7188	0.3509	0.0464
	0.5	79.25	1.9090	6.1012	5.8234	1.1184	96.05	2.5490	4.0887	0.4652	0.0729
	1	90.95	1.6350	2.6782	3.1582	0.7702	95.65	1.9694	2.5528	0.3304	0.0655
	2	97.35	1.3434	1.0020	0.7258	0.2355	95.15	1.4253	1.3429	0.0718	0.0196
	4	97.35	1.0236	0.5090	-0.5845	-0.2682	95.50	0.9783	0.6574	-0.0054	-0.0021
	10	96.10	0.5429	0.2341	-0.9868	-0.8436	92.85	0.5237	0.2063	0.0906	0.0632
(30,50)	0.1	66.55	2.3689	13.7289	8.9017	1.1681	95.55	3.2631	6.7404	-0.2340	-0.0285
	0.25	71.60	2.0676	9.2641	7.4151	1.2081	95.65	2.8288	5.0944	0.1081	0.0151
	0.5	78.45	1.8198	5.7186	5.6447	1.1214	94.95	2.3691	3.5501	0.1224	0.0205
	1	88.85	1.5549	2.6187	3.2164	0.8079	95.90	1.8259	2.1085	0.0823	0.0179
	2	95.90	1.2700	0.9982	0.9561	0.3174	95.50	1.3187	1.0974	-0.0588	-0.0178
	4	97.45	0.9597	0.4681	-0.2603	-0.1211	94.75	0.9141	0.5459	-0.0088	-0.0038
	10	95.60	0.5170	0.1867	-0.6006	-0.4892	90.55	0.4998	0.1916	0.1425	0.1035
(50,50)	0.1	53.15	1.9224	13.1762	9.7208	1.5919	94.50	2.7540	4.6338	-0.0417	-0.0061
	0.25	59.65	1.6745	8.4912	7.8100	1.5966	95.20	2.3773	3.5080	0.2059	0.0348
	0.5	71.20	1.4745	4.9081	5.7588	1.4431	95.05	1.9928	2.4764	0.2223	0.0447
	1	87.25	1.2624	1.9436	3.1020	0.9900	95.40	1.5421	1.4779	0.1278	0.0332
	2	96.25	1.0366	0.5969	0.6810	0.2902	95.60	1.1105	0.7396	-0.0388	-0.0143
	4	97.10	0.7914	0.3112	-0.6156	-0.3723	95.55	0.7728	0.3819	-0.0344	-0.0176
	10	92.75	0.4372	0.1705	-0.9461	-1.0511	94.80	0.4332	0.1310	0.0713	0.0624
(100,50)	0.1	36.85	1.5056	12.7793	10.1224	2.0107	94.20	2.2928	3.3812	-0.1158	-0.0199
	0.25	41.95	1.2939	7.8906	7.9331	1.9845	95.40	1.9562	2.3665	0.1761	0.0362
	0.5	55.75	1.1292	4.2902	5.6996	1.7655	95.75	1.6472	1.6262	0.1851	0.0459
	1	79.70	0.9600	1.4622	2.8969	1.1603	95.95	1.2815	0.9968	0.0240	0.0076
	2	96.50	0.7849	0.3483	0.3782	0.2069	96.35	0.9090	0.5115	-0.1065	-0.0471
	4	93.75	0.6001	0.2472	-0.9566	-0.7666	96.20	0.6151	0.2446	-0.0757	-0.0484
	10	73.60	0.3359	0.1990	-1.2556	-1.9520	95.90	0.3437	0.0762	-0.0012	-0.0014
(100,100)	0.1	29.40	1.3541	11.4811	9.7585	2.2047	95.50	1.9780	2.5226	0.0138	0.0027
	0.25	33.80	1.1793	7.4114	7.8461	2.2141	95.40	1.6812	1.8314	0.0636	0.0149
	0.5	47.30	1.0382	4.1878	5.7909	2.0042	95.00	1.4132	1.2862	0.0979	0.0273
	1	76.25	0.8888	1.4928	3.1296	1.3809	95.70	1.0905	0.7385	0.0426	0.0157
	2	95.20	0.729	0.3363	0.7052	0.4164	96.05	0.7816	0.3786	0.0025	0.0013
	4	96.25	0.5575	0.1763	-0.5945	-0.5007	96.60	0.5513	0.1926	-0.0210	-0.0151
	10	83.70	0.3167	0.1320	-0.9557	-1.4991	96.05	0.3221	0.0667	0.0150	0.0184

Table 4.19: Continued.

		$q(c_{GS})$									
$AUC = 0.90$		GPQ 95% CI		Parametric estimator			EL 95% CI		Nonparametric estimator		
(n_0, n_1)	rho	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$	$cover$ %	$width$ $\times 10^{-1}$	MSE $\times 10^{-3}$	$Bias$ $\times 10^{-2}$	$SBias$ $\times 10^{0}$
(30,30)	0.1	65.70	0.2494	0.1588	0.9802	1.2370	93.35	0.3441	0.0714	0.0004	0.0005
	0.25	70.45	0.5425	0.6414	1.9703	1.2379	95.55	0.7603	0.3574	0.0877	0.0464
	0.5	79.25	0.9545	1.5253	2.9117	1.1184	96.05	1.2745	1.0222	0.2326	0.0729
	1	90.95	1.6350	2.6782	3.1582	0.7702	95.65	1.9694	2.5528	0.3304	0.0655
	2	97.35	2.6868	4.0078	1.4516	0.2355	95.15	2.8505	5.3714	0.1437	0.0196
	4	97.35	4.0946	8.1439	-2.3379	-0.2682	95.50	3.9134	10.5186	-0.0215	-0.0021
	10	96.10	5.4291	23.4142	-9.8684	-0.8436	92.85	5.2365	20.6272	0.9064	0.0632
(30,50)	0.1	66.55	0.2369	0.1373	0.8902	1.1681	95.55	0.3263	0.0674	-0.0234	-0.0285
	0.25	71.60	0.5169	0.5790	1.8538	1.2081	95.65	0.7072	0.3184	0.0270	0.0151
	0.5	78.45	0.9099	1.4297	2.8224	1.1214	94.95	1.1846	0.8875	0.0612	0.0205
	1	88.85	1.5549	2.6187	3.2164	0.8079	95.90	1.8259	2.1085	0.0823	0.0179
	2	95.90	2.5400	3.9926	1.9123	0.3174	95.50	2.6374	4.3897	-0.1177	-0.0178
	4	97.45	3.8387	7.4900	-1.0411	-0.1211	94.75	3.6565	8.7341	-0.0351	-0.0038
	10	95.60	5.1695	18.6718	-6.0060	-0.4892	90.55	4.9983	19.1556	1.4252	0.1035
(50,50)	0.1	53.15	0.1922	0.1318	0.9721	1.5919	94.50	0.2754	0.0463	-0.0042	-0.0061
	0.25	59.65	0.4186	0.5307	1.9525	1.5966	95.20	0.5943	0.2193	0.0515	0.0348
	0.5	71.20	0.7372	1.2270	2.8794	1.4431	95.05	0.9964	0.6191	0.1111	0.0447
	1	87.25	1.2624	1.9436	3.1020	0.9900	95.40	1.5421	1.4779	0.1278	0.0332
	2	96.25	2.0732	2.3875	1.3621	0.2902	95.60	2.2211	2.9584	-0.0776	-0.0143
	4	97.10	3.1654	4.9792	-2.4626	-0.3723	95.55	3.0913	6.1099	-0.1375	-0.0176
	10	92.75	4.3723	17.0505	-9.4614	-1.0511	94.80	4.3318	13.1036	0.7134	0.0624
(100,50)	0.1	36.85	0.1506	0.1278	1.0122	2.0107	94.20	0.2293	0.0338	-0.0116	-0.0199
	0.25	41.95	0.3235	0.4932	1.9833	1.9845	95.40	0.4890	0.1479	0.0440	0.0362
	0.5	55.75	0.5646	1.0726	2.8498	1.7655	95.75	0.8236	0.4066	0.0925	0.0459
	1	79.70	0.9600	1.4622	2.8969	1.1603	95.95	1.2815	0.9968	0.0240	0.0076
	2	96.50	1.5697	1.3931	0.7564	0.2069	96.35	1.8180	2.0459	-0.2131	-0.0471
	4	93.75	2.4003	3.9546	-3.8266	-0.7666	96.20	2.4604	3.9129	-0.3027	-0.0484
	10	73.60	3.3595	19.9019	-12.5563	-1.9520	95.90	3.4373	7.6178	-0.0120	-0.0014
(100,100)	0.1	29.40	0.1354	0.1148	0.9759	2.2047	95.50	0.1978	0.0252	0.0014	0.0027
	0.25	33.80	0.2948	0.4632	1.9615	2.2141	95.40	0.4203	0.1145	0.0159	0.0149
	0.5	47.30	0.5191	1.0470	2.8954	2.0042	95.00	0.7066	0.3215	0.0489	0.0273
	1	76.25	0.8888	1.4928	3.1296	1.3809	95.70	1.0905	0.7385	0.0426	0.0157
	2	95.20	1.4580	1.3453	1.4103	0.4164	96.05	1.5632	1.5145	0.0051	0.0013
	4	96.25	2.2299	2.8201	-2.3781	-0.5007	96.60	2.2054	3.0817	-0.0841	-0.0151
	10	83.70	3.1670	13.1965	-9.5573	-1.4991	96.05	3.2208	6.6734	0.1504	0.0184

From Tables 4.14–4.16, we observe that the EL method has a good behaviour in terms of both interval coverage and width when estimating c_{GS}, $p(c_{GS})$ and $q(c_{GS})$. However, the GPQ method presents in some cases overcoverage or undercoverage when estimating c_{GS}, and for $\rho = 0.5, 1, 2, 4$ it presents most of the times a very pronounced undercoverage when estimating $p(c_{GS})$ and $q(c_{GS})$. This undercoverage gets worse as the sample sizes increase. From Tables 4.17–4.19, we observe that the EL method has a good behaviour in terms of both interval coverage and width when estimating c_{GS}, $p(c_{GS})$ and $q(c_{GS})$. However, the GPQ method shows most of the times a pronounced undercoverage when estimating c_{GS} for $\rho = 0.1, 0.25, 0.5, 1, 2, 4$, and sometimes it shows undercoverage when estimating $p(c_{GS})$ and $q(c_{GS})$ that gets worse as the AUC and the sample sizes increase, specially for small values of ρ. However, for large values of ρ, the GPQ method presents sometimes overcoverage when estimating $p(c_{GS})$ and $q(c_{GS})$. Besides, from Tables 4.14–4.19, we observe that in terms of $Bias$, the nonparametric point estimators proposed for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ behave much better than their parametric counterparts.

Therefore, although the parametric point estimators and the GPQ-based CIs turned out to perform well under the Box-Cox family models and the bigamma model, this was not the case for the mixtures scenarios where the distribution models differ between healthy and diseased populations (see plots e)–f) in Figure 4.7). So, we strongly recommend the use of the nonparametric point estimators and the EL-based CIs when the data do not follow the Box-Cox family.

4.5 Biomedical applications

In this section, we consider three real biomedical datasets on melanoma, prostate cancer, and coronary artery disease to illustrate the methodology previously discussed. See Section 1.3 of Chapter 1 for more details on these datasets and their basic descriptive statistics (mean, median, variance, standard deviation, minimum and maximum). In the analysis shown below, the Shapiro-Wilk normality test has been used to assess if the markers or a Box-Cox transformation of them can be assumed normally distributed at a 5% significance level.

4.5.1 Melanoma dataset

The empirical ROC curve corresponding to these data is displayed in the right panel of Figure 4.8. The AUC estimated for this dataset is equal to 0.906 (95% CI 0.833-0.979), which indicates a very good discrimination accuracy of CSS for differenciating between melanoma and non-melanoma pigmented lesions.

In the left panel of Figure 4.8 we show smooth density estimates of the clinical scoring scheme without dermoscope (CSS) in the two groups of interest (non-melanoma and melanoma groups). It seems that the non-melanoma group shows a slight bimodality while the melanoma group shows a more symmetric distribution. However, according to the Shapiro-Wilk normality test, both groups can be considered normally distributed (p-value $= 0.4719$ for the non-melanoma group and p-value $= 0.9084$ for the melanoma group). Therefore, since we can assume normality for the CSS in both groups and do not

need to previously transform the data by a monotone Box-Cox transformation, we show below (see Table 4.20) the parametric point estimates and 95%-confidence intervals for the Generalized Symmetry point of the CSS and its associated specificity and sensitivity indexes obtained with the GPQ approach. In principle, a FN decision will be more expensive than a FP one, because in a skin lesion leads to excision and then you will do the biopsy. Therefore, we will consider the following different values for the relative loss of a false positive misclassification compared with a false negative one: $\rho = 0.5, 0.25, 0.10$, which means that a false negative result is 2, 4 or even 10 times more serious than a false positive result, respectively. However, there may be some cases where it would be the contrary. For example, when it happens that there is a successful treatment but very aggressive or that can lead to an erroneous removal of something important. In this case, for example, the values of $\rho = 2, 4, 10$ (that is, a FP misclassification would be 2,4 or even 10 times more serious than a FN misclassification) would be more adequate. For the sake of illustration, we also consider here the value $\rho = 1$, that is, corresponding to the Symmetry point, where we assume the same cost for both FP and FN misclassifications.

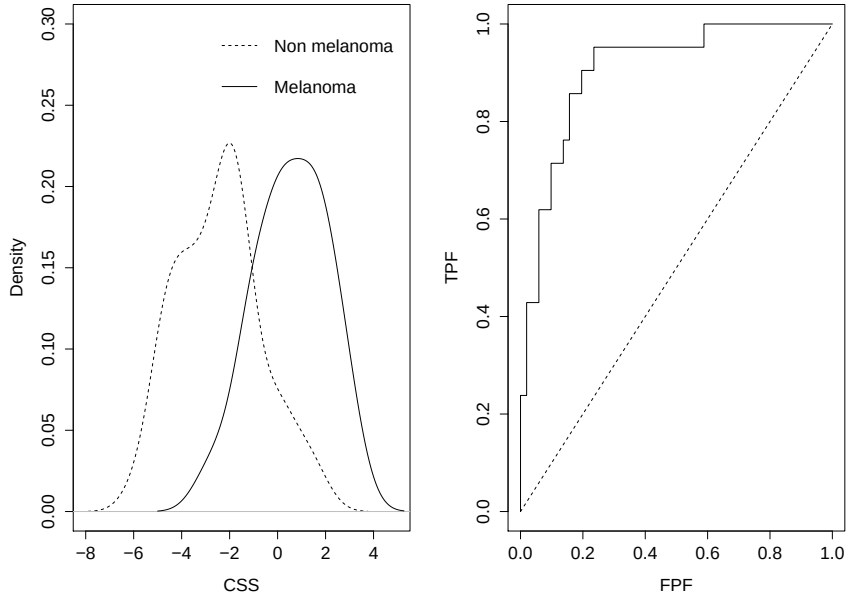


Figure 4.8: *Left panel*: Density estimates of the clinical scoring scheme without dermoscope (CSS) in melanoma (*solid line*) and non-melanoma (*dashed line*) groups. *Right panel*: ROC curve of CSS marker.

Firstly, if we do not take into account costs for the different misclassifications and therefore we assume the same cost for a false positive decision and for a false negative one, in this case, a Symmetry point of -0.827 is obtained, so that a pigmented lesion on the skin with a CSS value bigger than or equal to -0.827 is classified as a melanoma. With this optimal value, 81.8% of the patients with melanoma and 81.8% of the patients without melanoma are correctly classified. The sampling variability of the Symmetry point

is accounted by the corresponding 95% confidence interval given by $(-1.325; -0.298)$. If we consider, however, that we would assume the cost to perform 10 biopsies to patients without skin melanoma for avoiding to miss a patient with skin melanoma (that is, assuming a relative cost of $\rho = 0.1$), the Generalized Symmetry point obtained is -2.075 with a sensitivity of 0.958 and a specificity of 0.579. Finally, if we consider a ρ value of 2, the optimal value will be -0.434, with a sensitivity of 0.742 and a specificity of 0.871.

Table 4.20: Melanoma dataset: Parametric point estimates and GPQ-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of CSS marker.

Methods		GPQ	
Parameter	ρ	Parametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	0.496	(-0.002,1.095)
	4	-0.035	(-0.496,0.490)
	2	-0.434	(-0.905,0.075)
	1	-0.827	(-1.325,-0.298)
	0.5	-1.213	(-1.793,-0.628)
	0.25	-1.591	(-2.286,-0.941)
	0.1	-2.075	(-2.966,-1.322)
$p(c_{GS})$	10	0.952	(0.932,0.967)
	4	0.913	(0.872,0.943)
	2	0.871	(0.809,0.918)
	1	0.818	(0.727,0.887)
	0.5	0.755	(0.625,0.849)
	0.25	0.683	(0.515,0.805)
	0.1	0.579	(0.358,0.742)
$q(c_{GS})$	10	0.516	(0.317,0.672)
	4	0.651	(0.489,0.773)
	2	0.742	(0.617,0.836)
	1	0.818	(0.727,0.887)
	0.5	0.877	(0.812,0.924)
	0.25	0.921	(0.879,0.951)
	0.1	0.958	(0.936,0.974)

Since the EL method is a nonparametric method that does not require any parametric assumptions which has been proved appropriate under all the scenarios considered in the simulation study, we show below in Table 4.21 the nonparametric point estimates and the EL-based 95% CIs for the Generalized Symmetry point of the CSS and its associated specificity and sensitivity indexes.

As expected, the confidence intervals obtained with the EL approach turn out to be wider than those obtained with the GPQ approach. Additionally, it is worthwhile to mention here that the parametric (nonparametric) point estimates are included in the corresponding EL-based (GPQ-based) confidence intervals.

Therefore, taking into account the results obtained, we observe that the value of the

optimal cutpoint and its corresponding sensitivity and specificity indexes can be very different depending on the value of the relative cost assumed for the two types of misclassifications. Therefore, it is very important to know beforehand the specific clinical setting, the consequences of the different incorrect decisions and to incorporate the adequate costs for the different diagnostic decisions when computing the optimal cutpoint if we do not want to draw erroneous conclusions in clinical practice.

Table 4.21: Melanoma dataset: Nonparametric point estimates and EL-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of CSS marker.

Methods		EL	
Parameter	ρ	Nonparametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	0.794	(-0.108,1.495)
	4	0.122	(-0.871,0.933)
	2	-0.345	(-1.207,0.331)
	1	-1.025	(-1.520,-0.055)
	0.5	-1.238	(-1.840,-0.457)
	0.25	-1.616	(-2.567,-0.927)
	0.1	-2.226	(-2.943,-1.096)
$p(c_{GS})$	10	0.945	(0.927,0.971)
	4	0.905	(0.861,0.949)
	2	0.861	(0.799,0.931)
	1	0.833	(0.729,0.906)
	0.5	0.790	(0.633,0.897)
	0.25	0.725	(0.446,0.873)
	0.1	0.533	(0.342,0.848)
$q(c_{GS})$	10	0.451	(0.270,0.706)
	4	0.621	(0.445,0.797)
	2	0.723	(0.598,0.862)
	1	0.833	(0.729,0.906)
	0.5	0.895	(0.816,0.949)
	0.25	0.931	(0.861,0.968)
	0.1	0.953	(0.934,0.985)

4.5.2 Prostate cancer dataset

The empirical ROC curve corresponding to these data is shown in the right panel of Figure 4.9. From the analysis carried out by Le (2006), he found an estimated AUC value equal to 0.726, indicating a moderate-good classification performance, and, for practical use, obtained an optimal cutpoint of 0.75 based on the well-known optimality criterion given by the Youden index, with an associated specificity of 0.791 and an associated sensitivity of 0.554, which indicates that APBS has some good screening value but may be not good enough to be used alone for the prediction of nodal involvement.

In the left panel of Figure 4.9 we can observe that both distributions are non symmetric, in the way that the healthy group shows a slight bimodality and the diseased group presents a multimodality distribution. Indeed, according to the Shapiro-Wilk normal-

ity test, both groups can not be considered normally distributed (p-value < 0.0001 for healthy group and p-value $= 0.0232$ for diseased group). However, after transforming the data using the Box-Cox tranformation estimate given by $\hat{\lambda} = -1.2494$, the Shapiro-Wilk normality test indicates that both groups can be considered normally distributed (p-value $= 0.3641$ for group 0 and p-value $= 0.2118$ for group 1). Therefore, we show below the parametric point estimates and 95%-confidence intervals for the Generalized Symmetry point of APBS (considering firstly $\rho = 1$, that is, the Symmetry point) and its associated specificity and sensitivity indexes obtained with the GPQ approach.

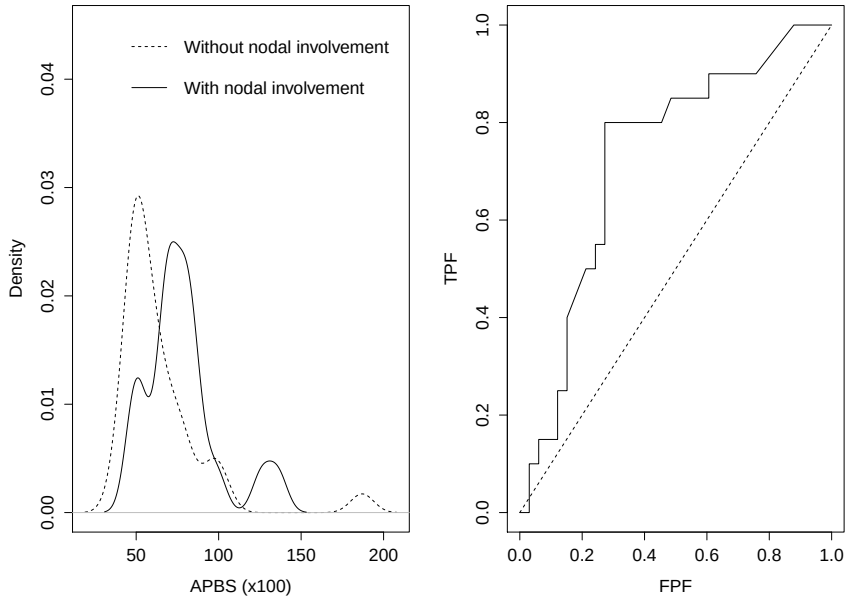


Figure 4.9: *Left panel*: Density estimates of the level of acid phosphatase in blood serum (APBS) $\times 100$ for individuals with (*solid line*) and without nodal involvement (*dashed line*). *Right panel*: ROC curve of APBS$\times 100$ marker.

Based on the GPQ method, the results obtained for the Symmetry point are:

$$\tilde{c}_S = 64.664, \quad 95\%\ \text{CI}(c_S) = (60.105, 70.261),$$

$$\tilde{p}(c_S) = \tilde{q}(c_S) = 0.659, \quad 95\%\ \text{CI}(p(c_S)) = 95\%\ \text{CI}(q(c_S)) = (0.549, 0.760).$$

Then, an individual suffering from prostate cancer would be classified as a patient with nodal involvement if his/her APBS value is bigger than or equal to the Symmetry point given by 64.664. The corresponding 95% confidence interval $(60.105, 70.261)$ measures the uncertainty associated with the point estimate of the Symmetry point. With this classification rule, it turns out that 65.9% of the patients with nodal involvement and 65.9% of the patients without nodal involvement are correctly classified.

Table 4.22: Prostate cancer dataset: Parametric point estimates and GPQ-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of APBS marker.

Methods		GPQ	
Parameter	ρ	Parametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	92.554	(80.006,118.783)
	4	78.527	(71.024,90.155)
	2	70.737	(65.271,78.041)
	1	64.664	(60.105,70.261)
	0.5	59.862	(55.438,64.995)
	0.25	56.001	(51.197,61.004)
	0.1	51.952	(46.635,57.127)
$p(c_{GS})$	10	0.920	(0.905,0.939)
	4	0.842	(0.798,0.885)
	2	0.759	(0.683,0.829)
	1	0.659	(0.549,0.760)
	0.5	0.552	(0.408,0.684)
	0.25	0.447	(0.277,0.609)
	0.1	0.323	(0.141,0.514)
$q(c_{GS})$	10	0.202	(0.052,0.395)
	4	0.370	(0.192,0.541)
	2	0.518	(0.367,0.657)
	1	0.659	(0.549,0.760)
	0.5	0.776	(0.704,0.842)
	0.25	0.862	(0.819,0.902)
	0.1	0.932	(0.914,0.951)

However, to find a cutpoint based on criteria that do not consider misclassification costs such as the Youden index or the Symmetry point may not be satisfactory in practice. In fact, in this biomedical example, a false negative result is much more harmful than a false positive result and therefore the incorporation of different costs associated to both types of decisions is essential in this clinical setting. For this reason, in the following, we analyze this dataset using our new approach for estimating the Generalized Symmetry point. In Table 4.22, we collect the point estimates and confidence intervals of the Generalized Symmetry point and associated specificity and sensitivity indexes for different values of ρ. Since a cutpoint with high sensitivity would be a good choice for the clinical use of APBS, we consider the following different values for the relative loss of a false-positive classification as compared with a false-negative classification: $\rho = 0.5, 0.25, 0.10$. These values mean that a false negative result is 2, 4 or even 10 times more serious than a false positive result, respectively. In addition, for the sake of illustration and comparison of the results obtained, the values $\rho = 1$ (the same cost for both missclassifications, that is, the Symmetry point already computed above), $\rho = 2$ (the cost associated to a false positive is double than the one associated to a false negative), $\rho = 4$ (the cost associated to a false positive is 4 times more than the one associated to a false negative), and even $\rho = 10$ (the cost associated to a false positive is 10 times more than the one associated to

a false negative) were also considered. Due to the fact that these data follow the Box-Cox family, we show firstly the 95%-confidence intervals obtained with the GPQ approach. For instance, for a ρ value of 0.1, (that is, the clinician would be willing to perform 10 laparoscopics to patients without nodal involvement in order to avoid missing a patient with nodal involvement), the point estimate of the Generalized Symmetry point is 51.952 with an associated sensitivity of 0.932 and an associated specificity of 0.323. The sampling variability of this optimal cutpoint is accounted for by the 95% confidence interval given by $(46.635, 57.127)$. Note that for $\rho = 2$, the optimal threshold value given by the Generalized Symmetry point is 70.737 with an associated sensitivity of 0.518 and an associated specificity of 0.759, and is the most similar value to that obtained in Le (2006) based on Youden index. However, this case yields a cutpoint with specificity greater than sensitivity, and this is exactly the opposite to what is aimed here. Therefore, we underline the importance of considering the misclassification costs in the selection of the optimal threshold value to avoid erroneous conclusions when using APBS in clinical practice.

Analogously to the previous analysis done on the melanoma dataset, we show below in Table 4.23 the nonparametric point estimates and the EL-based 95% CIs for the Generalized Symmetry point of APBS and its associated specificity and sensitivity accuracy indexes.

Table 4.23: Prostate cancer dataset: Nonparametric point estimates and EL-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of APBS marker.

Methods		EL	
Parameter	ρ	Nonparametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	97.152	(76.861,136.724)
	4	80.292	(70.782,95.212)
	2	72.598	(65.170,80.705)
	1	67.120	(59.136, 73.649)
	0.5	61.544	(52.659,68.235)
	0.25	52.167	(48.730,64.791)
	0.1	49.225	(45.406,59.975)
$p(c_{GS})$	10	0.914	(0.901,0.945)
	4	0.844	(0.783,0.899)
	2	0.761	(0.679,0.860)
	1	0.717	(0.580,0.803)
	0.5	0.600	(0.385,0.748)
	0.25	0.421	(0.219,0.689)
	0.1	0.245	(0.093,0.538)
$q(c_{GS})$	10	0.141	(0.013,0.446)
	4	0.378	(0.134,0.596)
	2	0.521	(0.358,0.719)
	1	0.717	(0.580, 0.803)
	0.5	0.800	(0.692,0.874)
	0.25	0.855	(0.805,0.922)
	0.1	0.925	(0.909,0.954)

4.5.3 Coronary artery disease dataset

The ROC curve corresponding to this dataset is shown in the right panel of Figure 4.10 and the estimated AUC is equal to 0.744 (95% CI 0.659, 0.828), which means that the elastase concentration has a moderate-good discrimination accuracy for discriminating between coronary and non-stenotic coronaries groups. Amaro et al. (1995) selected the optimal cutpoint as the value that maximized the correct classification rate (77.4%) which was achieved with a threshold of $20\ \mu gl^{-1}$ with a sensitivity of 95.8% and a specificity of 37.8%. So, blood leukocyte elastase concentration seems to be a sensitive diagnostic marker of CAD. In fact, if we used the elastase determination as a screening test previous to performing the coronary angiography for detecting CAD, we would search for a high sensitivity in order to be able to identify all the diseased patients.

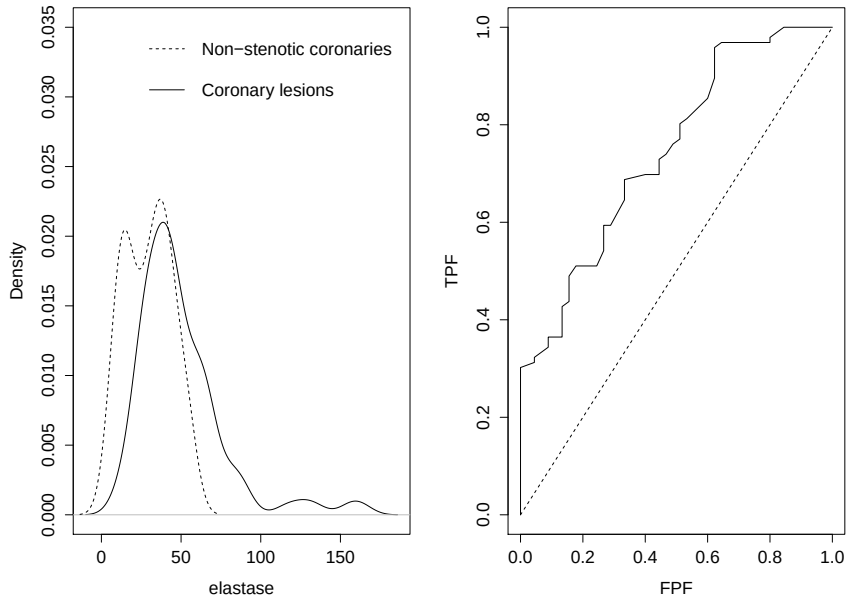


Figure 4.10: *Left panel*: Density estimates of peripheral blood leukocyte elastase concentration in coronary lesions group (*solid line*) and in non-stenotic coronaries group (*dashed line*). *Right panel*: ROC curve of elastase concentration marker.

In the following, we analyze this dataset using our approach. In Table 4.24, we collect the point estimates and 95%-confidence intervals of the Generalized Symmetry point of the elastase concentration and its associated specificity and sensitivity indexes for different values of ρ. According to the Shapiro-Wilk normality test, only the healthy group can be considered normally distributed (p-value $= 0.0746$ for group 0 and p-value < 0.0001 for group 1, that is, for patients with coronary lesions). After transforming the data using the Box-Cox transformation estimate given by $\hat{\lambda} = 0.1136$, the Shapiro-Wilk normality test indicates that only the diseased group can be considered normally distributed (p-value $= 0.0091$ for group 0 and p-value $= 0.0793$ for group 1). Therefore, we present

below the results obtained with the EL approach (that is, the nonparametric point estimates and the EL-based 95% CIs for the Generalized Symmetry point of the elastase concentration and its associated specificity and sensitivity indexes), which does not require the assumption of normality, because these data do not follow the Box-Cox family. Given that searching for a high sensitivity is equivalent to searching for a low number of false negatives, the same values as in the previous prostate cancer dataset were considered for ρ: $\rho = 0.5, 0.25, 0.10$ (a false negative result is regarded as more serious than a false positive one). For the sake of illustration, the values $\rho = 1, 2, 4, 10$ were also considered. Assuming, for example, $\rho = 0.1$ (that is, the clinician would be willing to perform 10 coronary angiographies to patients without CAD in order to avoid missing a patient with CAD), the point estimate of the Generalized Symmetry point is 23.136 with an associated sensitivity of 0.937 and an associated specificity of 0.374. Note that the optimal threshold value obtained in Amaro et al. (1995), despite having been selected by a different optimality criterion, is very close to the cutpoint obtained here considering $\rho = 0.1$.

Table 4.24: Coronary artery disease dataset: Nonparametric point estimates and EL-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of elastase marker.

Methods		EL	
Parameter	ρ	Nonparametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	51.960	(45.503,55.688)
	4	46.158	(41.386,51.235)
	2	41.383	(38.053,45.032)
	1	37.733	(33.692,40.291)
	0.5	31.597	(27.494,35.532)
	0.25	26.558	(23.991,29.672)
	0.1	23.136	(17.366,25.552)
$p(c_{GS})$	10	0.933	(0.924,0.948)
	4	0.858	(0.829,0.893)
	2	0.763	(0.718,0.823)
	1	0.666	(0.573, 0.733)
	0.5	0.515	(0.415,0.636)
	0.25	0.416	(0.302,0.569)
	0.1	0.374	(0.218,0.511)
$q(c_{GS})$	10	0.334	(0.243,0.481)
	4	0.433	(0.316,0.573)
	2	0.525	(0.435,0.646)
	1	0.666	(0.573, 0.733)
	0.5	0.757	(0.707,0.818)
	0.25	0.854	(0.825,0.892)
	0.1	0.937	(0.922,0.951)

If we erroneously assumed that a false negative misclassification has the same cost as a false positive one, we would obtain an optimal elastase value of $\hat{c}_S = 37.733$ and therefore in this case those patients with an elastase value bigger than or equal to the Symme-

try point given by 37.733 are identified as having CAD. The sampling variability of this Symmetry point is represented by the corresponding $95\%CI(c_S) = (33.692, 40.291)$. It should be noted that this optimal cutpoint is quite different to the previous one, with both sensitivity and specificity equal to $\hat{p}(c_S) = \hat{q}(c_S) = 0.666$ ($CI_{95\%}(p(c)) = CI_{95\%}(q(c)) = (0.573, 0.733)$). So, 66.6% of the patients with CAD and 66.6% of the individuals without CAD are correctly identified.

For the sake of illustration, we show below (see Table 4.25) the results obtained with the GPQ approach. However, since the data do not satisfy the assumption of normality (even after applying a Box-Cox transformation estimated by maximum likelihood), the following results have to be considered with caution because there is no guarantee that the coverage of these GPQ-based CIs is close to the nominal value of 95% (as we have seen in the simulation results).

Table 4.25: Coronary artery disease dataset: Parametric point estimates and GPQ-based 95% CI's for c_{GS}, $p(c_{GS})$ and $q(c_{GS})$ of elastase marker.

Methods		GPQ	
Parameter	ρ	Parametric point estimate	95% $CI(c_{GS})$
c_{GS}	10	59.394	(50.451,72.798)
	4	48.135	(42.236,55.896)
	2	40.894	(36.567,46.072)
	1	34.716	(31.346,38.424)
	0.5	29.513	(26.671,32.530)
	0.25	26.558	(23.991,29.672)
	0.1	20.521	(17.935,23.339)
$p(c_{GS})$	10	0.928	(0.914,0.941)
	4	0.860	(0.828,0.888)
	2	0.785	(0.736,0.831)
	1	0.693	(0.625, 0.759)
	0.5	0.590	(0.502,0.675)
	0.25	0.416	(0.302,0.569)
	0.1	0.354	(0.236,0.473)
$q(c_{GS})$	10	0.277	(0.139,0.411)
	4	0.438	(0.312,0.553)
	2	0.570	(0.471,0.662)
	1	0.693	(0.625, 0.759)
	0.5	0.795	(0.751,0.838)
	0.25	0.854	(0.825,0.892)
	0.1	0.935	(0.924,0.947)

Free software is software that respects your freedom and the social solidarity of your community. So it's free as in freedom.

- Richard Stallman

Chapter 5

Software development: R packages

In Chapters 3 and 4 we have described several criteria existing in the literature for selecting optimal cutpoints in continuous diagnostic tests. To facilitate this task of selecting optimal values in clinical practice, it is essential to have software that implements those optimal criteria in an environment that biomedical researchers find user-friendly and easily understandable. Important contributions to this issue have been made by the following R (R Core Team, 2013) packages: *DiagnosisMed* (Brasil, 2010), *pROC* (Robin et al., 2011) and *Epi* (Carstensen et al., 2013). The *DiagnosisMed* package includes the implementation of 10 different criteria for estimating optimal cutpoints such as, for instance, the criterion that selects the cutoff at which sensitivity and specificity are equal, that is, the Symmetry point (Amaro et al., 1995; Greiner, 1995; Hosmer and Lemeshow, 2000), or that based on maximizing the diagnostic odds ratio (Kraemer, 1992; Böhning et al., 2011). Moreover, in this package there is a function that draws the ROC curve and computes the test accuracy measures for each threshold. Even though their main objective is not the selection of an optimal cutpoint, the *pROC* and *Epi* packages also include some specific functions for selecting the optimal value using only one or two criteria. Specifically, the *pROC* package provides the cut-off point associated to the North-West corner criterion, that is, the optimal point on the ROC curve (Metz, 1978; Swets and Swets, 1979; Swets and Pickett, 1982) that is closest to the point (0,1) (Metz, 1978; Vermont et al., 1991), and the cut-off point associated to the Youden index (Youden, 1950), allowing the costs of the different diagnostic decisions to be incorporated into both criteria. The *Epi* package computes the threshold value that maximizes the sum of sensitivity plus specificity indexes, which is equivalent to the cut-off point given by the Youden index. There are other R packages that provide methods for obtaining classification rules for specific models or contexts other than medical settings. For instance, the *PresenceAbsence* package which includes 12 different criteria, and has two graphical functions to construct the ROC curve and plot accuracy measures, e.g., the sensitivity or specificity as a function of the threshold (Freeman, 2007; Freeman and Moisen, 2008); and the *SDMTools* package with 8 criteria implemented (VanDerWal et al., 2012).

However, all these packages have some limitations. Firstly, as pointed out above, some of them include very few selection criteria. Secondly, none of the packages include criteria based on predictive values and/or likelihoods ratios, i.e., they only consider optimal-cutpoint selection criteria based on sensitivity and specificity measures. This

entails an important limitation from the standpoint of clinical applicability: although sensitivity and specificity are considered the fundamental operating characteristics of a diagnostic test, there are nevertheless times when such measures may not be so useful in clinical practice, since clinical staff do not have prior information about the patient's true disease status. Indeed, the problem tends to be just the opposite, and involves the need to ascertain the probability of the patient being healthy (or diseased) in a case where the test result is negative (or positive). Hence, strategies for selecting the optimal cutpoint based on predictive values can be more useful in certain situations (see Vermont et al., 1991; Itoh et al., 1996; Gallop et al., 2003, among others).

To address some of the limitations or remaining gaps in the previous packages, we have implemented an R package called ***OptimalCutpoints*** (López-Ratón and Rodríguez-Álvarez, 2014; López-Ratón et al., 2014), specifically designed for selecting optimal cutpoints in continuous diagnostic tests. It is freely available from the Comprehensive R Archive Network (CRAN) at the URL http://CRAN.R-project.org/package=OptimalCutpoints. This package enables end-users to choose from among a considerable number of strategies (34) commonly used in clinical practice for optimal-cutpoint selection (see Table 5.2 in Section 5.1). *OptimalCutpoints* includes all the methods considered in the above mentioned packages plus others that have been proposed in the literature for selection of optimal values in diagnostic tests, such as criteria based on predictive values or likelihood ratios. Moreover, it incorporates several criteria that take into account the costs of the different diagnostic decisions as well as the prevalence of the disease under study.

All these packages cited above only consider the classical non-parametric empirical method for estimating the optimal cutpoint and accuracy measures, that is, none of these packages take into account recent methodology introduced in ROC curves (Molanes-López and Letón, 2011; Lai et al., 2012). So, in addition, we created the R package ***GsymPoint*** (López-Ratón et al., 2015b) for estimating the Generalized Symmetry point that includes two more efficient statistical methods than the empirical one, specifically it includes the two estimating methods previously presented in Chapter 4 for estimating the Generalized Symmetry point and its accuracy measures, a parametric method based on the Generalized Pivotal Quantity that assumes normality (Weerahandi, 1993, 1995; Lai et al., 2012) and a nonparametric method based on the Empirical Likelihood methodology (Thomas and Grunkemeier, 1975; Molanes-López and Letón, 2011). Therefore, we take into account in the estimation process if the assumption of normality is tenable, and we consider a cost-based generalization of one of the most widely used criteria for selecting the optimal cutpoint in clinical practice, the criterion where sensitivity and specificity accuracy measures are the same, taking into account the misclassification costs, a very important issue when selecting the optimal value in a specific clinical setting. The *GsymPoint* package is freely available at http://CRAN.R-project.org/package=GsymPoint.

In this chapter we explain in detail our two new R packages: *OptimalCutpoints* and *GsymPoint*.

5.1 The OptimalCutpoints Package

This section introduces ***OptimalCutpoints***, an `R` package in which most of the optimal criteria for selecting the threshold value previously presented in Chapter 3 have been incorporated in a way designed to be clear and user-friendly for the end-user. *OptimalCutpoints* provides numerical and graphical results. The numerical results include the optimal cutpoint according to the selected criterion, and the associated accuracy measures with their confidence intervals. The program's graphical output shows the ROC and PROC curves of the diagnostic test analyzed and, where possible, the plot of the pertinent criterion according to the different diagnostic marker values (candidates for the optimal cutpoint). In addition, *OptimalCutpoints* provides the option to automatically calculate the optimal thresholds according to the levels of certain (categorical) covariates and this will be illustrated in the next section with a biomedical example. This is of great interest because a diagnostic marker's discriminatory capacity can often depend on specific characteristics, such as a patient's age or gender, or the severity of the disease (Pepe, 2004). Moreover, no restriction has been imposed with respect to the range of values of the diagnostic test, i.e., it can take some values in a continuous range or a risk score obtained from a predictive diagnostic model (values from 0 to 1). Finally, insofar as computation is concerned, for all methods in this package, the optimal cut value obtained is always one of the observed diagnostic marker values, and the ROC and PROC curves and accuracy measures are empirically estimated.

In `R` language, programming is based on objects, and computations are basically functions that are specialized in performing specific calculations. Table 5.1 provides a summary of the main functions in the package.

5.1.1 `optimal.cutpoints()` function

The main function of the package is the `optimal.cutpoints` function, which uses the selected method(s) to compute the optimal cutpoint with its accuracy measures, and creates an object of class `optimal.cutpoints`. Usage is as follows:

```
optimal.cutpoints(X, status, tag.healthy, methods, data,
direction = c("<",">"), categorical.cov = NULL, pop.prev = NULL,
control = control.cutpoints(), ci.fit = FALSE, conf.level = 0.95,
trace = FALSE, ...)
```

The `X` argument is either a character string with the name of the diagnostic test variable or a formula. When `X` is a formula, it must be an object of class `formula`. The right side of $\sim$ must contain the name of the variable that distinguishes healthy from diseased individuals, and the left side of $\sim$ must contain the name of the diagnostic test variable. The `status` argument only applies when the `X` argument contains the name of the diagnostic test variable, and is a character string with the name of the variable that distinguishes healthy from diseased individuals. The `tag.healthy` argument is the value codifying healthy individuals in the `status` variable.

Function	Description
`optimal.cutpoints`	Computes the optimal cutpoint with its accuracy measures and, optionally, the pertinent confidence intervals for such measures.
`control.cutpoints`	Function used to set several parameters that control the optimal-cutpoint computing process.
`print`	Print method for objects fitted with `optimal.cutpoints`.
`summary`	Produces a summary of an `optimal.cutpoints` object.
`plot`	Plot method for objects fitted with `optimal.cutpoints`. Includes the plots of the ROC and PROC curves, indicating the optimal cutpoint on these plots.

Table 5.1: Summary of functions in the *OptimalCutpoints* package.

The `methods` argument is a character vector specifying which method/s is/are used for selecting optimal cutpoints. A total of 34 methods have been implemented in *OptimalCutpoints* (see Table 5.2). Various optimal-cutpoint selection methods can be selected simultaneously.

The `data` argument is a data frame which must, at minimum, contain the following variables: the variable that indicates the diagnostic marker; the true disease status (diseased/healthy); and whether adjustment is to be made for any (categorical) covariate of interest, a variable that indicates the levels of this covariate. A standard-type data input structure is used, with each row of the database indicating a patient/case and each column referring to a variable.

The `direction` argument is a character string specifying the direction in which the ROC curve must be computed. By default, individuals with a test value lower than the cutoff are classified as healthy (negative test), whereas patients with a test value greater than (or equal to) the cutoff are classified as diseased (positive test). If this is not the case, however, and the high values are related to health, this argument should be established at "`>`".

The `categorical.cov` argument is an optional argument, a character string with the name of the categorical covariate according to which optimal cutpoints are to be calculated. By default it is `NULL`, i.e., no categorical covariate is considered in the analysis.

Criterion name	Description
Criteria based on sensitivity and specificity measures	
`ValueSe`	A value set for sensitivity (Rutter and Miglioretti, 2003): the cutpoint c fulfilling the condition $Se(c) = \texttt{valueSe}$.
`ValueSp`	A value set for specificity (Rutter and Miglioretti, 2003): the cutpoint c fulfilling the condition $Sp(c) = \texttt{valueSp}$.
`SpEqualSe`	Sensitivity = specificity: the cutpoint c minimizing $\{\|Sp(c) - Se(c)\|\}$ (Amaro et al., 1995; Greiner, 1995; Hosmer and Lemeshow, 2000).
`MaxSe`	Maximizes sensitivity: the cutpoint c maximizing $Se(c)$ (Filella et al., 1995; Hoffman et al., 2000; Álvarez García et al., 2003).
`MaxSp`	Maximizes specificity: the cutpoint c maximizing $Sp(c)$ (Bortheiry et al., 1994; Hoffman et al., 2000).
`MaxSpSe`	Maximizes sensitivity and specificity simultaneously: the cutpoint c maximizing $\{min\{Sp(c), Se(c)\}\}$ (Riddle and Stratford, 1999; Gallop et al., 2003).
`Youden`	Youden index: the cutpoint c maximizing $YI(c) = Se(c) + Sp(c) - 1$ (Youden, 1950; Aoki et al., 1997; Greiner et al., 2000) or generalized Youden index: the cutpoint c maximizing $GYI(c) = Se(c) + \frac{1-p}{p}\frac{C_{FN}}{C_{FP}}Sp(c) - 1$ (Geisser, 1998; Greiner et al., 2000).
`MaxProdSpSe`	Maximizes the product of sensitivity and specificity: the cutpoint c maximizing $\{Sp(c)Se(c)\}$ (Lewis et al., 2008).
`Minimax`	Minimizes the most frequent error (Hand, 1987): the cutpoint c minimizing $\{\max\{p(1 - Se(c)), (1 - p)(1 - Sp(c))\}\}$.
`MinValueSe`	A minimum value set for sensitivity: the cutpoint c fulfilling the condition $Se(c) \geq \texttt{minValueSe}$ (Schäfer, 1989; Vermont et al., 1991; Gallop et al., 2003).
`MinValueSp`	A minimum value set for specificity: the cutpoint c fulfilling the condition $Sp(c) \geq \texttt{minValueSp}$ (Schäfer, 1989; Vermont et al., 1991; Gallop et al., 2003).
`MinValueSpSe`	A minimum value set for specificity and sensitivity (Schäfer, 1989): the cutpoint c fulfilling the condition $Sp(c) \geq \texttt{minValueSp}$ and $Se(c) \geq \texttt{minValueSe}$.

Criterion name	Description
`ROC01`	Minimizes distance between ROC plot and point (0,1): the cutpoint c minimizing $\{(Sp(c)-1)^2+(Se(c)-1)^2\}$ (Metz, 1978; Vermont et al., 1991).
`MaxDOR`	Maximizes Diagnostic Odds Ratio: the cutpoint c maximizing $DOR(c) = \frac{Se(c)}{1-Se(c)}\frac{Sp(c)}{1-Sp(c)}$ (Kraemer, 1992; Böhning et al., 2011).
`MaxEfficiency`	Maximizes efficiency or accuracy: the cutpoint c maximizing $Ef(c) = pSe(c)+(1-p)Sp(c)$ (Feinstein, SH, 1975; Galen, 1986; Greiner, 1995, 1996).
`MaxKappa`	Maximizes Kappa index (Cohen, 1960; Greiner et al., 2000) or Weighted Kappa index (Kraemer, 1992; Kraemer et al., 2002).
Criteria based on predictive values	
`ValueNPV`	A value set for negative predictive value: the cutpoint c fulfilling the condition $NPV(c) =$ `valueNPV`.
`ValuePPV`	A value set for positive predictive value: the cutpoint c fulfilling the condition $PPV(c) =$ `valuePPV`.
`NPVEqualPPV`	Negative predictive value = positive predictive value (Vermont et al., 1991): the cutpoint c minimizing $\|NPV(c) - PPV(c)\|$.
`MaxNPVPPV`	Maximizes negative predictive value and positive predictive value simultaneously: the cutpoint c maximizing $\{min\{NPV(c), PPV(c)\}\}$.
`MaxSumNPVPPV`	Maximizes the sum of negative predictive value and positive predictive value: the cutpoint c maximizing $\{NPV(c) + PPV(c)\}$.
`MaxProdNPVPPV`	Maximizes the product of negative predictive value and positive predictive value: the cutpoint c maximizing $\{NPV(c)PPV(c)\}$.
`MinValueNPV`	A minimum value set for negative predictive value (Vermont et al., 1991): the cutpoint c fulfilling the condition $NPV(c) \geq$ `minValueNPV`.
`MinValuePPV`	A minimum value set for positive predictive value (Vermont et al., 1991): the cutpoint c fulfilling the condition $PPV(c) \geq$ `minValuePPV`.

Criterion name	Description
`MinValueNPVPPV`	A minimum value set for predictive values (Vermont et al., 1991): the cutpoint c fulfilling the condition $NPV(c) \geq$ `minValueNPV` and $PPV(c) \geq$ `minValuePPV`.
`PROC01`	Minimizes distance between PROC plot and point (0,1) (Vermont et al., 1991; Gallop et al., 2003): the cutpoint c minimizing $\{(NPV(c)-1)^2 + (PPV(c)-1)^2\}$.
Criteria based on diagnostic likelihood ratios	
`ValueDLR.Negative`	A value set for negative diagnostic likelihood ratio: the cutpoint c fulfilling the condition $DLR_{-}(c) =$ `valueDLR.Negative` (Boyko, 1994; Rutter and Miglioretti, 2003).
`ValueDLR.Positive`	A value set for positive diagnostic likelihood ratio: the cutpoint c fulfilling the condition $DLR_{+}(c) =$ `valueDLR.Positive` (Boyko, 1994; Rutter and Miglioretti, 2003).
Criteria based on cost-benefit analysis of the diagnosis	
`CB`	Cost-benefit method, computing slope of ROC curve at optimal cutpoint, as $S = \frac{1-p}{p}\frac{C_{FP}-C_{TN}}{C_{FN}-C_{TP}}$ (McNeill et al., 1975; Metz et al., 1975; Metz, 1978).
`MCT`	Minimizes Misclassification Cost Term: the cutpoint c minimizing $MCT(c) = \frac{C_{FN}}{C_{FP}}p(1-Se(c)) + (1-p)(1-Sp(c))$ (Smith, 1991; Greiner, 1995, 1996).
Maximum Chi-squared or minimum p value criterion	
`MinPvalue`	Minimizes p value associated with the statistical χ^2 test which measures the association between the marker and the binary result obtained on using the cutpoint (Miller and Siegmund, 1982; Altman et al., 1994).
Prevalence-based methods	
`MeanPrev`	The closest value to the mean of the diagnostic test values. This criterion is usually used in cases where the diagnostic test takes values in the interval (0,1), i.e., the mean probability of occurrence, e.g., based on the results of a statistical model (Manel et al., 2001; Kelly et al., 2008).

Criterion name	Description
`ObservedPrev`	The closest value to the observed prevalence: the cutpoint c minimizing $\|c-p\|$, with p being prevalence estimated from the sample. This criterion is thus indicated/valid in cases where the diagnostic test takes values in the interval (0,1) (Manel et al., 2001).
`PrevalenceMatching`	The value for which predicted prevalence is practically equal to observed prevalence: the cutpoint c minimizing $\{\|p(1-Se(c))-(1-p)(1-Sp(c))\|\}$. This criterion is usually used in cases where the diagnostic test takes values in the interval (0,1), i.e., the predicted probability, e.g., based on a statistical model (Manel et al., 2001; Kelly et al., 2008).

Table 5.2: Available methods in the ***OptimalCutpoints*** package.

The `pop.prev` argument is the value of the disease's prevalence. By default it is `NULL`, and in such a case, when no value is introduced, prevalence is estimated on the basis of sample prevalence, taking into account the number of patients in the sample (cross-sectional study). However, the end-user can also specify a given value for prevalence, as, say, in other types of studies (case-control study) where it cannot be estimated on the basis of the sample. Where the `categorical.cov` is not `NULL`, the prevalence value can be specified by a single value if the same prevalence is assumed for the different levels of the covariate, or by a vector having as many components as levels if different values are assumed.

The `control` argument indicates the output of the `control.cutpoints` function, which controls the whole optimal-cutpoint calculation process. This function will be explained in detail in the following subsection.

The `ci.fit` argument is a logical value, and if it is `TRUE` then inference is performed on the accuracy measures at the optimal cutpoint (by default it is `FALSE`). Finally, `conf.level` is the value of the confidence level (1-α), and by default is equal to 0.95.

Summarizing, the `X`, `status`, `tag.healthy`, `methods` and `data` arguments of the `optimal.cutpoints` function are essential arguments, and if one or more is not introduced, this will lead to error. The remaining arguments (`direction`, `categorical.cov`, `pop.prev`, `control`, `ci.fit`, `conf.level` and `trace`) are optional and, in the event of having no value, will operate with the values established by default.

5.1.2 `control.cutpoints()` function

It should be noted that there are arguments that are specific to each method. We decided to include all of these in the `control` argument; `control` is a list of control values for the selection process designed to replace the default values yielded by the `control.cutpoints` function. The arguments of the `control.cutpoints` function, as well as the methods for which they apply, are shown in Table 5.3.

The values of the costs in general (necessary in criteria which make use of cost/benefit-based methodology, i.e., ``CB'', ``MCT'', ``Youden'' and ``MaxEfficiency''), and the costs ratio and costs of incorrect classifications in particular, must be indicated in the `costs.ratio`, `CFP` and `CFN` arguments, respectively. By default, the value 1 is established for all, a situation equivalent to classification costs not being considered. The values established by default for the accuracy measures (necessary in the ``MinValueSp'', ``ValueSp'', ``MinValueSe'', ``ValueSe'', ``MinValueSpSe'', ``MinValue NPV'', ``ValueNPV'', ``MinValuePPV'', ``ValuePPV'', ``MinValueNPVPPV'', ``ValueDLR.Positive'' and ``ValueDLR.Negative'' methods) are indicated in the `valueSp`, `valueSe`, `valueNPV`, `valu ePPV`, `valueDLR.Positive` and `valueDLR.Negative` arguments, respectively. By default, a value of 0.85 appears for sensitivity, specificity and predictive values measures, a value of 2 for the positive likelihood ratio, and a value of 0.5 for the negative likelihood ratio. These values were set on the basis of values usually indicated in the literature but end-users will have to set them in line with their own goals.

The `adjusted.pvalue` argument of the `control.cutpoints` function should be used in the ``MinPvalue'' method to indicate whether the Miller and Siegmund method (``PADJMS'' option) or Altman method (``PALT5'' and ``PALT10'' options) is selected for adjusting the p value (Miller and Siegmund, 1982; Altman et al., 1994). The default is ``PADJMS''. The first method uses the minimum p value (p_{min}) observed and the proportion (ϵ) of sample data which is below the lowest (ϵ_{low}) (or above the highest ϵ_{high}) cutpoint considered:

$$p_{acor} = \phi(z)(z - \frac{1}{z})log\left(\frac{\epsilon_{high}(1-\epsilon_{low})}{(1-\epsilon_{high})\epsilon_{low}}\right) + 4\frac{\phi(z)}{z},$$

where z is the $(1 - p_{min}/2)$ quantile of the standard normal distribution, and ϕ denotes the density function of the standard normal. Altman et al. (1994) furnished the following simplifications of the above formula that work well for low minimum p values $(0.0001 < p_{min} < 0,1)$ and are easily applicable: For $\epsilon = \epsilon_{low} = \epsilon_{high} = 5\% : p_{alt5} = -3.13p_{min}(1 + 1.65ln(p_{min}))$ and for $\epsilon = \epsilon_{low} = \epsilon_{high} = 10\% : p_{alt10} = -1.63p_{min}(1 + 2.35ln(p_{min}))$.

Various approaches are considered in *OptimalCutpoints* for calculating the confidence intervals of the accuracy measures. The `ci.SeSp`, `ci.PV` and `ci.DLR` arguments in the `control.cutpoints` function indicate the methods selected for computing confidence intervals for Se/Sp, PPV/NPV and DLR_+/DLR_-, respectively. They are meaningful only when the argument `ci.fit` is `TRUE`.

Argument	Method	Description

`CFP`	`MCT`
	`Youden`
	`MaxKappa`
		A numerical value specifying the cost of a false positive decision C_{FP}.
		The default value is 1.

`CFN`	`MCT`
	`Youden`
	`MaxKappa`
		A numerical value specifying the cost of a false negative decision C_{FN}.
		The default value is 1.

`costs.ratio`	`CB`	A numerical value specifying the cost ratio $CR = \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}$. The default value is 1.
`costs.benefits.Youden`	`Youden`	A logical value. If `TRUE`, the optimal cutpoint based on cost-benefit methodology is computed. The default is `FALSE`.
`costs.benefits.Efficiency`	`MaxEfficiency`	A logical value. If `TRUE`, the optimal cutpoint based on cost-benefit methodology is computed. The default is `FALSE`.
`generalized.Youden`	`Youden`	A logical value. If `TRUE`, the generalized Youden index is computed. The default is `FALSE`.
`weighted.Kappa`	`MaxKappa`	A logical value. If `TRUE`, the Weighted Kappa index is computed. The default is FALSE.

`valueSe`	`MinValueSe`
	`ValueSe`
	`MinValueSpSe`
		A numerical value specifying the (minimum or specific) value set for sensitivity.
		The default value is 0.85.

Argument	Method	Description

`valueSp`	`MinValueSp`
	`ValueSp`
	`MinValueSpSe`
		A numerical value specifying the (minimum or specific) value set for specificity. The default value is 0.85.

`valueNPV`	`MinValueNPV`
	`ValueNPV`
	`MinValueNPVPPV`
		A numerical value specifying the minimum value set for negative predictive value. The default value is 0.85

`valuePPV`	`MinValuePPV`
	`ValuePPV`
	`MinValueNPVPPV`
		A numerical value specifying the minimum value set for positive predictive value. The default value is 0.85.

`valueDLR.Positive`	`ValueDLR.Positive`	A numerical value specifying the value set for the positive diagnostic likelihood ratio. The default value is 2.
`valueDLR.Negative`	`ValueDLR.Negative`	A numerical value specifying the value set for the negative diagnostic likelihood ratio. The default value is 0.5.
`maxSp`	`MinValueSpSe`	A logical value meaningful only in a case where there is more than one cutpoint fulfilling the conditions. If `TRUE`, those of the cutpoints which yield maximum specificity are computed. Otherwise, the cutoff that yields maximum sensitivity is computed. The default is `TRUE`.
`maxNPV`	`MinValueNPVPPV`	A logical value meaningful only in the case where there is more than one cutpoint fulfilling the conditions. If `TRUE`, those of the cutpoints which yield the maximum negative predictive value are computed. Otherwise, the cutoff that yields the maximum positive predictive value is computed. The default is `TRUE`.

Argument	Method	Description
ci.SeSp	All methods	A character string meaningful only when the argument ci.fit of the optimal.cutpoints function is TRUE. It indicates the method for estimating the confidence interval for sensitivity and specificity measures. Options are ''Exact'' (Clopper and Pearson, 1934), ''Quadratic'' (Fleiss, 1981), ''Wald'' (Wald and Wolfowitz, 1939), ''AgrestiCoull'' (Agresti and Coull, 1998) and ''RubinSchenker'' (Rubin and Schenker, 1987). The default is ''Exact''.
ci.PV	All methods	A character string meaningful only when the argument ci.fit of the optimal.cutpoints() function is TRUE. It indicates the method for estimating the confidence interval for predictive values. Options are ''Exact'' (Clopper and Pearson, 1934), ''Quadratic'' (Fleiss, 1981), ''Wald'' (Wald and Wolfowitz, 1939), ''AgrestiCoull'' (Agresti and Coull, 1998), ''RubinSchenker'' (Rubin and Schenker, 1987), ''Transformed'' (Simel et al., 1991), ''NotTransformed'' (Koopman, 1984) and ''GartNam'' (Gart and Nam, 1998). The default is ''Exact''.
ci.DLR	All methods	A character string meaningful only when the argument ci.fit of the optimal.cutpoints function is TRUE. It indicates the method for estimating the confidence interval for diagnostic likelihood ratios. Options are ''Transformed'' (Simel et al., 1991), ''NotTransformed'' (Koopman, 1984) and ''GartNam'' (Gart and Nam, 1998). The default is ''Transformed''.
adjusted.pvalue	MinPvalue	A character string specifying the method for adjusting the p value, i.e., ''PADJMS'' for the Miller and Siegmund method (Miller and Siegmund, 1982), and ''PALT5'', ''PALT10'' for the Altman method (Altman et al., 1994). The default is ''PADJMS''.

Argument	Method	Description
standard.deviation.accuracy	MaxEfficiency	A logical value. If TRUE, standard deviation associated with accuracy at the optimal cutpoint is computed. The default is FALSE.

Table 5.3: Summary of arguments of the control.cutpoints function.

In the ci.SeSp argument, the options are ''Exact'', ''Quadratic'', ''Wald'', ''AgrestiCoull'' and ''RubinSchenker''. ''Exact'' is the exact confidence interval based on the exact distribution of a proportion (Clopper and Pearson, 1934). It should be noted that this method cannot be applied for proportions where the numerator or the difference between the denominator and the numerator is equal to zero. If this occurs for any value of the corresponding accuracy measure, i.e., the sensitivity or the specificity, the program shows a warning message and returns a NaN for the limit of the confidence interval that could not be computed. It is worth noting, however, that this problem only happens for values of sensitivity/specificity equal to zero or one, i.e., on values which are not of interest in clinical practice. ''Quadratic'' refers to Fleiss' quadratic confidence interval (Fleiss, 1981), based on the asymptotic normality of the estimator of a proportion but adding a continuity correction, and this approach is valid in a situation where both the numerator and the difference between the denominator and the numerator of the proportion are greater than 5. ''Wald'' indicates Wald's confidence interval (Wald and Wolfowitz, 1939) with continuity correction, based on maximum-likelihood estimation of a proportion, and adding a continuity correction; it is valid where the numerator and the difference between the denominator and numerator are greater than 20. Similarly to the ''Exact'' method, when ''Quadratic'' or ''Wald'' approaches are not valid for any value of the corresponding accuracy measure, the program shows a warning message. However, in these cases the confidence intervals are computed. We therefore recommend the user to check the conditions under which these methods are valid at the optimal cutpoint. The ''AgrestiCoul l'' option computes the confidence interval proposed by Agresti and Coull (1998), and is a score confidence interval that does not use the standard calculation for the binomial proportion. Finally, ''RubinSchenker'' means Rubin and Schenker's logit confidence interval (1987), and uses logit transformation and bayesian arguments with an a priori Jeffreys distribution. The default is ''Exact''.

In the ci.DLR argument, ''Transformed'', ''NotTransformed'' and ''Gart Nam'' are the available options. ''Transformed'' indicates the confidence interval based on the logarithmic transformation of the diagnostic likelihood ratios (Simel et al., 1991), ''NotTransformed'' is the confidence interval without transformation (Koopman, 1984), and ''GartNam'' is the confidence interval based on the calculation of the interval for the ratio of two independent proportions (Gart and Nam, 1998). The default

is ``Transformed''. Inference of the predictive values depends on the type of study, i.e., whether cross-sectional (prevalence can be estimated on the basis of the sample) or case-control. In the former case, the approaches for calculating the confidence intervals of the predictive values are the same as for the sensitivity and specificity measures. Accordingly, in such a case, the possible options for the `ci.PV` argument are ``Exact'', ``Quadratic'', ``Wald'', ``AgrestiCoull'' and ``RubinSchenker''. In a case-control study, however, the confidence intervals of the predictive values should be based on the intervals of the likelihood ratios. Therefore, the available options are ``Transformed'', ``NotTransformed'' and ``GartNam''. The default value is ``Exact''.

For greater detail, the Help Manual of the `optimal.cutpoints` and `control.cutpoints` functions can be consulted. In addition, the last subsection gives an illustration of the use of these two functions, based on the real biomedical CAD example.

Numerical and graphical summaries of the created object can be obtained by using the `summary.optimal.cutpoints` function or `summary` method, `print.optimal.cutpoints` function or `print` method and `plot.optimal.cutpoints` function or `plot` method.

5.1.3 `summary.optimal.cutpoints()` function

Numerical results are printed on the screen, and the output yielded by the `summary.optimal.cutpoints` function or `summary` method always includes:

- The matched call to the main function `optimal.cutpoints`
- The AUC value with its confidence interval (Delong et al., 1988)
- Information relating to the optimal cutpoint:
  - The method used for selecting the optimal value together with the number of optimal cutpoints (in some cases there may be more than one value)
  - The optimal cutoff(s) and its/their accuracy-measure estimates (Se, Sp, etc.). Furthermore, accuracy measures will be accompanied by their confidence intervals, if the `se.fit` argument is `TRUE`
  - The number of false positive and false negative classifications; and, where possible, the value of the optimal criterion.

  All this information will be shown for each level of categorical covariate, if specified.

The call to the function is as follows:

```
R> summary(object, ...)
```

The `object` argument is of class `optimal.cutpoints` as produced by `optimal.cutpoints` function.

`...` indicates further arguments passed to or from other methods. None are used in this method.

5.1.4 `plot.optimal.cutpoints()` function

The graphical output is yielded by the `plot.optimal.cutpoints` function or the `plot` method. This function plots empirical ROC and PROC curves and, where possible, the plot of the chosen optimal criterion versus all the different test values (candidates for the optimal cutpoint). Usage is as follows:

```
R> plot(x, legend = TRUE, which = c(1,2), ...)
```

The `x` argument is an object of class `optimal.cutpoint` as produced by `optimal.cutpoints` function.

The argument `legend` is a logical value for including the legend of optimal coordinates with specific characteristics and by default is `TRUE`.

`which` is a numeric vector with the required plots and by default, both the ROC and the PROC curves are plotted.

`...` indicates further arguments passed to method `plot.default`.

5.1.5 Technical features

In this subsection, certain specific characteristics of some methods and the behavior of the package in such cases are briefly explained. The methods in which a minimum value is set for sensitivity, specificity or the predictive values (the ``MinValueSe'', ``MinValueSp'', ``MinValuePPV'' and ``MinValueNPV'' methods, respectively), can take several or even zero values. In the latter case, an error message is shown and the user can enter a new minimum value, if desired. In a case where there is more than one cutpoint fulfilling the condition, that which maximizes the other measure is chosen as the optimal cutpoint(s). For example, in the ``MinValueSp'' method, if there is more than one cutpoint with $Sp \geq$ `minValueSp`, that which yields the maximum sensitivity is chosen. So, the cutpoint(s) that achieves the highest sensitivity and specificity under the condition $Sp \geq$ `minValueSp` are finally chosen.

The same behavior has been used for the methods that set minimum values for both the sensitivity and specificity measures (``MinValueSpSe'' method) or for both predictive values (``MinValueNPVPPV'' method). The only difference is that if there is more than one cutpoint fulfilling these conditions, those which yield maximum sensitivity or maximum specificity (in ``MinValueSpSe'') or maximum predictive positive value or negative predictive value (in ``MinValueNPVPPV'') are chosen. The user can select one of the two options by means of the `maxSp` and `maxNPV` arguments in the `control.cutpoints` function, respectively (see Table 5.3). If `TRUE` (the default value), the cutpoint/s yielding maximum specificity or maximum negative predictive

value is/are computed as the optimal cutpoint(s).

Finally, it should be noted that, in the last version of the package, only criteria proposed in the literature that provide different optimal values have been included. For instance, the ``Youden'' method is identical (from an optimization point of view) to the method that maximizes the sum of sensitivity and specificity (Albert and Harris, 1987; Zweig and Campbell, 1993) and to the criterion that maximizes concordance, which is a function of the AUC defined as $Se+Sp-0.5$ (Begg et al., 2000; Gönen and Sima, 2013). So, these last two methods are not included in the package. Similarly, ``MaxProdSpSe'' is the same as the method which maximizes the accuracy area just defined as the product of sensitivity and specificity (Lewis et al., 2008). Moreover, the method that maximizes efficiency or accuracy (``MaxEfficiency'' method in *OptimalCutpoints*) provides the same optimal cutpoint as the method that minimizes the classification error rate (Metz, 1978). Thus, the method based on the accuracy area and the method of minimum error rate have not been incorporated in *OptimalCutpoints*.

5.1.6 Biomedical application

This subsection describes the application of the *OptimalCutpoints* R package. To illustrate the use of this package, we shall consider the study that sought to investigate the clinical usefulness of leukocyte elastase for diagnosis of coronary artery disease (CAD) (Amaro et al., 1995) as example for illustrating the use of this package. Usefulness refers to the practical value of information when it comes to managing patients. The main research question here is to select optimal cutpoints for leukocyte elastase concentrations at the date of diagnosing patients with CAD. Depending on a predetermined elastase concentration cutpoint, subjects are chosen for coronary angiography. We remind here that 96 individuals had coronary lesions (diseased patients) and 45 had non-stenotic coronaries (non-diseased patients). Since it is well established that elastase concentrations behave differently according to gender, the analyses were performed here separately for male s and females.

The first step for using the package consists of loading the *OptimalCutpoints* package and the data set in R (in this case, the elas data is included in the package):

```
R> library("OptimalCutpoints")

R> data("elas")
```

If we want to view here summary statistics of the variables included in the data set:

```
R> summary(elas)

      elas            status          gender
 Min.   :  5.00   Min.   :0.0000   Female: 37
 1st Qu.: 27.00   1st Qu.:0.0000   Male  :104
 Median : 39.00   Median :1.0000
 Mean   : 43.28   Mean   :0.6809
 3rd Qu.: 51.00   3rd Qu.:1.0000
```

```
 Max.   :163.00   Max.   :1.0000
```

To compute the optimal cutpoint using the elas data set, simply use the syntax shown below:

```
R> cutpoint1 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ categorical.cov = "gender", pop.prev = NULL, control =
+ control.cutpoints(), ci.fit = TRUE)
```

In this case, by way of example, two methods are considered for calculating the optimal cutpoint, namely, the method based on the Youden index, and the sensitivity-specificity equality criterion, since these are two of the best-known and most widely used methods in clinical practice (Youden, 1950; Greiner, 1995; Aoki et al., 1997; Shapiro, 1999; Greiner et al., 2000). If no adjustment is made for any categorical covariate, one is left with the default value of the `categorical.cov` argument: `categorical.cov=NULL`, and so there is no need for it to be indicated. As the intention here, however, is to perform separate analyses for males and females, the `categorical.cov = ''gender''` must be indicated. As this example involves a cross-sectional study, disease prevalence is estimated on the basis of sample prevalence, and so the default value, `pop.prev = NULL`. In another type of study, end-users could indicate a given value for the population prevalence. Fin ally, as the argument `ci.fit` is `TRUE`, the confidence intervals for the accuracy measures are computed. By default, the `''Exact''` method is used for the sensitivity, the specificity and the predictive values. As pointed out before, this method cannot be applied in those situations where the TPs, the FPs, the TNs or the FNs are equal to zero. In these cases, the program produces a warning, and returns a `NaN` for the limit of the confidence interval that could not be computed. Specifically, in this example eight warning messages appear: four warnings for females (for the sensitivity, the specificity, and the positive and negative predictive values), and four similar warnings for males.

The `cutpoint1` object is a list that consists of the following components:

```
R> names(cutpoint1)

[1] "Youden" "SpEqualSe"  "methods"  "levels.cat" "call"  "data"
```

The component `''methods''` is a character vector with the value of the argument `methods` used in the call; `''levels.cat''` is a character vector indicating the levels of the categorical covariate; `''call''` is the matched call; and finally, `''data''` is the data frame used in the analysis. The first two components (`''Youden''` and `''SpEqualSe''`) contain the results associated with each of the methods selected for computing the optimal cutpoint. In this case, each of these components is itself a two-component list (for `''Male''` and `''Female''`) containing:

```
R> names(cutpoint1$Youden$Male)

"measures.acc" "optimal.cutoff" "criterion" "optimal.criterion"
```

Each of the previous components contains the following information:

(i) ``measures.acc'': a list with all cutoffs, their accuracy measures (Se, Sp, PPV, NPV, DLR_+ and DLR_-), the AUC, and prevalence and sample size in healthy and diseased populations:

```
R> names(cutpoint1$Youden$Male$measures.acc)

[1] "cutoffs"  "Se"   "Sp"   "PPV"   "NPV"
[6] "DLR.Positive" "DLR.Negative" "AUC"  "pop.prev"  "n"
```

(ii) ``optimal.cutoff'': a list with the optimal cutoff(s), its/their accuracy measures (Se, Sp, PPV, NPV, DLR_+ and DLR_-), and the number of false positive and false negative decisions:

```
R> names(cutpoint1$Youden$Male$optimal.cutoff)

[1] "cutoff"   "Se"   "Sp"   "PPV"     "NPV"
[6] "DLR.Positive" "DLR.Negative" "FP"   "FN"
```

(iii) ``criterion'': the numerical value of the method considered for selecting the optimal cutpoint for each cutoff:

```
R> cutpoint1$Youden$Male$criterion

      1          2          3          4          5          6          7
0.00000000 0.04347826 0.08695652 0.10574342 0.09339775 0.13687601 0.18035427
         8          9         10         11         12         13         14
0.22383253 0.25496511 0.23027375 0.21792807 0.19323671 0.14385400 0.22490607

[...]

        50         51         52         53         54         55         56
0.08641975 0.07407407 0.06172840 0.04938272 0.03703704 0.02469136 0.01234568
```

(iv) ``optimal.criterion": the numerical value of the criterion at the optimal cutpoint:

```
R > cutpoint1$Youden$Male$optimal.criterion
[1] 0.3934514
```

Accordingly, end-users can easily access each of these components, e.g., to see the sensitivity values for all cut values in males:

```
R> cutpoint1$Youden$Male$measures.acc$Se
```

or

```
R> cutpoint1$SpEqualSe$Male$measures.acc$Se
```

Numerical results

A numerical summary of the results can be obtained by calling up the `print.optimal.cutpoints` or `summary.optimal.cutpoints` functions, which can be abbreviated by `print` and `summary` methods:

```
R> summary(cutpoint1)

Call:
optimal.cutpoints.default(X = "elas", status = "status",
    tag.healthy = 0, methods = c("Youden", "SpEqualSe"),
    data = elas, categorical.cov = "gender", pop.prev = NULL,
    control = control.cutpoints(), ci.fit = TRUE)

**************************************************************
Female
**************************************************************

Area under the ROC curve (AUC):  0.818 (0.684, 0.952)

CRITERION: Youden
Number of optimal cutoffs: 1

                   Estimate 95% CI lower limit 95% CI upper limit
cutoff           46.0000000                 -                 -
Se                0.6666667         0.3838037         0.8817589
Sp                0.8181818         0.5971542         0.9481327
PPV               0.7142857         0.4516107         0.9031160
NPV               0.7826087         0.5285570         0.9359958
DLR.Positive      3.6666667         1.4096667         9.5373214
DLR.Negative      0.4074074         0.1939348         0.8558589
FP                4.0000000                 -                 -
FN                5.0000000                 -                 -
Optimal criterion 0.4848485                 -                 -

CRITERION: SpEqualSe
Number of optimal cutoffs: 1

                     Estimate 95% CI lower limit 95% CI upper limit
cutoff           41.00000000                 -                 -
Se                0.73333333         0.4489968         0.9221285
Sp                0.68181818         0.4512756         0.8613535
PPV               0.61111111         0.3762084         0.8712446
NPV               0.78947368         0.5263331         0.9157684
DLR.Positive      2.30476190         1.1634481         4.5656764
DLR.Negative      0.39111111         0.1611869         0.9490094
FP                7.00000000                 -                 -
```

```
FN                 4.00000000                -                -
Optimal criterion  0.05151515                -                -

****************************************************************
Male
****************************************************************

Area under the ROC curve (AUC):  0.722 (0.612, 0.831)

CRITERION: Youden
Number of optimal cutoffs: 1

                   Estimate 95% CI lower limit 95% CI upper limit
cutoff           38.0000000                -                -
Se                0.6543210        0.5404147        0.7565737
Sp                0.7391304        0.5159480        0.8977139
PPV               0.8983051        0.7686835        0.9355009
NPV               0.3777778        0.2738718        0.6528595
DLR.Positive      2.5082305        1.2382379        5.0807845
DLR.Negative      0.4676834        0.3180320        0.6877538
FP                6.0000000                -                -
FN               28.0000000                -                -
Optimal criterion 0.3934514                -                -

CRITERION: SpEqualSe
Number of optimal cutoffs: 1

                     Estimate 95% CI lower limit 95% CI upper limit
cutoff           36.00000000                -                -
Se                0.66666667        0.5531734        0.7675667
Sp                0.60869565        0.3854190        0.8029236
PPV               0.85714286        0.7075091        0.9083152
NPV               0.34146341        0.2429772        0.5759222
DLR.Positive      1.70370370        1.0003372        2.9016279
DLR.Negative      0.54761905        0.3492857        0.8585711
FP                9.00000000                -                -
FN               27.00000000                -                -
Optimal criterion 0.05797101                -                -
```

In this case, the `summary.optimal.cutpoints` function or the `summary` method displays the information relating to the optimal cutpoint, i.e., the methods used for selecting the optimal value here were ``Youden'' and ``SpEqualSe'', and the optimal cutpoints as well as their accuracy measures are shown for both methods and for males and females. Optimal criteria refer to the Youden index (i.e., the value $Se + Sp - 1$ at optimal cutpoint) and the $Se - Sp$ difference at optimal cutoff. It must be borne in mind that, as the choice of the cutpoint is made on the basis of the empirical ROC curve, it is

not always possible to obtain the value c of the sample for which $Se(c)$=$Sp(c)$; here the cutpoint $c/min_c\{|\widehat{Se}(c) - \widehat{Sp}(c)|\}$ is obtained.

As in the call to the function, `ci.fit = TRUE`, the accuracy measures -as pointed out above- appear accompanied by their corresponding confidence intervals. By default, confidence intervals for the AUC and accuracy measures are calculated for a confidence level α of 0.95, though this value can be changed in the `conf.level` argument of the main `optimal.cutpoints` function. Moreover, the exact confidence interval (Clopper and Pearson, 1934) is being calculated by default for the sensitivity, specificity and predictive values measures (in this case the inference for these values is correct, since it is a cross-sectional study; if it were a case-control study, however, the method of calculating the confidence interval for the predictive values would have to be changed), and for the interval based on the transformation logarithm for the likelihood ratios (Simel et al., 1991). These methods for calculating confidence intervals can be changed by means of the `ci.SeSp`, `ci.PV` and `ci.DLR` arguments, which appear in the `control.cutpoints` function. Hence, if one wanted to change the method of calculating the confidence interval for the sensitivity and specificity measures, one would only have to indicate this using the following syntax:

```
R> cutpoint2 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ pop.prev = NULL, categorical.cov = "gender", control =
+ control.cutpoints(ci.SeSp = "AgrestiCoull"), ci.fit = TRUE)

R> summary(cutpoint2)
```

In this case, the results would be as follows:

```
Call:
optimal.cutpoints.default(X = "elas", status = "status",
    tag.healthy = 0, methods = c("Youden", "SpEqualSe"),
    data = elas, categorical.cov = "gender", pop.prev = NULL,
    control = control.cutpoints(ci.SeSp = "AgrestiCoull"),
    ci.fit = TRUE)

***********************************************
Female
***********************************************
Area under the ROC curve (AUC):  0.818 (0.684, 0.952)

CRITERION: Youden
Number of optimal cutoffs: 1

                   Estimate 95% CI lower limit 95% CI upper limit
cutoff           46.0000000                   -                   -
Se                0.6666667           0.4171355           0.8482368
Sp                0.8181818           0.6148339           0.9269312
```

```
[...]

CRITERION: SpEqualSe
Number of optimal cutoffs: 1

                    Estimate 95% CI lower limit 95% CI upper limit
cutoff           41.00000000                 -                 -
Se                0.73333333         0.4804957         0.8910255
Sp                0.68181818         0.4731860         0.8363941
[...]

************************************************
Male
************************************************
Area under the ROC curve (AUC):  0.722 (0.612, 0.831)

CRITERION: Youden
Number of optimal cutoffs: 1

                     Estimate 95% CI lower limit 95% CI upper limit
cutoff           38.0000000                 -                 -
Se                0.6543210         0.5458938         0.7487735
Sp                0.7391304         0.5353000         0.8745138
[...]

CRITERION: SpEqualSe
Number of optimal cutoffs: 1

                    Estimate 95% CI lower limit 95% CI upper limit
cutoff           36.00000000                 -                 -
Se                0.66666667         0.5585284         0.7597123
Sp                0.60869565         0.4078552         0.7784238
[...]
```

Note that the exploration of the usefulness of medical information involves many factors which, rather than being properties of the test system, are instead properties of the circumstances of the clinical application. With respect to clinical interpretation, the following result was obtained: leukocyte elastase determination was shown to be a test that displays good "diagnostic accuracy" or ability to discriminate between patients with and without CAD, particularly in women (AUC=0.82 in women versus AUC=0.72 in men). The cutpoint obtained using the criterion based on the Youden index for women was 46 μgl^{-1}. Accordingly, women with an elastase value higher than or equal to 46 were classified as patients with CAD. Using this cutpoint, 82% of women without CAD and 67% of women with CAD were correctly classified (4 false positive and 5 false negative classifications). Furthermore, 78% of women who registered a negative test result (i.e., an elastase value lower tha n 46) did not really present with CAD, while 71% of women with a positive result did in fact present with the disease. The likelihood of a female having

CAD increased by 3.67 in the case of a positive test result and, conversely, decreased by 0.41 in the case of a negative test result.

In the case of men, the cutpoint obtained using the criterion based on the Youden index was 38 μgl^{-1}, a value lower than that obtained for women. This means that men with elastase values higher than or equal to 38 μgl^{-1} were classified as CAD patients. On the basis of this value, 65% of the men who presented with the disease were correctly classified by determination of elastase (positive value) and 74% of those who did not present with CAD were likewise correctly classified (negative value, elastase lower than 38 μgl^{-1}). Of the men who registered a positive elastase value, almost 90% really had CAD but among those who registered a negative elastase value, only 37% did not really have the disease. Moreover, the likelihood of a male having CAD was 2.5-fold if the test result was positive (elastase $\geq$ 38 μgl^{-1}) and 0.47-fold if the test result was negative.

If, instead of the Youden criterion, one applies the method that selects the cutoff at which sensitivity is equal to specificity, one obtains optimal cutpoints lower than those previously obtained (equal to 41 μgl^{-1} in women and 36 μgl^{-1} in men), and the same conclusions can be drawn as in the previous case. The only difference is that, when the optimal value falls, the measures of sensitivity (and thus the false positive decisions) and negative predictive value increase, while the measures of specificity (and false negatives), positive predictive value and likelihood ratios all decrease.

The Youden index can also be interpreted from a cost-benefit analysis perspective. The slope of the ROC curve at the optimal cutpoint obtained using this index is equal to 1 (Perkins and Schisterman, 2006), which is equivalent to having a prevalence equal to 0.5 and a costs ratio equal to 1. If one wished to calculate the optimal value taking this into account, one would have to specify that the `costs.benefits.Youden` argument of the `control.cutpoints` function was set as `TRUE`:

```
R> cutpoint3 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ pop.prev = NULL, categorical.cov = "gender", control =
+ control.cutpoints(costs.benefits.Youden = TRUE), ci.fit = TRUE)
```

The Youden index gives equal weight to sensitivity and specificity. Sometimes, however, different weights are suitable (based on the cost of the different types of error and the prevalence of the disease), and in such a case the generalized Youden index can be used. This possibility was also implemented in the *OptimalCutpoints* package, within the `''Youden''` method. For the purpose, this only has to be indicated in the `generalized.Youden` argument of the `control.cutpoints` function.
If `generalized.Youden = TRUE`, the generalized Youden index is computed. In the absence of a value being indicated for costs of incorrect diagnostic decisions or prevalence, $C_{FP} = C_{FN} = 1$ is considered by default (a situation equivalent to having no costs), and prevalence is estimated on the basis of the sample, in this case p = 0.77885 in males, and p = 0.40541 in females. If the end-user wishes to specify any given costs, these must be in dicated in the C_{FP} and C_{FN} arguments of the `control.cutpoints` function, within the `control` argument in the main function, e.g.:

```
R> cutpoint4 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ pop.prev = NULL, categorical.cov = "gender", control =
+ control.cutpoints(generalized.Youden = TRUE, CFP = 1, CFN = 3),
+ ci.fit = TRUE)
```

So, in this case, assuming that an FN result has triple the cost of an FP result, the Youden index would yield some cutpoints that were lower than those obtained without considering misclassification costs, (25 μgl^{-1} in women and 13 μg^{-1} in men). Hence, with these optimal values, the presence of false negatives (zero false negative decisions) is avoided, and some maximum values (equal to 1) are attained for sensitivity and also the negative predictive value.

To change the value of the population prevalence, this only has to be directly indicated in the `pop.prev` argument of the `optimal.cutpoints` function. For instance, with a prevalence equal to 0.5 and costs equal to 1, the generalized Youden index is equivalent to the Youden index (the results would be the same):

```
R> cutpoint5 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ pop.prev = 0.5, categorical.cov = "gender", control =
+ control.cutpoints(generalized.Youden = TRUE), ci.fit = TRUE)
```

It could also be indicated as follows, by using a two-component vector (number of levels of the covariate) to specify the prevalence:

```
R> cutpoint5 <- optimal.cutpoints(X = "elas", status = "status",
+ tag.healthy = 0, methods = c("Youden","SpEqualSe"), data = elas,
+ pop.prev = c(0.5,0.5), categorical.cov = "gender", control =
+ control.cutpoints(generalized.Youden = TRUE), ci.fit = TRUE)
```

Graphical results

The graphical output of the results can be obtained by calling up the `plot.optimal.cutpoints` function, which can be abbreviated by the `plot` method:

```
R> plot(cutpoint1)
```

By default, the `plot` method depicts the plots of the ROC and PROC curves (`which = c(1,2)`). However, the plot of the values of the optimal criterion as a function of the cutoffs can, where applicable, be obtained by specifying the argument `which = 3`:

```
R> plot(cutpoint1, which = 3, ylim = c(0,1))
```

Figures 5.1 and 5.2 show the figures that appear as a result of the above calls in females and males, respectively. This is the default output but the end-user can add specific graphic parameters, such as color, legend, etc.

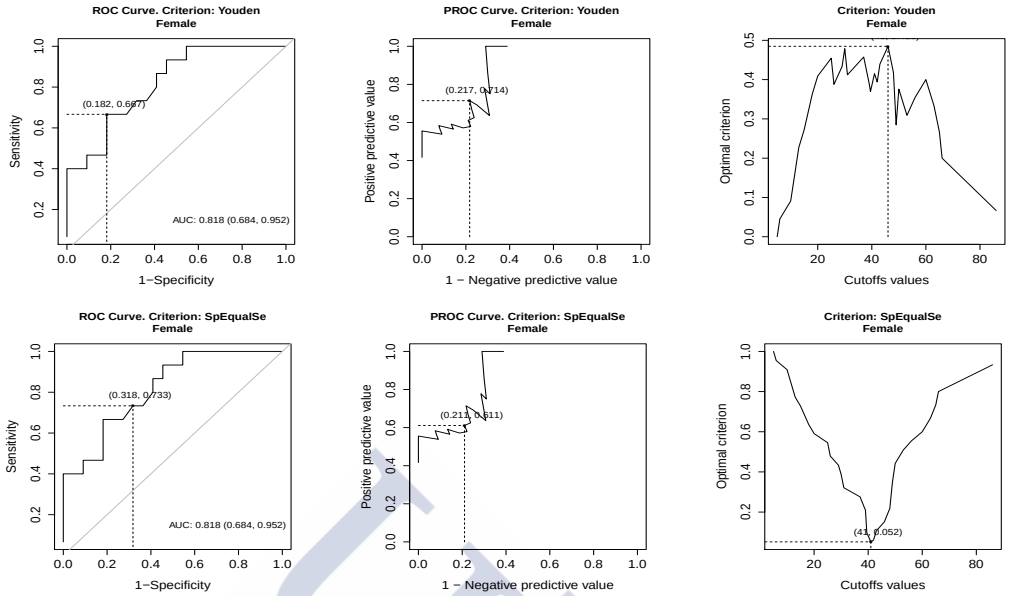


Figure 5.1: Graphical output in females. *From left to right*: ROC curve, PROC curve and optimal criterion (according to cutpoints). *From top to bottom*: the Youden index and sensitivity-specificity equality criteria.

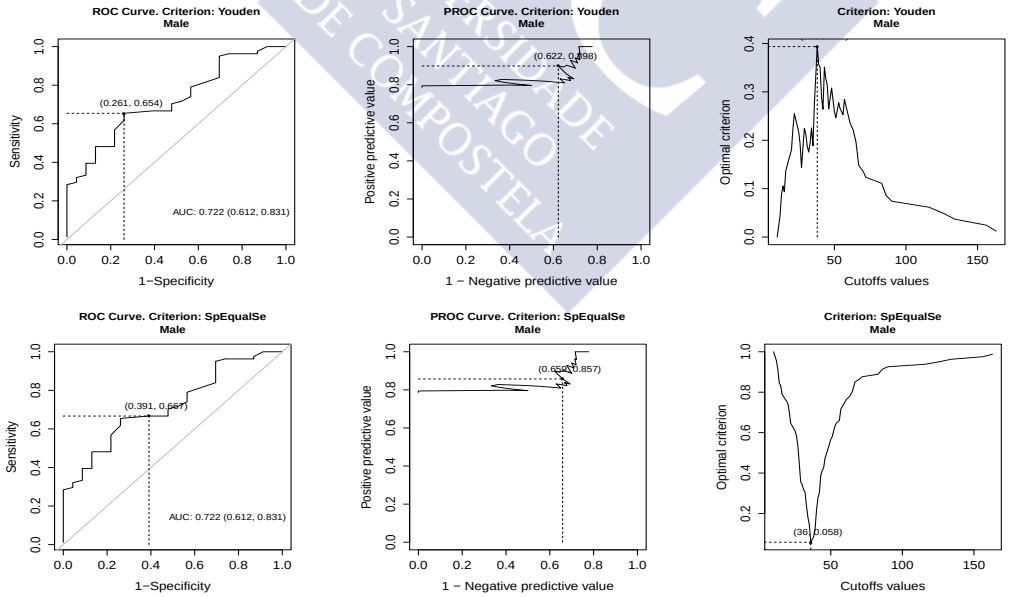


Figure 5.2: Graphical output in males. *From left to right*: ROC curve, PROC curve and optimal criterion (according to cutpoints). *From top to bottom*: the Youden index and sensitivity-specificity equality criteria.

Furthermore, other figures are also possible, e.g., in the method in which sensitivity and specificity are equal, the plot of these measures (jointly) according to the cutpoint may be of interest. This graph can be created on the basis of the pertinent components yielded with the `optimal.cutpoints` function. For instance, for males, the code is as follows:

```
R> plot(cutpoint1$SpEqualSe$Male$measures.acc$cutoffs,
+ cutpoint1$SpEqualSe$Male$measures.acc$Se[,1],
+ xlab = "Cutpoint", ylab = "Sensitivity and Specificity",
+ type = "l", lty = 2, main = "Sensitivity and Specificity
+ Males")

R> lines(cutpoint1$SpEqualSe$Male$measures.acc$cutoffs,
+ cutpoint1$SpEqualSe$Male$measures.acc$Sp[,1], xlab ="Cutpoint",
+ ylab = "Sensitivity and Specificity", type = "l")

R> legend("topright", legend = c("Se","Sp"), lty = c(2,1),
+ bty = "n")
```

And similarly for females (see Figure 5.3).

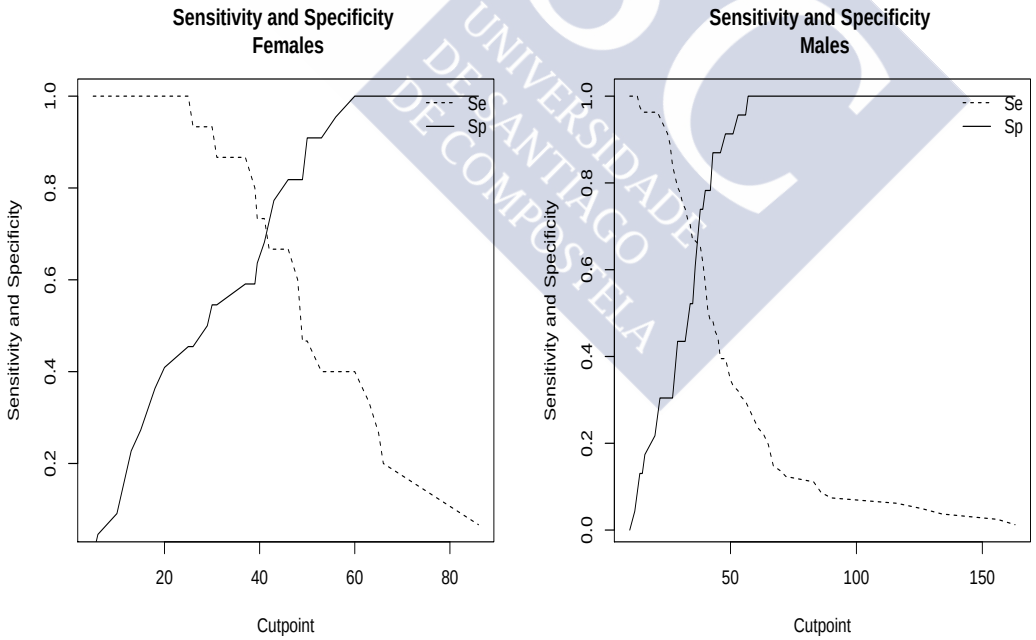


Figure 5.3: Sensitivity (*dotted line*) and specificity (*solid line*) plots according to cutpoints for females (*left panel*) and males (*right panel*).

5.2 The GsymPoint Package

In this section we present ***GsymPoint***, a package written in `R` for estimating the Generalized Symmetry point that is freely available from the Comprehensive `R` Archive Network (CRAN) at `http://CRAN.R-project.org/package=GsymPoint`. This package enables end-users to obtain the point estimates and the $(1-\alpha)\%$ confidence intervals (with α the signification level) using the recent methodologies introduced in ROC curves (Molanes-López and Letón, 2011; Lai et al., 2012) for the Generalized Symmetry point and its corresponding sensitivity and specificity indexes, taking into account the misclassification costs. Specifically, the two estimating methods presented in the previous chapter, the Generalized Pivotal Quantity method (Weerahandi, 1993, 1995; Lai et al., 2012) and the Empirical Likelihood method (Thomas and Grunkemeier, 1975; Molanes-López and Letón, 2011) have been incorporated in a clear and user-friendly way for the end-user Similarly to the *OptimalCutpoints* package (López-Ratón and Rodríguez-Álvarez, 2014; López-Ratón et al., 2014), the estimation of the optimal value can be computed straightforwardly by the levels of given (categorical) covariates since the discrimination capacity of a marker may depend on certain characteristics, such as the gender or age of the patient or the severity of disease (Pepe, 2004); and the *GsymPoint* package only requires a data-entry file where each column indicates a variable and each row indicates a patient/case. This dataset must, at least, contain the variable that indicates the diagnostic marker, the variable that indicates the true disease status (diseased/healthy) and if the optimal value is computed according to a (categorical) covariate, a variable that indicates the levels of such covariate. The numerical output of *GsymPoint* package includes the Generalized Symmetry point and its corresponding sensitivity and specificity accuracy measures with their associated $(1-\alpha)\%$ confidence intervals. In basis on the graphical interpretation of the Generalized Symmetry point, the graphical output shows the intersection point between the ROC curve and the line y = 1- ρ x, that is, the operating point in the ROC curve corresponding to the Generalized Symmetry point. Table 5.4 provides a summary of the most important functions included in the package.

5.2.1 `gsym.point()` function

The main function of the package is the `gsym.point` function, which uses the selected method(s) (GPQ and/or EL) to obtain (parametric and/or nonparametric) confidence intervals and point estimates for the Generalized Symmetry point and its corresponding sensitivity and specificity accuracy measures, and creates a class `gsym.point` object.

The code to use the `gsym.point` function is as follows:

```
gsym.point (methods, data, marker, status, tag.healthy,
categorical.cov = NULL, CFN = 1, CFP = 1, control =
control.gsym.point(), confidence.level = 0.95,
trace = FALSE, seed = FALSE, value.seed = 3)
```

The `methods` argument is a character vector selecting the method(s) to be used for estimating the Generalized Symmetry point and its corresponding accuracy measures.

Function	Description
`gsym.point`	Computes the Generalized Symmetry point and its sensitivity and specificity accuracy measures with their corresponding confidence intervals.
`control.gsym.point`	Function used to set several parameters that control the optimal-cutpoint estimating process.
`print`	Print method for objects fitted with `gsym.point`.
`summary`	Produces a summary of a `gsym.point` object.
`plot`	Plot method for objects fitted with `gsym.point`. Includes the plot of the ROC and the line y=1-ρ x. The intersection point between them is the operating point associated to the Generalized Symmetry point.

Table 5.4: Summary of functions in the *GsymPoint* package.

The possible options are: `"GPQ"`, `"EL"`, `c("GPQ","EL")` or `c("EL","GPQ")`.

The `data` argument is the data frame containing all the needed variables: the diagnostic marker, the true disease status and when it is neccesary, the categorical covariate; `marker` and `status` arguments are character strings with the names of the diagnostic test variable and the variable that distinguishes healthy from diseased individuals, respectively. The value codifying healthy individuals in this last variable `status` is indicated in the `tag.healthy` argument.

The `categorical.cov` argument is a character string with the name of the categorical covariate according to which the Generalized Symmetry point is automatically computed for each of its levels. By default it is `NULL`, that is, no categorical covariate is considered in the analysis.

The `CFN` and `CFP` arguments are the misclassification costs of false negative and false positive classifications, respectively. The default value is 1 for both.

The `control` argument indicates the output of the `control.gsym.point` function, which controls the whole optimal-cutpoint calculation process. This function will be explained in detail in the following subsection.

The `confidence.level` argument is the numerical value of the confidence level $1-\alpha$ for the construction of the confidence intervals. By default it is equal to 0.95.

The `trace` argument is a logical value that shows information on the progress if `TRUE`. By default it is `FALSE`.

The `seed` argument is a logical value, such that if `TRUE`, a seed is fixed for generating the trials when computing the confidence intervals. The default value is `FALSE`.

The `value.seed` argument is the numerical value for the fixed seed if `seed` is `TRUE`, and the default value is equal to 3.

Some of these arguments, `methods`, `data`, `marker`, `status` and `tag.healthy`, are essential and therefore they must be specified in the call to the `gsym.point` function. The other arguments, `categorical.cov`, `control`, `conf.level` and `trace`, are optional and therefore, if they are not specified, the default values will be considered.

5.2.2 `control.gsym.point()` function

It should be noted that there are some extra arguments, specific to each estimation method. We considered to include all of them in the `control` argument, which is a list of control values specified by calling to the `control.gsym.point` function, designed to replace the default values yielded by the `control.gsym.point` function. The arguments of the `control.gsym.point` function are presented in Table 5.5.

5.2.3 `summary.gsym.point()` function

Numerical results are printed on the screen, and the output yielded by the `summary.gsym.point` function or the `summary` method always includes:

- The matched call to the main function `gsym.point`
- The area under the ROC curve (AUC) estimated value.
- Information relating to the Generalized Symmetry point as the optimal cutpoint:
  - The method(s) used for estimating the optimal value (`EL` or `GPQ` method)
  - The Generalized Symmetry point, that is, the optimal cutoff and its accuracy-measure estimates Se and Sp, with their confidence intervals.

  All this information will be shown for each level of categorical covariate, if specified.

The call to this function is as follows:

```
R> summary(object, ...)
```

Argument	Description
`B`	The number of simulations in the Empirical Likelihood ("EL") method. The default value is 499.
$c_{sampling}$	The constant needed for resampling in the Empirical Likelihood ("EL") method. The default value is 0.25.
c_F	The constant needed for estimating the distribution in the Empirical Likelihood ("EL") method. The default value is 0.25.
c_{ELq}	The constant needed for estimating the empirical likelihood function in the Empirical Likelihood ("EL") method. The default value is 0.25.
c_R	The constant needed for estimating the ROC curve in the Empirical Likelihood ("EL") method. The default value is 0.25.
`I`	The number of replicates in the Generalized Pivotal Quantity ("GPQ") method. The default value is 2500.

Table 5.5: Summary of arguments of the `control.gsym.point` function.

The `object` argument is an object of class `gsym.point` as produced by `gsym.point` function.

`...` indicates further arguments passed to or from other methods. None are used in this method.

5.2.4 `plot.gsym.point()` function

The graphical output of the *GsymPoint* package is yielded by the `plot.gsym.point` function or by the `plot` method. This function plots the ROC curve and the line y = 1-ρx. Usage is as follows:

```
R> plot(x, legend = TRUE,...)
```

The `x` argument is an object of class `gsym.point` as produced by the `gsym.point` function.

The argument `legend` is a logical value for including the legend of the AUC value with its confidence interval when it is `TRUE` (value by default).

`...` indicates further arguments passed to the method `plot.default`.

5.2.5 Biomedical applications

This section illustrates the use of the `R`-based *GsymPoint* package by means of three biomedical datasets on melanoma, prostate cancer, and coronary artery disease, which have already been analyzed in previous chapters.

Melanoma dataset

We have analyzed in the previous chapter a dataset on 72 patients with suspicious lesions of being a melanoma, considering the CSS as the diagnostic marker for discriminating if a suspicious pigmented lesion on the skin is (or not) a melanoma (Venkatraman and Begg, 1996). In this case, a biopsy detected 21 melanomas. So, the main objective here is to illustrate the use of the *GsymPoint* package for estimating the Generalized Symmetry point as the optimal cutpoint for CSS to diagnose melanoma lesions on the skin. The first step consists on downloading the *GsymPoint* package and the melanoma dataset in `R`:

```
R> library(GsymPoint)

R> data(melanoma)
```

To view summary statistics of the variables included in the data set:

```
R> summary(melanoma)

       X                 group
 Min.   :-5.88100   Min.   :0.0000
 1st Qu.:-3.22100   1st Qu.:0.0000
 Median :-1.69550   Median :0.0000
 Mean   :-1.55642   Mean   :0.2917
 3rd Qu.: 0.00675   3rd Qu.:1.0000
 Max.   : 3.03200   Max.   :1.0000
```

To estimate the Generalized Symmetry point using the melanoma dataset, we have to use the commands shown below. For the misclassification costs we have considered firstly $CFN = 2$ and $CFP = 1$, taking into account therefore that a false negative decision is 2 times more serious than a false positive decision. Since the CSS values can be considered normally distributed in both melanoma and non-melanoma groups, according to the Shapiro-Wilk normality test, as we pointed out in the previous chapter, the GPQ method is the more adequate in this case for estimating the Generalized Symmetry point.

```
R> cutpoint1 <- gsym.point(methods = "GPQ", data = melanoma,
+ marker = "X", status = "group", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 2, CFP = 1, control =
+ control.gsym.point(),confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)
```

The `cutpoint1` object is a list that consists of the following components: `"GPQ"`; `"methods"`; `"call"`; and `"data"`:

```
R> names(cutpoint1)

"GPQ"     "methods"  "call"    "data"
```

The component `"methods"` is a character vector with the value of the argument `methods` used in the call; `"call"` is the matched call; and finally, `"data"` is the data frame used in the analysis (in this case the melanoma dataset). The first component (`"GPQ"`) contains the results associated with the GPQ method for estimating the Generalized Symmetry point:

```
R> names(cutpoint1$GPQ)
[1] "Global"

R> names(cutpoint1$GPQ$Global)
[1] "optimal.result"  "AUC"    "rho"
[4] "pvalue.healthy"  "pvalue.diseased"
```

1) `"optimal.result"` is a list with the Generalized Symmetry point and its associated sensitivity and specificity indexes with the corresponding $(1-\alpha)\%$ confidence intervals,

2) `"AUC"`: is a list with the numerical value of the area under the ROC curve;

3) `"rho"`: is the numerical value of the costs ratio $\rho = \frac{C_{FP}}{C_{FN}}$;

4)`"pvalue.healthy"`: is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in healthy population; and

5) `"pvalue.diseased"`: is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in diseased population.

```
R>cutpoint1$GPQ

$Global
$Global$optimal.result
$Global$optimal.result$cutoff
```

```
      Value        ll         ul
1 -1.213237 -1.7929079 -0.6283236

$Global$optimal.result$Specificity
    Value        ll        ul
1 0.75465 0.6249716 0.8485824

$Global$optimal.result$Sensitivity
     Value        ll        ul
1 0.877325 0.8124858 0.9242912

$Global$AUC
$Global$AUC$AUC
[1] 0.9056956

$Global$rho
[1] 0.5

$Global$pvalue.healthy
[1] 0.4719117

$Global$pvalue.diseased
[1] 0.9084176
```

If we do not visualizate all this information, the end-user can directly access each of these components, for example, to see only the value of the AUC:

```
R> cutpoint1$GPQ$Global$AUC

$AUC
[1] 0.9056956
```

A numerical summary of the results can be obtained by calling up the `print.gsym.point` or `summary.gsym.point` functions, which can be abbreviated by the `print` and `summary` methods, respectively:

```
R> summary(cutpoint1)

**********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
**********************************************

Call:
gsym.point(methods = "GPQ", data = melanoma, marker = "X",
status = "group", tag.healthy = 0, categorical.cov = NULL,
CFN = 2, CFP = 1, control = control.gsym.point(),
confidence.level = 0.95, trace = TRUE, seed = TRUE, value.seed = 3)
```

```
According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

Area under the ROC curve (AUC):  0.906

METHOD: GPQ

             Estimate 95% CI lower limit 95% CI upper limit
cutoff      -1.213237         -1.7929079         -0.6283236
Specificity  0.754650          0.6249716          0.8485824
Sensitivity  0.877325          0.8124858          0.9242912
```

As the parameter `seed` in the above call is `seed` = TRUE, if the user needs to replicate this example, he (or she) just need to run again the same call to obtain the same output, that is, the same confidence intervals and point estimates previously obtained for the Generalized Symmetry point and its corresponding sensitivity and specificity indexes. In this case, the output provided by the `summary.gsym.point` function shows:

1) an informative message indicating that the marker can be considered normally distributed in both groups, according to the Shapiro-Wilk normality test,

2) the Shapiro-Wilk test p-values indicating normality in both groups,

3) the AUC value and information relating to the Generalized Symmetry point, that is, the point estimates and the GPQ based $(1-\alpha)\%$-confidence intervals for the Generalized Symmetry point and its corresponding sensitivity and specificity indexes. By default, confidence intervals for the AUC and accuracy indexes are computed for a default confidence level, $1-\alpha$, equal to 0.95, although this value can be changed in the `confidence.level` argument of the main `gsym.point` function. Moreover, if the user wants to change some of the specific parameters of each method, as for instance, the number of replicates in the "GPQ" method, this value can be changed by means of the `I` argument, which appear in the `control.gsym.point` function. For example, if the number of replicates has to be equal to 1500, the user would only have to indicate this value using the following syntax:

```
R> cutpoint2 <- gsym.point(methods = "GPQ", data = melanoma,
+ marker = "X", status = "group", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 2, CFP = 1, control =
+ control.gsym.point(I = 1500), confidence.level = 0.95,
+ trace = FALSE, seed = TRUE, value.seed = 3)

R> summary(cutpoint2)
```

In this case, the results would be as follows:

```
***********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
***********************************************

Call:
gsym.point(methods = "GPQ", data = melanoma, marker = "X",
    status = "group", tag.healthy = 0, categorical.cov = NULL,
    CFN = 2, CFP = 1, control = control.gsym.point(I = 1500),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

Area under the ROC curve (AUC):  0.906

METHOD: GPQ

             Estimate 95% CI lower limit 95% CI upper limit
cutoff      -1.213237          -1.7973143         -0.6206575
Specificity  0.754650           0.6263304          0.8504338
Sensitivity  0.877325           0.8131652          0.9252169
```

If we wanted to change the method for estimating the Generalized Symmetry point and to consider the Empirical Likelihood method rather than the Generalized Pivotal Quantity method, since the diagnostic marker in this case can be assumed normally distributed in both groups, the program would show an informative message indicating that the Generalized Pivotal Quantity method is more adequate in this case:

```
R> cutpoint3 <- gsym.point(methods = "EL", data = melanoma,
+ marker = "X", status = "group", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 2, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace =
+ FALSE, seed = TRUE, value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.
```

```
Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084
```

In this case, the results would be the following:

```
R> summary(cutpoint3)

***********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
***********************************************

Call:
gsym.point(methods = "EL", data = melanoma, marker = "X",
    status = "group", tag.healthy = 0, categorical.cov = NULL,
    CFN = 2, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

Area under the ROC curve (AUC):  0.906

METHOD: EL

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      -1.2382325         -1.8403497         -0.4565671
Specificity  0.7901833          0.6326184          0.8973174
Sensitivity  0.8950916          0.8163092          0.9486587
```

And the same behaviour if we consider both estimating methods GPQ and EL simultaneously:

```
R> cutpoint4 <- gsym.point(methods = c("EL","GPQ"), data =
+ melanoma, marker = "X", status = "group", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 2, CFP = 1,
+ control = control.gsym.point(), confidence.level = 0.95,
+ trace = FALSE, seed = TRUE, value.seed = 3)
```

```
According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

R> summary(cutpoint4)

************************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
************************************************

Call:
gsym.point(methods = c("EL", "GPQ"), data = melanoma, marker = "X",
    status = "group", tag.healthy = 0, categorical.cov = NULL,
    CFN = 2, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = FALSE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

Area under the ROC curve (AUC):  0.906

METHOD: EL

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      -1.2382325         -1.8403497         -0.4565671
Specificity  0.7901833          0.6326184          0.8973174
Sensitivity  0.8950916          0.8163092          0.9486587

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      -1.213237         -1.8118902         -0.6817740
```

```
Specificity  0.754650          0.6302884          0.8484081
Sensitivity  0.877325          0.8151442          0.9242040
```

In case there is a successful treatment for treating the diseased individuals, other misclassification costs would be more adequate. In fact, this situation would be the opposite to that shown previously, that is, a FP decision would be more serious than a FN classification. If we assume, for instance, that the appropriate misclassification costs are $CFN = 1$ and $CFP = 2$, we have to change these values directly in the call to the main function `gsym.point`:

```
R> cutpoint5 <- gsym.point(methods = "GPQ", data = melanoma,
+ marker = "X", status = "group", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 1, CFP = 2, control =
+ control.gsym.point(), confidence.level = 0.95,
+ trace = FALSE, seed = TRUE, value.seed = 3)

R> summary(cutpoint5)

***********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
***********************************************

Call:
gsym.point(methods = "GPQ", data = melanoma, marker = "X",
    status = "group", tag.healthy = 0, categorical.cov = NULL,
    CFN = 1, CFP = 2, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.4719  0.9084

Area under the ROC curve (AUC):  0.906

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      -0.4341588         -0.9050401         0.07453657
Specificity  0.8711315          0.8085293         0.91787228
Sensitivity  0.7422630          0.6170586         0.83574457
```

The graphical output can be obtained by calling up the `plot.gsym.point` function, which can be abbreviated by the `plot` method:

```
R> plot(cutpoint1)
```

The graphical output shows the plot of the Receiver Operating Characteristic (ROC) curve and the line y = 1-$\rho\, x$, where the intersection between them indicates the operating point corresponding to the Generalized Symmetry point of the CSS marker for discriminating between patients with and without melanoma. Figure 5.4 shows the plot that appears as a result of the above call. This is the default output but the end-user can add specific graphic parameters, such as color, legend, etc.

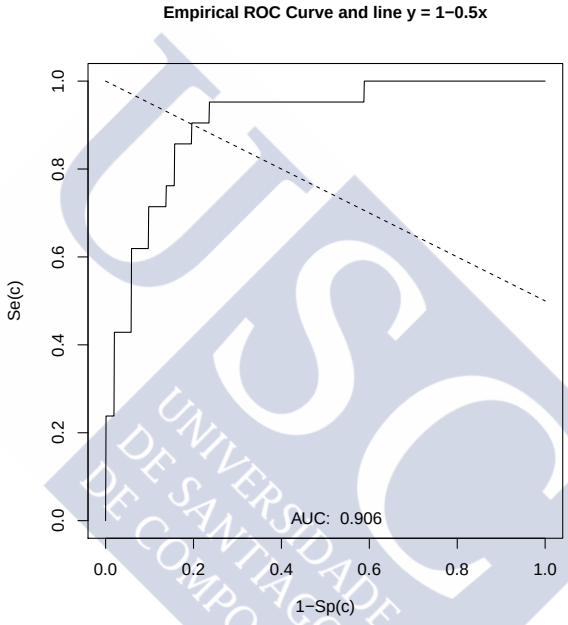


Figure 5.4: Graphical output of GsymPoint package for melanoma dataset

Prostate cancer dataset

We have considered here again the dataset on prostate cancer analyzed by Le (2006). We remind that in this data, 20 patients (of the total of 55 patients) has nodal involvement, and the level of acid phosphatase in blood serum (APBS) ($\times 100$) is considered as the (continuous) diagnostic marker for predicting nodal involvement. We illustrate below how to apply the *GsymPoint* package for estimating the Generalized Symmetry point of the APBS marker that will be used for discriminating between individuals with and without nodal involvement. To do so, we first have to download the package and the corresponding prostate dataset in `R`:

```
R> library(GsymPoint)
```

```
R> data(prostate)
```

To view summary statistics of the variables included in this data set:

```
R> summary(prostate)

    marker           status
 Min.   : 40.00   Min.   :0.0000
 1st Qu.: 50.00   1st Qu.:0.0000
 Median : 65.00   Median :0.0000
 Mean   : 69.42   Mean   :0.3774
 3rd Qu.: 78.00   3rd Qu.:1.0000
 Max.   :187.00   Max.   :1.0000
```

After loading the package and the prostate cancer data, for estimating the Generalized Symmetry point of the APBS marker by means of the `GsymPoint` package, we will use the following syntax:

```
R> cutpoint1 <- gsym.point (methods = "GPQ", data = prostate,
+ marker = "marker", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)
```

In this biomedical example, a FN result is much more harmful than a FP result and therefore we have considered that specifically a false negative classification is exactly 10 times more serious than a false positive classification ($CFN = 10$, $CFP = 1$) because cancer is a very serious disease which can cause death. Moreover, since the Shapiro-Wilk normality test indicated that both groups could be assumed normally distributed (after a Box-Cox transformation of the data), we have selected the GPQ method as the more adequate for computing the point estimates and 95% confidence intervals for the Generalized Symmetry point of APBS and its associated sensitivity and sepecificity indexes.

The numerical results will be obtained by means of the `print.gsym.point` or `summary.gsym.point` functions, which can be abbreviated by `print` and `summary` methods, respectively:

```
R> summary(cutpoint1)

**********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
**********************************************

Call:
gsym.point(methods = "GPQ", data = prostate, marker = "marker",
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)
```

```
According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                            Group 0 Group 1
Original marker              0.0000  0.0232
Box-Cox transformed marker   0.3641  0.2118

Area under the ROC curve (AUC):  0.725

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      51.9522523          46.6348952          57.1269332
Specificity  0.3233012           0.1411610           0.5140256
Sensitivity  0.9323301           0.9141161           0.9514026
```

In this case, the numerical output obtained with the `summary.gsym.point` function shows:

1) an informative message indicating that the original data can not be assumed normally distributed in both groups, but the Box-Cox transformed data can be considered normally distributed in both groups, according to the Shapiro-Wilk normality test;
2) the estimated value of the Box-Cox power lambda;
3) the estimated value of the area under the ROC curve (AUC), and
4) information corresponding to the Generalized Symmetry point: the point estimates and the GPQ based $(1-\alpha)\%$ confidence intervals (where α is the signification level) for the Generalized Symmetry point (the optimal value of APBS for discriminating patients with prostate cancer) and its corresponding sensitivity and specificity measures. By default, confidence intervals for the AUC and accuracy measures are computed for a confidence level $1-\alpha$ of 0.95, although this value can be changed in the `confidence.level` argument of the `main gsym.point` function.

The components of the `cutpoint1` object (which is a list) are: `"GPQ"`; `"methods"`; `"call"`; and `"data"`:

```
R> names(cutpoint1)
"GPQ"     "methods" "call"    "data"
```

The first component (`"GPQ"`) contains the results associated with the GPQ method for estimating the Generalized Symmetry point:

```
R> names(cutpoint1$GPQ)
[1] "Global"

R> names(cutpoint1$GPQ$Global)

[1] "optimal.result"                "AUC"
[3] "rho"                           "lambda"
[5] "normality.transformed"         "pvalue.healthy"
[7] "pvalue.diseased"               "pvalue.healthy.transformed"
[9] "pvalue.diseased.transformed"
```

1) "optimal.result" is a list with the Generalized Symmetry point and its sensitivity and specificity accuracy measures with the corresponding $(1-\alpha)\%$ confidence intervals:

```
R> cutpoint1$GPQ$Global$optimal.result

$cutoff
     Value      ll       ul
1 51.95225 46.6349 57.12693

$Specificity
      Value       ll        ul
1 0.3233012 0.141161 0.5140256

$Sensitivity
      Value        ll       ul
1 0.9323301 0.9141161 0.9514026
```

2) "AUC": is a list with the numerical value of the area under the ROC curve:

```
R> cutpoint1$GPQ$Global$AUC

$AUC
[1] 0.725
```

3) "rho" is the numerical value of $\rho = C_{FP}/C_{FN}$ (in this case $\rho = 0.1$):

```
R> cutpoint1$GPQ$Global$rho
[1] 0.1
```

4) "lambda": is the numerical value of the power used in the Box-Cox transformation of the GPQ method:

```
R> cutpoint1$GPQ$Global$lambda
[1] -1.249428
```

5) "normality.transformed": is a character string indicating if the transformed marker values by the Box-Cox transformation are normally distributed ("yes") or not ("no"):

```
R> cutpoint1$GPQ$Global$normality.transformed
[1] "yes"
```

6) "pvalue.healthy": is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the healthy population:

```
R> cutpoint1$GPQ$Global$pvalue.healthy
[1] 3.276498e-07
```

7) "pvalue.diseased": is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the diseased population:

```
R> cutpoint1$GPQ$Global$pvalue.diseased
[1] 0.02323895
```

8) "pvalue.healthy.transformed": is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the healthy population:

```
R> cutpoint1$GPQ$Global$pvalue.healthy.transformed
[1] 0.3640662
```

9) "pvalue.diseased.transformed": is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the diseased population:

```
R> cutpoint1$GPQ$Global$pvalue.diseased.transformed
[1] 0.2118137
```

The component "methods" is a character vector with the value of the argument methods used in the call; "call" is the matched call; and finally, "data" is the data frame used in the analysis (in this case the data set on prostate cancer):

```
R> cutpoint1$methods
[1] "GPQ"

R> cutpoint1$call
gsym.point(methods = "GPQ", data = prostate, marker = "marker",
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

R> cutpoint1$data
   marker status
```

```
1       40       0
2       40       0
3       46       0

[...]

51      99       1
52     126       1
53     136       1
```

We remind that in the above command for computing the value `cutpoint1`, the number of the replicates used in the GPQ method was $I = 2500$ (the value established by default). If we want to change such value and to select, for instance, a value of I equal to 2000, this can be changed in the `control` argument as follows:

```
R> cutpoint2 <- gsym.point (methods = "GPQ", data = prostate,
+ marker = "marker", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
+ control.gsym.point(I = 2000), confidence.level = 0.95,
+ trace = FALSE, seed = TRUE, value.seed = 3)
```

So, if we use 2000 replicates for estimating the Generalized Symmetry point, the results would be the following:

```
R> summary(cutpoint2)

**********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
**********************************************

Call:
gsym.point(methods = "GPQ", data = prostate, marker = "marker",
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(I = 2000),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                           Group 0 Group 1
```

```
Original marker             0.0000  0.0232
Box-Cox transformed marker  0.3641  0.2118

Area under the ROC curve (AUC):  0.725

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      51.9522523         46.8233675         57.2607208
Specificity  0.3233012          0.1411278          0.5190622
Sensitivity  0.9323301          0.9141128          0.9519062
```

If we consider the EL method instead of the GPQ method for estimating the Generalized Symmetry point and its accuracy measures, an informative message is shown by the package, advising the user that the GPQ method would be more suitable in this case, due to the normality of the Box-Cox transformed marker in both groups:

```
R> cutpoint3 <- gsym.point (methods = "EL", data = prostate,
+ marker = "marker", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                           Group 0 Group 1
Original marker             0.0000  0.0232
Box-Cox transformed marker  0.3641  0.2118
```

The results for EL method would be the following, where the values of the optimal cutpoint and specificity are lower than for GPQ method:

```
R> summary(cutpoint3)

**********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
**********************************************
```

```
Call:
gsym.point(methods = "EL", data = prostate, marker = "marker",
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                          Group 0 Group 1
Original marker            0.0000  0.0232
Box-Cox transformed marker 0.3641  0.2118

Area under the ROC curve (AUC):  0.725

METHOD: EL

             Estimate 95% CI lower limit 95% CI upper limit
cutoff      49.2249839         45.4055971         59.9753792
Specificity  0.2451690          0.0927334          0.5376739
Sensitivity  0.9245169          0.9092733          0.9537674
```

And similarly if we consider both the GPQ and EL methods simultaneously. The program shows a message indicating that the results will be computed and presented by the two methods, although the GPQ method would be more adequate in this situation of normality:

```
R> cutpoint4 <- gsym.point(methods = c("GPQ","EL"), data = prostate,
+ marker = "marker", status = "status", tag.healthy = 0,
+ CFN = 10, CFP = 1, control = control.gsym.point(),
+ confidence.level = 0.95, trace = FALSE, seed = TRUE, value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.
```

```
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                           Group 0 Group 1
Original marker             0.0000  0.0232
Box-Cox transformed marker  0.3641  0.2118

R> summary(cutpoint4)

************************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
************************************************

Call:
gsym.point(methods = c("GPQ", "EL"), data = prostate,
    marker = "marker", status = "status", tag.healthy = 0,
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
    value.seed = 3)

According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -1.2494

Shapiro-Wilk test p-values
                           Group 0 Group 1
Original marker             0.0000  0.0232
Box-Cox transformed marker  0.3641  0.2118

Area under the ROC curve (AUC):  0.725

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      51.9522523         46.6348952         57.1269332
Specificity  0.3233012          0.1411610          0.5140256
```

```
Sensitivity  0.9323301          0.9141161          0.9514026

METHOD: EL

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      49.2249839        45.38400377         57.6847668
Specificity  0.2451690         0.07525967          0.4999079
Sensitivity  0.9245169         0.90752597          0.9499908
```

The graphical results can be obtained by the `plot.gsym.point` function, which can be abbreviated by the `plot` method:

```
R> plot(cutpoint1)
```

With the above command, the ROC curve and the line y = 1-ρ x are represented by default, and the operating point associated to the Generalized Symmetry point of the APBS marker for discriminating patients with prostate cancer is indicated by the intersection between the ROC curve and that line (see Figure 5.5). In addition, in this figure the user can change several characteristics as legend, color, ..., specifying/adding the corresponding graphical arguments into the syntax below.

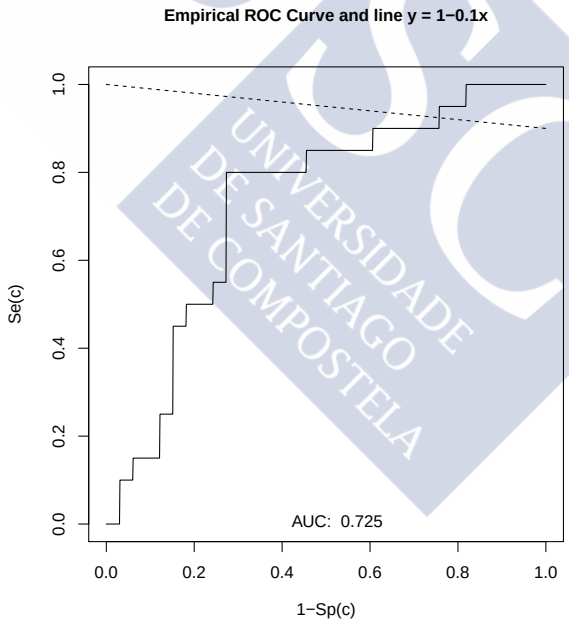


Figure 5.5: Graphical output of GsymPoint package for prostate dataset

Coronary artery disease dataset

In this subsection we consider the galician study conducted on 141 patients (96 with coronary lesions and 45 with non-stenotic coronaries) admitted to the Cardiology Department of a Teaching Hospital in Galicia (northwest Spain) for evaluating chest pain or

cardiovascular disease, where the leukocyte elastase determination was investigated as a potential clinical marker for the diagnosis of coronary artery disease (Amaro et al., 1995). Our proposal in this case is to illustrate the practical application of *GsymPoint* package to these data, that is, to compute the optimal value (by means of the Generalized Symmetry point) of elastase concentration to diagnose patients with coronary artery disease (CAD). From here on, we will refer to this dataset as elastase.

First of all, we need to download the *GsymPoint* package and the corresponding elastase dataset in R:

```
R> library(GsymPoint)

R> data(elastase)
```

We can see the nature (continuous or categorical) and summary statistics of the variables of this dataset:

```
R> summary(elastase)
      elas            status          gender
 Min.   :  5.00   Min.   :0.0000   Female: 37
 1st Qu.: 27.00   1st Qu.:0.0000   Male  :104
 Median : 39.00   Median :1.0000
 Mean   : 43.28   Mean   :0.6809
 3rd Qu.: 51.00   3rd Qu.:1.0000
 Max.   :163.00   Max.   :1.0000
```

Then, to compute the Generalized Symmetry point in the elastase dataset, we simply have to use the syntax indicated below. In this case we are interested in having a high sensitivity (elastase as a screening test), and given that searching for a high sensitivity is the same as searching for a low number of false negatives, the same values as in the previous prostate cancer dataset are considered for the misclassification costs $CFN = 10$ and $CFP = 1$ (a false negative result is regarded as more serious than a false positive one). In addition, since these data do not follow the Box-Cox family in both CAD and non-CAD groups according to the Shapiro-Wilk normality test, we use the Empirical Likelihood method, that does not require the normality assumption, rather than the Generalized Pivotal Quantity method, for computing the Generalized Symmetry point and its accuracy measures Se and Sp :

```
R> cutpoint1 <- gsym.point (methods = "EL", data = elastase,
+ marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace =
+ FALSE, seed = TRUE, value.seed = 3)
```

The object obtained by means of the execution of the above command (`cutpoint1` object) is a list with the following components:

```
R> names(cutpoint1)
[1] "EL"      "methods" "call"    "data"
```

The component "methods" is a character vector with the value of the argument methods used in the call; "call" is the matched call; "data" is the data frame used in the analysis (in this case the elastase dataset). The first component ("EL") includes the results associated with the EL method used to compute the Generalized Symmetry point and it contains the following elements:

```
R> names(cutpoint1$EL)
[1] "Global"

R> names(cutpoint1$EL$Global)
[1] "optimal.result"                "AUC"
[3] "rho"                           "lambda"
[5] "normality.transformed"         "pvalue.healthy"
[7] "pvalue.diseased"               "pvalue.healthy.transformed"
[9] "pvalue.diseased.transformed"
```

where

"optimal.result" is a list with the Generalized Symmetry point and its sensitivity and specificity indexes with the corresponding (1-α)% confidence intervals for the elastase concentration, the potential diagnostic marker of CAD:

```
R> cutpoint1$EL$Global$optimal.result

$cutoff
         [,1]     [,2]    [,3]
[1,] 23.13564 17.36626 25.5516

$Specificity
          [,1]      [,2]      [,3]
[1,] 0.3735075 0.2181514 0.5111823

$Sensitivity
          [,1]      [,2]      [,3]
[1,] 0.9373507 0.9218151 0.9511182
```

"AUC" is a list with the numerical value of the area under the ROC curve

```
R> cutpoint1$EL$Global$AUC

$AUC
[1] 0.7436343
```

"rho" is the numerical value of the ratio $\rho = \frac{C_{FP}}{C_{FN}}$,

```
R> cutpoint1$EL$Global$rho
[1] 0.1
```

"lambda" is the estimated numerical value of the power in the Box-Cox transformation

```
R> cutpoint1$EL$Global$lambda
[1] 0.1136496
```

"normality.transformed" is a character string indicating if the transformed marker values by the Box-Cox transformation are normally distributed ("yes") or not ("no"):

```
R> cutpoint1$EL$Global$normality.transformed
[1] "no"
```

"pvalue.healthy" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the healthy population:

```
R> cutpoint1$EL$Global$pvalue.healthy
[1] 0.07463537
```

"pvalue.diseased" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the diseased population:

```
R> cutpoint1$EL$Global$pvalue.diseased
[1] 2.192607e-09
```

"pvalue.healthy.transformed" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the healthy population:

```
R> cutpoint1$EL$Global$pvalue.healthy.transformed
[1] 0.009062022
```

"pvalue.diseased.transformed" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the diseased population:

```
R> cutpoint1$EL$Global$pvalue.diseased.transformed
[1] 0.07933026
```

A summary of the numerical results can be obtained by calling up the `print.gsym.point` or `summary.gsym.point` functions, which can be abbreviated by `print` and `summary` methods, respectively:

```
R> summary(cutpoint1)

***********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
***********************************************

Call:
gsym.point(methods = "EL", data = elastase, marker = "elas",
```

```
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
value.seed = 3)

Area under the ROC curve (AUC):  0.744

METHOD: EL

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      23.1356401         17.3662628         25.5516003
Specificity  0.3735075          0.2181514          0.5111823
Sensitivity  0.9373507          0.9218151          0.9511182
```

Similarly to previous biomedical examples, in this case, the output provided by the `summary.gsym.point` function shows: firstly, the AUC value and, secondly, information relating to the Generalized Symmetry point, that is, the point estimates and (1-α)%-confidence intervals obtained by the EL method for the Generalized Symmetry point of the elastase concentration and its corresponding sensitivity and specificity measures in males and females. By default, confidence intervals for the AUC and accuracy measures are computed for a confidence level 1-α of 0.95, but this value can be changed directly in the `confidence.level` argument of the main `gsym.point` function. Moreover, if the user wants to change some of the specific parameters of the EL method, as for instance, the number of simulations executed in the Empirical Likelihood method, this value can be changed by means of the `control` argument of the main function `gsym.po int`, indicating the new value of `B` inside the `control.gsym.point` function. For example, if the user wants to set the number of simulations equal to 999, this can be simply indicated by using the following syntax:

```
R> cutpoint2 <- gsym.point(methods = "EL", data = elastase,
+ marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1,
+ control = control.gsym.point(B = 999),
+ confidence.level = 0.95, trace = FALSE, seed = TRUE,
+ value.seed = 3)

R> summary(cutpoint2)

**********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
**********************************************

Call:
gsym.point(methods = "EL", data = elastase, marker = "elas",
    status = "status", tag.healthy = 0, categorical.cov = NULL,
    CFN = 10, CFP = 1, control = control.gsym.point(B = 999),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
```

```
    value.seed = 3)

Area under the ROC curve (AUC):  0.744

METHOD: EL

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      23.1356401         17.4291826         25.7166528
Specificity  0.3735075          0.2234065          0.4887079
Sensitivity  0.9373507          0.9223407          0.9488708
```

For this dataset, if we select the GPQ method for computing the Generalized Symmetry point and its accuracy measures, the *GsymPoint* package shows a message indicating that in this case, since the data marker in both diseased and healthy populations are not normally distributed, the only valid method or the two implemented in the package is the EL method, which does not require the assumption of normality:

```
R> cutpoint3 <- gsym.point (methods = "GPQ", data = elastase,
+ marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)

According to the Shapiro-Wilk normality test, the original marker
can not be considered normally distributed in both groups.
After transforming the marker using the Box-Cox transformation
estimate the Shapiro-Wilk normality test indicates that the
transformed marker can not be considered normally distributed
in both groups.
Therefore, the results obtained with the GPQ method may not be
reliable. You must use the EL method instead.

Box-Cox lambda estimate = 0.1136

Shapiro-Wilk test p-values
                            Group 0 Group 1
Original marker              0.0746  0.0091
Box-Cox transformed marker   0.0000  0.0793
```

In a similar way, if we introduce both GPQ and EL methods simultaneously, the program shows a message indicating that the GPQ method is not valid and that may yield not reliable results in this case, and that consequently you must only use the results obtained by the EL method:

```
R> cutpoint4 <- gsym.point (methods = c("EL","GPQ"), data =
+ elastase, marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = NULL, CFN = 10, CFP = 1, control =
```

```
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)

According to the Shapiro-Wilk normality test, the original marker
can not be considered normally distributed in both groups.
After transforming the marker using the Box-Cox transformation
estimate the Shapiro-Wilk normality test indicates that the
transformed marker can not be considered normally distributed
in both groups.
Therefore, although the results of both methods will be shown,
the results obtained with the GPQ method may not be reliable.
You must use the EL method instead.

Box-Cox lambda estimate = 0.1136

Shapiro-Wilk test p-values
                           Group 0 Group 1
Original marker             0.0746  0.0091
Box-Cox transformed marker  0.0000  0.0793
```

As in the above biomedical applications, the graphical output of the results can be obtained with the `plot.gsym.point` function or with the `plot` method.

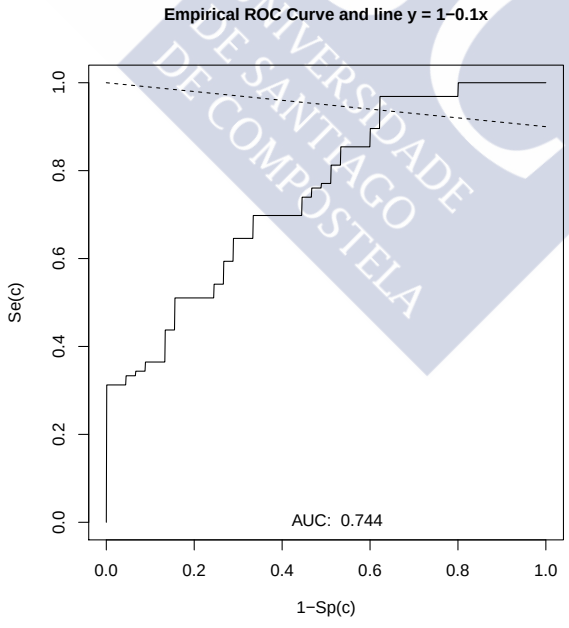


Figure 5.6: Graphical output of GsymPoint package for cardiology coronary dataset

In Figure 5.6 the ROC curve of the elastase concentrations, the potential marker of CAD,

is represented, together with the line y=1-ρ x. The intersection point between them is the operating point associated to the Generalized Symmetry point, that is, the optimal cutpoint for discriminating patients with CAD. This is the graphical output that appears by default as a result of the above call, but the end-user can change several graphical parameters, as the legends, colors, etc.

In addition, we now consider the categorical covariate gender (male or female) for evaluating if the optimal cutpoint is different between males and females. For computing the Generalized Symmetry point in the elastase data set taking into account this categorical covariate, we have to use the following syntax:

```
R> cutpoint1 <- gsym.point (methods = "GPQ", data = elas, marker =
+ "elas", status = "status", tag.healthy = 0, categorical.cov =
+ "gender", CFN = 10, CFP = 1, control = control.gsym.point(),
+ confidence.level = 0.95, trace = FALSE, seed = TRUE,
+ value.seed = 3)
```

Since these data in males and females follow the Box-Cox family in both CAD and non-CAD groups according to the Shapiro-Wilk normality test, the GPQ method would be more adequate under this situation of normality.

In this case, the `cutpoint1` object would be a list with the following components:

```
R> names(cutpoint1)

[1] "GPQ"         "methods"     "levels.cat" "call"
[5] "data"
```

where the component `"levels.cat"` is a character vector indicating the levels of the categorical covariate, that is, `"Male"` and `"Female"`. The rest of the components were already explained before, but in this case the first component `"GPQ"` is itself a two-component list (for `"Male"` and `"Female"`) containing:

```
R> names(cutpoint1$GPQ)
[1] "Female" "Male"

R> names(cutpoint1$GPQ$Male)

[1] "optimal.result"                "AUC"
[3] "rho"                           "lambda"
[5] "normality.transformed"         "pvalue.healthy"
[7] "pvalue.diseased"               "pvalue.healthy.transformed"
[9] "pvalue.diseased.transformed"
```

where
`"optimal.result"` is a list with the Generalized Symmetry point and its sensitivity and specificity indexes with the corresponding $(1-\alpha)\%$ confidence intervals (for males):

```
R> cutpoint1$GPQ$Male$optimal.result
```

```
$cutoff
     Value       ll      ul
1 20.72776 17.95095 23.2821

$Specificity
      Value        ll        ul
1 0.2739826 0.1320574 0.4206069

$Sensitivity
      Value        ll        ul
1 0.9273983 0.9132057 0.9420607
```

"AUC" is a list with the numerical value of the area under the ROC curve (for males)

```
R> cutpoint1$GPQ$Male$AUC
$AUC
[1] 0.7216855
```

"rho" is the numerical value of the ratio $\rho = \frac{C_{FP}}{C_{FN}}$,

```
R> cutpoint1$GPQ$Male$rho
[1] 0.1
```

"lambda" is the estimated numerical value of the power in the Box-Cox transformation obtained by considering only the males individuals:

```
R> cutpoint1$GPQ$Male$lambda
[1] -0.04277911
```

"normality.transformed" is a character string indicating if the transformed marker values by the Box-Cox transformation (for males) are normally distributed ("yes") or not ("no"):

```
R> cutpoint1$GPQ$Male$normality.transformed
[1] "yes"
```

"pvalue.healthy" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the healthy population of males:

```
R> cutpoint1$GPQ$Male$pvalue.healthy
[1]  0.5866506
```

"pvalue.diseased" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the marker in the diseased population of males:

```
R> cutpoint1$GPQ$Male$pvalue.diseased
[1] 5.44323e-09
```

"pvalue.healthy.transformed" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the healthy population of males:

```
R> cutpoint1$GPQ$Male$pvalue.healthy.transformed
[1] 0.06656483
```

"pvalue.diseased.transformed" is the numerical value of the p-value obtained by the Shapiro-Wilk normality test for checking the normality assumption of the Box-Cox transformed marker in the diseased population of males:

```
R> cutpoint1$GPQ$Male$pvalue.diseased.transformed
[1] 0.2147409
```

The numerical results obtained by the summary.gsym.point function or the summary method show the Generalized Symmetry point with its sensitivity and specificity indexes for each level of the categorical covariate, in this case, for males and females:

```
R> summary(cutpoint1)

***********************************************
OPTIMAL CUTOFF: GENERALIZED SYMMETRY POINT
***********************************************

Call:
gsym.point(methods = "GPQ", data = elastase, marker = "elas",
    status = "status", tag.healthy = 0, categorical.cov = "gender",
    CFN = 10, CFP = 1, control = control.gsym.point(),
    confidence.level = 0.95, trace = FALSE, seed = TRUE,
    value.seed = 3)

***********************************************
Female
***********************************************
According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.0837  0.9077

Area under the ROC curve (AUC):  0.818

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      25.0929510         12.3370641         34.0526540
```

```
Specificity  0.4246091          0.1460251          0.6618634
Sensitivity  0.9424609          0.9146025          0.9661863

************************************************
Male
************************************************
According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.

Box-Cox lambda estimate = -0.0428

Shapiro-Wilk test p-values
                          Group 0 Group 1
Original marker            0.5867  0.0000
Box-Cox transformed marker 0.0666  0.2147

Area under the ROC curve (AUC):  0.722

METHOD: GPQ

              Estimate 95% CI lower limit 95% CI upper limit
cutoff      20.7277556          17.9509453         23.2821020
Specificity  0.2739826           0.1320574          0.4206069
Sensitivity  0.9273983           0.9132057          0.9420607
```

If we use instead the EL method for computing the Generalized Symmetry point and its accuracy measures, the *GsymPoint* package shows a message indicating that, since the data marker in both diseased and healthy populations for males and females are normally distributed according to the Shapiro-Wilk test, the GPQ method would be more adequate in this case:

```
R> cutpoint2 <- gsym.point (methods = "EL", data = elastase,
+ marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = "gender", CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)

Female :
According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.

Shapiro-Wilk test p-values
```

```
                Group 0 Group 1
Original marker  0.0837  0.9077

Male :
According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
indicates that the transformed marker can be considered
normally distributed in both groups.
Therefore the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -0.0428

Shapiro-Wilk test p-values
                          Group 0 Group 1
Original marker            0.5867  0.0000
Box-Cox transformed marker 0.0666  0.2147
```

Similarly, if we introduce both GPQ and EL methods simultaneously, the *GsymPoint* package shows a message indicating that the results of the two methods will be shown, although the GPQ method would be more adequate in this case because of the normality assumption:

```
R> cutpoint3 <- gsym.point (methods = c("EL","GPQ"), data =
+  elastase, marker = "elas", status = "status", tag.healthy = 0,
+ categorical.cov = "gender", CFN = 10, CFP = 1, control =
+ control.gsym.point(), confidence.level = 0.95, trace = FALSE,
+ seed = TRUE, value.seed = 3)

Female :
According to the Shapiro-Wilk normality test, the marker can be
considered normally distributed in both groups.
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Shapiro-Wilk test p-values

                Group 0 Group 1
Original marker  0.0837  0.9077

Male :
According to the Shapiro-Wilk normality test, the marker can not
be considered normally distributed in both groups.
However, after transforming the marker using the Box-Cox
transformation estimate, the Shapiro-Wilk normality test
```

```
indicates that the transformed marker can be considered
normally distributed in both groups.
Therefore, although the results of both methods will be shown,
the GPQ method would be more suitable for this dataset.

Box-Cox lambda estimate = -0.0428

Shapiro-Wilk test p-values
                            Group 0 Group 1
Original marker              0.5867  0.0000
Box-Cox transformed marker   0.0666  0.2147
```

The graphical results are obtained by the `plot.gsym.point` function or with the `plot` method.

```
R> plot(cutpoint1)
```

In Figure 5.7 the ROC curve of the elastase concentrations is represented separately for males and females, together with the line $y = 1-\rho x$. The intersection point between them is the operating point associated to the Generalized Symmetry point separately obtained in males and females for discriminating patients with CAD. As a result of the above call, this is the graphical output that appears by default for males and females, respectively, but as usual the end-user can change several graphical parameters, as the legends, colors, etc.

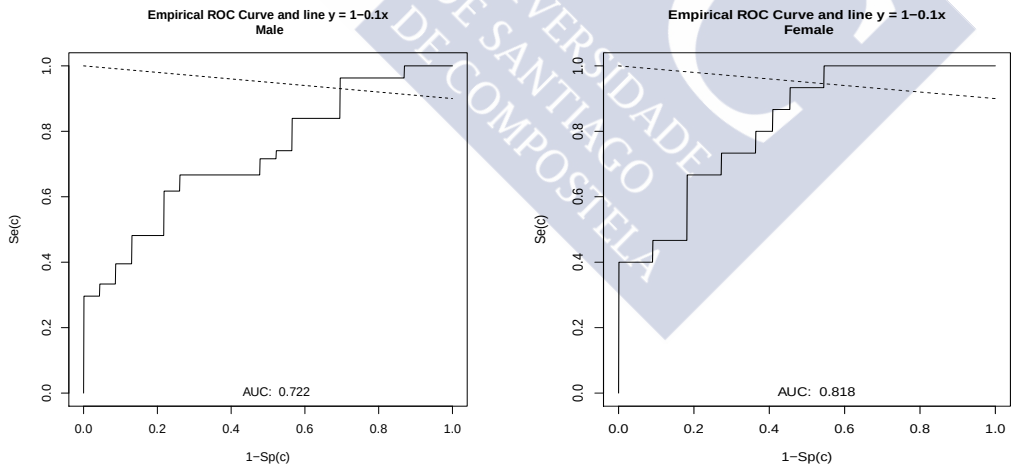


Figure 5.7: Graphical output of GsymPoint package for cardiology coronary dataset in males (*left panel*) and females (*right panel*).

Be gifted with wise flexibility.

- Angelica Hopes

Chapter 6

Improvement of the discriminatory capacity via generalized additive models

6.1 Introduction

In classical ROC analysis, marker levels above and below a given cut-off value result in individuals being labeled as diseased or nondiseased, respectively. However, in cases where the marker shows an irregular distribution, with a dominance of diseased subjects in non-contiguous regions, classification by reference to a cut-off value is neither feasible nor logical. Indeed, use of such an analysis would lead to erroneous conclusions, and a modification of the classification rule is therefore necessary (Lustres-Pérez et al., 2010). An intuitive solution to this problem would be to estimate for each individual the conditional probability of belonging to one of the status (e.g., diseased) given the observed value in the original scale of the marker, and consider these estimated probabilities as the observed values of a new continuous marker (the transformed marker) that will be used for classification purposes. In order to estimate such conditional probabilities we propose to use generalized additive models (GAMs) for binary data. GAMs are modern regression techniques that have the advantage of not assuming a parametric relationship between status and marker, and eliminate the need for the researcher to impose functional assumptions (Hastie and Tibshirani, 1990). As we have previously discussed, the selection of an optimal cut-off value is an important clinical task that will be affected by the classification rule adopted (Altman et al., 1994; Lausen and Schumacher, 1996; Mazumdar and Glassman, 2000; Klotsche et al., 2009). Therefore, in this chapter, we describe firstly a procedure for improving the discriminatory capacity of a continuous marker that consists on transforming the original marker by using generalized additive models for binary data, and by transforming back to the original scale the optimal cut-off value obtained in the transformed scale (López-Ratón et al., 2015d). As it will be clear soon, this procedure may yield more than one cut-off value in the original scale, and more specifically, an interval of values. Then, we carry out a simulation study to check the practical behaviour of this approach and finally, we illustrate its applicability using a real dataset.

6.2 Transformed marker through logistic GAM regression

Let Y be a continuous marker. According to Neyman and Pearson (1933) and McIntosh and Pepe (2002), the best Y-based classifier with a single cut-off is that based on the conditional probability of one of the status (e.g., diseased), given the values of Y. Let $\tilde{Y}$ denote such a conditional probability function:

$$\tilde{Y} \equiv f(Y) = \Pr(D|Y) \subset (0,1) \tag{6.1}$$

In practice, however, the function $f(.)$ of (6.1) is unknown, and it is required to estimate it. In this study, $f(.)$ is estimated by using the logistic GAM regression model given by:

$$\tilde{Y} \equiv f(Y) = \Pr(D|Y) = g^{-1}(\alpha + h(Y)) = \frac{exp(\alpha + h(Y))}{1 + exp(\alpha + h(Y))} \tag{6.2}$$

where $g(.)$ is the logit link function, and $h(.)$ is an unknown smooth function.

To date, several approaches have been proposed in the statistical literature for estimating (6.2), including methods based on penalized regression splines (Eilers and Marx, 1996; Wood, 2003) or Bayesian versions of them (Lang and Brezger, 2004). Alternatively, the local scoring algorithm with kernel-type smoothers can also be used (McCullagh and Nelder, 1989; Wand and Jones, 1995). In this study, we use penalized regression with B(asic)-splines as smoothers (Eilers and Marx, 1996), for estimating the function $h(.)$. When estimating $h(.)$, a crucial step is the selection of the smoothing parameter that controls the smoothness of the resulting estimate. In this study, the optimal smoothing parameter is chosen automatically by use of the Un-Biased Risk Estimator (UBRE) criterion (Wood, 2004).

Once model (6.2) has been fitted, the estimated probabilities are used as the new marker, and the corresponding ROC curve and AUC are obtained. It should be noted that the logistic GAM regression model is a suitable tool for estimating the odds ratio (OR) function, denoted by $OR(y, y_{ref})$, which generalizes the concept of OR per unit increase in the continuous marker in the sense that represents the ratio of the odds of an event (in this case, presence of disease) occurring when the individual has $Y = y$ to the odds of that event occurring when the individual has $Y = y_{ref}$, where y_{ref} is a value clinically accepted as reference value. Assuming model (6.2), the OR function, denoted by $OR(y, y_{ref})$, can be expressed as (Figueiras and Cadarso-Suárez, 2001; Zhao et al., 2006):

$$\begin{aligned}
OR(y, y_{ref}) &= \frac{\Pr(D|Y=y)/(1-\Pr(D|Y=y))}{\Pr(D|Y=y_{ref})/(1-\Pr(D|Y=y_{ref}))} \\
&= \frac{\frac{\exp(\alpha+h(y))}{1+\exp(\alpha+h(y))} / \frac{1}{1+\exp(\alpha+h(y))}}{\frac{\exp(\alpha+h(y_{ref}))}{1+\exp(\alpha+h(y_{ref}))} / \frac{1}{1+\exp(\alpha+h(y_{ref}))}} \\
&= \frac{\exp(\alpha + h(y))}{\exp(\alpha + h(y_{ref}))} \\
&= \exp(h(y) - h(y_{ref})),
\end{aligned}$$

that is,

$$OR(y, y_{ref}) = exp(h(y) - h(y_{ref})). \tag{6.3}$$

The simulation study and the real data application detailed in the following sections have been carried out using R (R Core Team, 2015). The logistic GAM regression model was fitted using the `gam` function of the `mgcv` package (Wood, 2006) and both the ROC curve and AUC were empirically estimated.

6.3 Simulation study

In this section, we carry out a simulation study in `R` (R Core Team, 2015) to compare the performance of the transformed marker previously introduced in the former section, compared to the classical ROC analysis based on the original or crude marker. In this simulation study, we focus on the mean, mean squared error (MSE), bias ($Bias$) and standardized bias ($SBias$) of the nonparametric point estimators proposed in Chapter 4 for the Generalized Symmetry point c_{GS} and its accuracy measures $p(c_{GS})$ and $q(c_{GS})$ based on the EL methodology.

In the following subsections, we describe first the theoretical scenarios considered in this simulation study and then we discuss the results obtained.

6.3.1 Scenarios

The proposed methodology, previously introduced in Section 6.2, has been applied to several controlled scenarios. Specifically, we have analized three different models in which the distribution of the marker Y among the healthy and diseased populations was assumed to be known, and assessed the improvement in the classificatory capacity of the transformed marker $\tilde{Y} = \Pr(D|Y)$ over that of the original marker. In all cases a 50% disease prevalence was assumed.

We give below the specific details of the three distributional models considered in this simulation study:

Model a). Normal distributions with the same dispersion in healthy and diseased subjects:

$$Y|\bar{D} \sim N(0, 0.5); \quad Y|D \sim N(0.5, 0.5) \quad \text{(see Figure 6.1)}.$$

Model b). Normal distributions with different dispersion in healthy and diseased subjects:

$$Y|\bar{D} \sim N(0, 0.5); \quad Y|D \sim N(0.3, 0.9) \quad \text{(see Figure 6.2)}.$$

Model c). Normal distribution in healthy subjects and mixture of normal distributions in diseased subjects:

$$Y|\bar{D} \sim N(1.5, 0.8); \quad Y|D \sim 0.5\ N(0, 0.7) + 0.5\ N(3, 0.6) \quad \text{(see Figure 6.3)}.$$

Additionally, we collect in Table 6.1 the AUC values for Y and $\widetilde{Y}$, and for three different optimality criteria (the criterion that sets an FPF = 0.20, the Youden index, and

the Symmetry point) we collect the corresponding optimal threshold values c_{Opt} for each scale (Y and $\widetilde{Y}$). Besides, for the transformed scale $\widetilde{Y}$, we also collect in Table 6.1 the value of c or the interval of values (c_1, c_2) obtained after transforming back to the original scale the corresponding optimal threshold value c_{Opt} on the transformed scale.

Table 6.1: Parameters associated to the theoretical models studied in Subsection 6.3.1.

			FPR = 0.2			Youden index			Sym point		
		AUC	c_{Opt}	c	$q(c_{Opt})$	c_{Opt}	c	YI_{Opt}	c_{Opt}	c	$q(c_{Opt})$
a)	$\widetilde{Y}$	0.76	0.58	0.42	0.56	0.50	0.25	0.39	0.50	0.25	0.69
	Y	0.76	0.42		0.56	0.25		0.39	0.25		0.69
		AUC	c_{Opt}	(c_1, c_2)	$q(c_{Opt})$	c_{Opt}	(c_1, c_2)	YI_{Opt}	c_{Opt}	(c_1, c_2)	$q(c_{Opt})$
b)	$\widetilde{Y}$	0.70	0.49	(-0.80, 0.53)	0.51	0.50	(-0.84, 0.57)	0.32	0.41	(-0.61, 0.34)	0.64
	Y	0.61	0.42		0.44	0.56		0.26	0.11		0.58
c)	$\widetilde{Y}$	0.86	0.52	(0.46, 2.51)	0.77	0.50	(0.48, 2.49)	0.57	0.49	(0.50, 2.48)	0.78
	Y	0.51	2.17		0.46	2.49		0.30	1.51		0.50

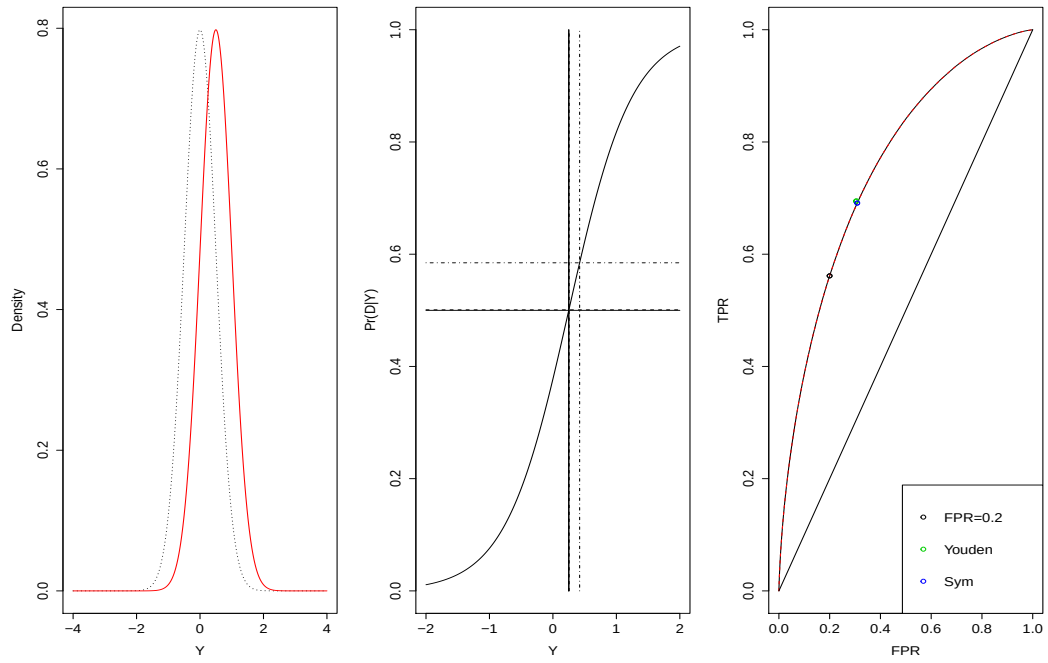


Figure 6.1: Theoretical model a). *Left panel*: Density functions of Y among diseased (*solid red line*) and healthy individuals (*dashed black line*). *Central panel*: Conditional probability of being diseased given Y ($\widetilde{Y} \equiv f(Y) = \Pr(D|Y)$) and optimal cut-off points derived from three different optimality criteria, FPR = 0.20 (*horizontal dashed-dotted line*), Youden index (*horizontal dashed line*) and Symmetry point (*horizontal solid line*), based on the transformed marker $\widetilde{Y}$. Vertical lines depict the corresponding cut-off points in the original scale of Y. *Right panel*: ROC curves for Y (*dashed red line*) and $\widetilde{Y}$ (*solid black line*), and operating points, $(1 - p(c), q(c))$, associated to the optimal cut-off points given by FPR = 0.20 (*black*), Youden index (*green*) and Symmetry point (*blue*).

Figure 6.1 shows the situation corresponding to model a). As can be seen, there is no interspersion in the marker's distribution among healthy and diseased subjects, so that the probability of being ill increases linearly with any increase in the value of Y. This situation is consistent with a linear logistic regression model, and so the classic approach is acceptable. In fact, both Y and $\tilde{Y}$ share the same ROC curve and consequently they yield the same results in terms of AUC values ($AUC_Y = AUC_{\tilde{Y}} = 0.76$) and optimal cut-off points. From the central panel of Figure corresponding to model a), it is observed that for each of the three optimality criteria represented in the transformed scale of the marker Y (the criterion that sets $FPR = 0.20$, the Youden index and the Symmetry point), there is a single cut-off point in the original scale Y. Besides, the horizontal lines (equivalently, the vertical lines) corresponding to the Youden index and the Symmetry point concur because these two criteria yield the same optimal cut-off point when the marker is normally distributed for both populations and both have the same standard deviation.

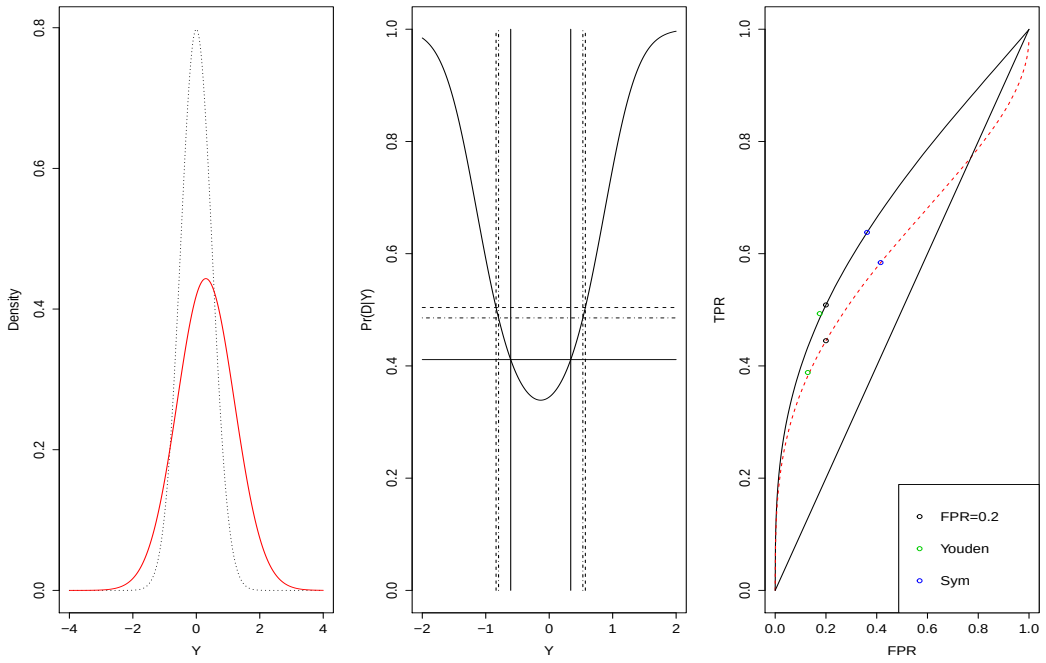


Figure 6.2: Theoretical model b). *Left panel*: Density functions of Y among diseased (*solid red line*) and healthy individuals (*dashed black line*). *Central panel*: Conditional probability of being diseased given Y ($\widetilde{Y} \equiv f(Y) = \Pr(D|Y)$) and optimal cut-off points derived from three different optimality criteria, $FPR = 0.20$ (*horizontal dashed-dotted line*), Youden index (*horizontal dashed line*) and Symmetry point (*horizontal solid line*), based on the transformed marker $\widetilde{Y}$. Vertical lines depict the corresponding cut-off points in the original scale of Y. *Right panel*: ROC curves for Y (*dashed red line*) and transformed marker $\widetilde{Y}$ (*solid black line*), and operating points, $(1 - p(c), q(c))$, associated to the optimal cut-off points given by $FPR = 0.20$ (*black*), Youden index (*green*) and Symmetry point (*blue*).

Under models b) and c), however, (see the left panel of Figures 6.2 and 6.3, respec-

tively) there is interspersion in the marker's distribution among healthy and diseased subjects. Hence, the probability of being ill does not show a monotone increasing behavior vis-á-vis the values of Y. This can be clearly seen in the central panel of Figures 6.2 and 6.3, where it is also observed that for any of the three optimality criteria represented in the transformed scale of Y (the criterion that sets FPR = 0.20, the Youden index and the Symmetry point), there are two cut-off points in the original scale of Y. Under these two models, $\tilde{Y}$ performs better than Y in terms of higher classificatory capacity. This can be seen in the right panel of Figures 6.2 and 6.3, which depict the ROC curves corresponding to Y and $\tilde{Y}$ under both models. The ROC curve of $\tilde{Y}$ is always above the ROC curve of Y. In other words, regardless of the FPF value chosen, the corresponding TPF is always higher (or the same) for the transformed marker. Lastly, whereas AUC values for the original marker Y are AUC_Y = 0.61 and AUC_Y = 0.51 in models b) and c), respectively, these values increase to $AUC_{\tilde{Y}}$ = 0.70 and $AUC_{\tilde{Y}}$ = 0.86 when using the transformed marker $\tilde{Y}$.

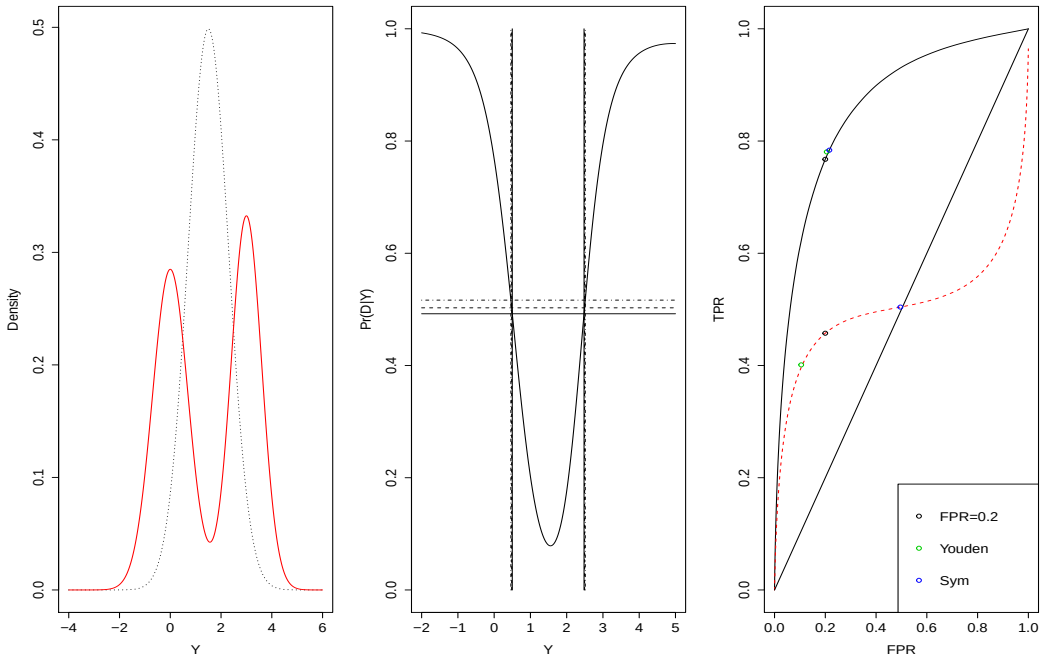


Figure 6.3: Theoretical model c). *Left panel*: Density functions of Y among diseased (*solid red line*) and healthy subjects (*dashed black line*). *Central panel*: Conditional probability of being diseased given Y ($\widetilde{Y} \equiv f(Y) = \Pr(D|Y)$) and optimal cut-off points derived from three different optimality criteria, FPR = 0.20 (*horizontal dashed-dotted line*), Youden index (*horizontal dashed line*) and Symmetry point (*horizontal solid line*), based on the transformed marker $\widetilde{Y}$. Vertical lines depict the corresponding cut-off points in the original scale of Y. *Right panel*: ROC curves for Y (*dashed red line*) and $\widetilde{Y}$ (*solid black line*), and operating points, $(1 - p(c), q(c))$, associated to the optimal cut-off points given by FPR = 0.20 (*black*), Youden index (*green*) and Symmetry point (*blue*).

In general, the suitability or unsuitability of using the transformed marker for discriminatory purposes is dictated by the form adopted by the function $f(Y) = \Pr(D|Y)$. If this function is monotone (either monotone increasing as in model a) or monotone decreasing), then using the transformed marker yields no gain in terms of discrimination capacity, since the ROC curve is invariant to this type of transformations (Swets et al., 1961; Egan, 1975).

Impact on selecting the optimal threshold

Situations in which the risk marker/disease relationship is not monotone increasing (or decreasing) give rise to a new challenge linked to the choice of the "optimal" cut-off point. In such cases, rather than proposing a single cut-off value as the classification rule, it would seem more reasonable to use an optimal interval that will define individuals with lower (higher) risk of presenting with the disease. Hence, unlike the classical approach, which consists of identifying a single cut-off point, c, for the marker, the proposed methodology enables a range of marker values to be obtained which are applicable to a wider variety of risk situations.

Based on the theoretical scenarios considered here, we propose the use of the transformed marker $\tilde{Y}$ for the choice of optimal cut-off point(s). Cut-off values are calculated in two stages, as follows: 1) the optimality criterion is applied to the transformed marker $\tilde{Y} = f(Y) = \Pr(D|Y)$, thereby obtaining the "optimal" probability or risk cut-off: and, 2) the original marker value/s that corresponds/correspond to this probability is/are then computed. For this purpose, the inverse function of f is calculated, $f^{-1}(\tilde{Y})$, for which an explicit form does not always exist but which can be calculated by means of numerical approximation techniques.

In the literature, as we pointed out in the previous chapters, there are numerous criteria for selecting optimal cut-off values in continuous diagnostic tests. By way of illustration, we considered in this study three different criteria: 1) the criterion that sets an FPF = 0.20, 2) the Youden index, and 3) the Symmetry point. We decided to select these methods because this choice 1) enables direct observation of the gain in sensitivity resulting from the use of the transformed marker, 2) the gain in terms of the Youden index value as a summary measure of the accuracy, and 3) the gain in both sensitivity and specificity (see Table 6.1 where we collect the parameters associated to the above theoretical models considered).

As can be seen from the central and right panels of Figure 6.1 and Table 6.1, for model a) the choice of optimal cut-off points is independent of the use of Y or $\tilde{Y}$, with a single cut-off point c_{Opt} being obtained in both cases, i.e., c_{Opt} = 0.42, with a corresponding TPF or $q(c_{Opt})$ = 0.56, considering the criterion that sets an FPF = 0.2, and c_{Opt} = 0.25, with a corresponding sensitivity of $q(c_{Opt})$ = 0.39, according to both the Youden index and Symmetry point that under this model provide the same cut-off point.

Under models b) and c), in contrast, (see the central and right panels of Figures 6.2 and 6.3, and Table 6.1) the choice of cut-off points changes according to whether the

decision is based on the original or transformed marker. In model b), the optimal cut-off point obtained by the classic method using the method that sets FPF = 0.2, is 0.42, which corresponds to a TPF = 0.44. When the choice of the cut-off point is based on the values of the transformed marker, the optimal probability value is c_{Opt} = 0.49 which corresponds to a TPF = 0.51, i.e., the transformed marker yields a 7% sensitivity gain over that obtained with the original marker. The optimal probability value corresponds to an interval of original marker values ranging from c_1 = -0.80 to c_2 = 0.53. The classification rule derived from this optimal interval would lay down that individuals within this interval are to be deemed healthy and the remainder, diseased. If we consider instead the method based on Youden index, we obtain a higher optimal cutpoint c_{Opt} = 0.56 for the original marker with a value of Youden index equal to 0.26, and an optimal probability c_{Opt} = 0.50 for the transformed marker which corresponds to a YI_{Opt} = 0.32, obtaining thus a 6% gain in the Youden index when using the transformed marker compared to the original marker. In this case, this optimal $\widetilde{Y}$ value of 0.50 corresponds to an interval of original marker values varying from c_1 = -0.84 to c_2 = 0.57. Finally, based on the Symmetry point, the optimal cut point obtained by the classical approach is c_{Opt} = 0.11, a lower value than the value obtained using previous criteria, with a true positive fraction $q(c_{Opt}) = p(c_{Opt})$ = 0.58. However, based on the transformed marker, we obtain an optimal value c_{Opt} = 0.41 corresponding to a true positive fraction $q(c_{Opt}) = p(c_{Opt})$ = 0.64, which means a 6% sensitivity gain and, equivalently, a 6% specificity gain. The corresponding interval of original marker values associated to this optimal probability is given by (c_1, c_2) = (-0.61,0.34).

Under model c), the phenomenon, albeit analogous, is more pronounced in terms of sensitivity or Youden index gain. On one hand, if one were to base oneself on the original marker and take the marker value showing an FPF = 0.20 as being optimal, a cut-off point of 2.17 associated with a TPF = 0.46 would be obtained. On the other hand, if the transformed marker point with an FPF = 0.20 were taken as optimal, a value of 0.52 would be obtained, which has a TPF = 0.77 and yields an interval of original marker values given by (0.46,2.51). In other words, the latter case would result in a 31 % sensitivity gain. Similar values of gain are obtained using the other optimal-cutpoint selection criteria here considered, a 27% Youden index gain in the case of the Youden index and a 28 % sensitivity (equivalently, specificity) gain in the case of the Symmetry point.

In brief, in any case where the risk is not monotone (models b) and c)), applying the optimal cut-off point selection criteria to the transformed marker ensures better results in terms of either higher values of sensitivity or higher values of Youden index.

In the simulation study, for every model previously specified, 1000 trials were considered. For each trial, a sample of n_0 i.i.d. observations, $\{Y_{01}, \ldots, Y_{0n_0}\}$, and a sample of n_1 i.i.d. observations, $\{Y_{11}, \ldots, Y_{1n_1}\}$, were independently drawn from markers in both healthy and diseased populations Y_0 and Y_1, respectively. Specifically, we have considered the following pairs of sample sizes that include both balanced and unbalanced designs to mimic real situations that are frequently seen in clinical practice: $(n_0, n_1) = (30, 60), (45, 45), (60, 30)$, with a total sample size of $n = n_0 + n_1 = 90$, and $(n_0, n_1) = (50, 100), (75, 75), (100, 50)$, with a total sample size of $n = n_0 + n_1 = 150$.

6.3.2 Results

For illustration, we present here the results obtained when the methodology proposed in Section 6.2 was applied to the theoretical scenarios considered previously in Subsection 6.3.1, that is, the transformed marker values were estimated by the corresponding logistic GAM regression model (6.2). For the sake of illustration, we have only consider the optimality criterion based on the Symmetry point, because in terms of gain it is similar to the criterion that sets an FPF = 0.20 and it is the optimal criterion on which we have focused our attention more. In fact, we have considered the Empirical Likelihood method (EL) previously explained in Chapter 4 to estimate the Symmetry point and its corresponding pair of sensitivity and specificity indexes.

Table 6.2: Mean, MSE, $Bias$ and $SBias$ of the nonparametric point estimators of c_S and $p(c_S) = q(c_S)$ for the estimated transformed marker based on the logistic GAM regression model and for the original marker under model a).

	Estimated transformed marker $\widetilde{Y}$							
	c_S				$p(c_S) = q(c_S)$			
	c							
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-3}$	SBias $\times 10^{-2}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-3}$	SBias $\times 10^{-1}$
(30,60)	0.4987 0.2518	0.1538 1.2844	0.1637 4.6304	0.4173 4.0870	0.6939	2.4344	3.2134	0.6523
(45,45)	0.4999 0.2497	0.1321 0.8917	1.3435 2.5130	3.6977 2.6608	0.6916	2.1521	0.9408	0.2027
(60,30)	0.5023 0.2487	0.1406 1.2526	3.7289 1.5493	9.9900 1.3837	0.6911	2.1513	0.4654	0.1003
(50,100)	0.5006 0.2532	0.0887 0.5431	1.9815 5.9822	6.6643 8.1402	0.6909	1.5772	0.2616	0.0658
(75,75)	0.4993 0.2491	0.0797 0.3828	0.7200 1.9292	2.5505 3.1180	0.6946	1.3160	3.9640	1.0988
(100,50)	0.5008 0.2477	0.0878 0.4542	2.1726 0.4813	7.3462 0.7139	0.6906	1.5065	-0.0345	-0.0089

	Original marker Y							
	c_S				$p(c_S) = q(c_S)$			
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-3}$	SBias $\times 10^{-2}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-3}$	SBias $\times 10^{-1}$
(30,60)	0.2499	0.5317	2.4495	3.3593	0.6916	2.4879	0.8381	0.1680
(45,45)	0.2504	0.5140	2.9798	4.1576	0.6895	2.2631	-1.2879	-0.2707
(60,30)	0.2488	0.5192	1.3326	1.8487	0.6882	2.2391	-2.5825	-0.5463
(50,100)	0.2509	0.3495	3.4447	5.8341	0.6899	1.6048	-0.9032	-0.2254
(75,75)	0.2486	0.2887	1.1695	2.1762	0.6939	1.3333	3.0626	0.8413
(100,50)	0.2481	0.3031	0.7150	1.2983	0.6898	1.5101	-1.0560	-0.2717

For every model, we collect in the same table the results obtained for the Symmetry point, c_S, and the associated specificity and sensitivity indexes, $p(c_S) = q(c_S)$ for both the estimated transformed marker $\widetilde{Y}$ and the original marker Y. Specifically, in Tables 6.2–6.4 we collect the results obtained for models a)–c) considering the six pairs of sample sizes (n_0, n_1) previously specified. It should be noted that depending on the marker scale we

have used different scientific notation ($\times 10^0$, $\times 10^{-1}$, $\times 10^{-2}$, etc.) in these tables to better represent significant figures.

Table 6.3: Mean, MSE, $Bias$ and $SBias$ of the nonparametric point estimators of c_S and $p(c_S) = q(c_S)$ for the estimated transformed marker based on the logistic GAM regression model and for the original marker under model b).

	Estimated transformed marker $\widetilde{Y}$							
	c_S c_1 c_2				$p(c_S) = q(c_S)$			
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$
(30,60)	0.4200	0.0964	0.8692	2.9148				
	-0.5972	5.9172	0.9836	0.4045	0.6598	2.8167	2.1897	4.5271
	0.3504	2.3865	1.1268	0.7310				
(45,45)	0.4247	0.0972	1.3425	4.7699				
	-0.5902	6.0822	1.6761	0.6808	0.6576	2.5163	1.9691	4.2658
	0.3664	2.4416	2.7280	1.7722				
(60,30)	0.4291	0.1311	1.7827	5.6549				
	-0.5846	5.8338	2.2353	0.9289	0.6584	2.9252	2.0465	4.0859
	0.3643	2.6710	2.5172	1.5581				
(50,100)	0.4190	0.5921	0.7736	3.3513				
	-0.6008	4.4907	0.6210	0.2930	0.6506	1.4804	1.2642	3.4771
	0.3478	1.7084	0.8639	0.6620				
(75,75)	0.4213	0.0625	1.0007	4.3656				
	-0.6154	4.5392	-0.8367	-0.3928	0.6479	1.3232	1.0001	2.8581
	0.3543	1.6128	1.5146	1.2006				
(100,50)	0.4255	0.0867	1.4184	5.4951				
	-0.6115	4.6127	-0.4466	-0.2079	0.6509	1.7497	1.2931	3.2489
	0.3584	1.8503	1.9294	1.4322				

	Original marker Y							
	c_S				$p(c_S) = q(c_S)$			
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$
(30,60)	0.1079	0.7902	0.0995	0.1118	0.5860	2.3976	0.1727	0.3527
(45,45)	0.1024	0.7437	-0.4503	-0.5226	0.5811	2.7954	-0.3168	-0.6000
(60,30)	0.1050	0.7472	-0.1925	-0.2227	0.5817	3.1191	-0.2572	-0.4608
(50,100)	0.1078	0.4773	0.0833	0.1205	0.5842	1.5467	-0.0017	-0.0042
(75,75)	0.1068	0.4320	-0.0142	-0.0215	0.5833	1.5938	-0.0924	-0.2314
(100,50)	0.1059	0.4998	-0.0982	-0.1389	0.5843	2.2918	0.0084	0.0175

Under model a) we observe from Table 6.2 that as the total sample size, n, increase the MSE for all the estimators decreases for both balanced and unbalanced designs. However, for a fixed total sample size, the MSE is in general smaller when $n_0 = n_1$ and larger when $n_1 > n_0$. Besides, there is a slight loss in accuracy when estimating the Symmetry point of Y if we base the estimation on $\widetilde{Y}$ (the values of MSE obtained for the estimates of c using $\widetilde{Y}$ are larger than those obtained for c_S using Y). However, this effect is the opposite regardig the estimates of $p(c_S) = q(c_S)$. Under model a), we can conclude that

very similar results are obtained, either based on Y or $\widetilde{Y}$.

Table 6.4: Mean, MSE, $Bias$ and $SBias$ of the nonparametric point estimators of c_S and $p(c_S) = q(c_S)$ for the estimated transformed marker based on the logistic GAM regression model and for the original marker under model c).

	Estimated transformed marker $\widetilde{Y}$							
	c_S c_1 c_2				$p(c_S) = q(c_S)$			
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$
(30,60)	0.4543 0.4994 2.4798	0.4122 3.1799 2.5871	-3.5859 -0.2931 0.3461	-6.7298 -0.1643 0.2151	0.7932	2.0824	1.0605	2.3883
(45,45)	0.4591 0.5009 2.4893	0.3554 2.3856 2.2689	-3.1056 -0.1504 1.2993	-6.0995 -0.0973 0.8654	0.7907	1.7589	0.8079	1.9621
(60,30)	0.4650 0.4811 2.5065	0.3925 3.2622 3.3432	-2.5175 -2.1273 3.0168	-4.3861 -1.1855 1.6720	0.7921	2.0600	0.9462	2.1305
(50,100)	0.4586 0.5046 2.4842	0.3200 1.9442 1.5209	-3.1545 0.2245 0.7922	-6.7141 0.1610 0.6434	0.7877	1.2246	0.5046	1.4564
(75,75)	0.4624 0.4980 2.4846	0.2756 1.7721 1.7218	-2.7797 -0.4350 0.8280	-6.2384 -0.3268 0.6319	0.7881	1.0549	0.5457	1.7036
(100,50)	0.4657 0.4995 2.4984	0.2838 1.7937 1.8853	-2.4519 -0.2839 2.2095	-5.1816 -0.2119 1.6296	0.7885	1.1884	0.5837	1.7172

	Original marker Y							
	c_S				$p(c_S) = q(c_S)$			
(n_0, n_1)	Mean	MSE $\times 10^{-2}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$	Mean	MSE $\times 10^{-3}$	Bias $\times 10^{-2}$	SBias $\times 10^{-1}$
(30,60)	1.5060	3.9372	-0.2012	-0.1014	0.5049	3.6768	0.0396	0.0653
(45,45)	1.5048	3.4118	-0.3258	-0.1763	0.5024	4.5835	-0.2104	-0.3107
(60,30)	1.5085	4.1851	0.0466	0.0228	0.5026	6.4486	-0.1896	-0.2361
(50,100)	1.5065	2.5132	-0.1488	-0.0938	0.5063	2.2535	0.1849	0.3897
(75,75)	1.5120	2.0443	0.3946	0.2760	0.5067	2.6557	0.2205	0.4280
(100,50)	1.5045	2.5184	-0.3485	-0.2196	0.5022	4.0704	-0.2319	-0.3635

From Tables 6.3–6.4, we observe that as the total sample size increases, the general tendency is that the MSE for all the estimators decreases. However, unlike model a), now there is no a so clear pattern in the sense that not always the smallest MSE is obtained for the balanced design when the total sample size is fixed. Unlike model a), it does not make sense to compare, for instance, the MSE obtained for estimating $p(c_S)$ using Y and that obtained using $\widetilde{Y}$ because the target pursued is different under models b) and c) (that is, the specificities involved are not the same). Under models b)-c), it is clearly seen that transformed marker estimated through the logistic GAM regression model yields a

better performance in terms of higher specificity and sensitivity. Besides, it approximates well the behaviour of the theoretical transformed marker.

As a general conclusion, we strongly recommend the use of the estimated transformed marker when $f(Y) = \Pr(D|Y)$ function is not monotone. Besides, even in cases where this relationship is monotone and therefore there is no gain in using the transformed marker, its use yields similar results to those obtained using Y.

6.4 Biomedical application

The methodology introduced in Section 6.2 is applied here to the postoperative infection dataset introduced in Section 1.3 of Chapter 1 in order to evaluate the discriminatory capacity of the plasma glucose as a potential postoperative infection biomarker. Specifically, the following logistic GAM regression model was fitted:

$$\Pr(POI = 1|Glucose) = g^{-1}(\alpha + h(Glucose)) = \frac{exp(\alpha + h(Glucose))}{1 + exp(\alpha + h(Glucose))}, \qquad (6.4)$$

where $Glucose$ is the continuous variable that represents the plasma glucose levels, that is, the original marker, POI is the binary indicator variable of the presence (POI = 1) or absence (POI = 0) of postoperative infection, $g(.)$ is the logit link function, and $h(.)$ is an unknown smooth function.

To assess the classificatory capacity of the plasma glucose levels, the ROC curve and AUC were calculated for both the original marker ($Glucose$) and the GAM transformed marker obtained from the fit of model (6.4). Besides, glucose cut-off points were obtained by the Symmetry point using the original marker and the transformed marker.

The left panel of Figure 6.4 shows smooth estimates of the marker density among healthy and diseased subjects. It is observed that glucose distribution in healthy and diseased subjects displays a certain level of interspersion, with characteristics similar to those displayed under Scenario 2. Individuals with POI registered extreme glucose values (high and low) compared with those who failed to develop postoperative infection (intermediate values).

In the central panel of Figure 6.4 we represent the probability of POI according to the values of $Glucose$. Similarly to Scenario 2, such probability function is not monotone increasing and this means that when using the transformed marker, the classificatory capacity improves. The ROC curves for the original glucose marker and the GAM transformed marker (see model (6.4)) are shown in the right panel of Figure 6.4. It is observed that the estimated transformed marker $\widetilde{Glucose}$ yields an ROC curve which is always above the ROC curve of $Glucose$. Based on the original marker, the AUC is 0.60 and its corresponding 95% bootstrap confidence interval (CI) (Efron, 1979) is (0.49, 0.70). In the light of this result, there would seem to be no evidence to indicate that glucose values can be used for discerning the states of POI. However, this is not the case when looking at the results obtained in the transformed scale: an AUC value of 0.69 and a CI equal to (0.60, 0.78), with the value of 0.5 lying outside this interval. Therefore, these results allow for plasma glucose to be regarded as a possible marker for POI.

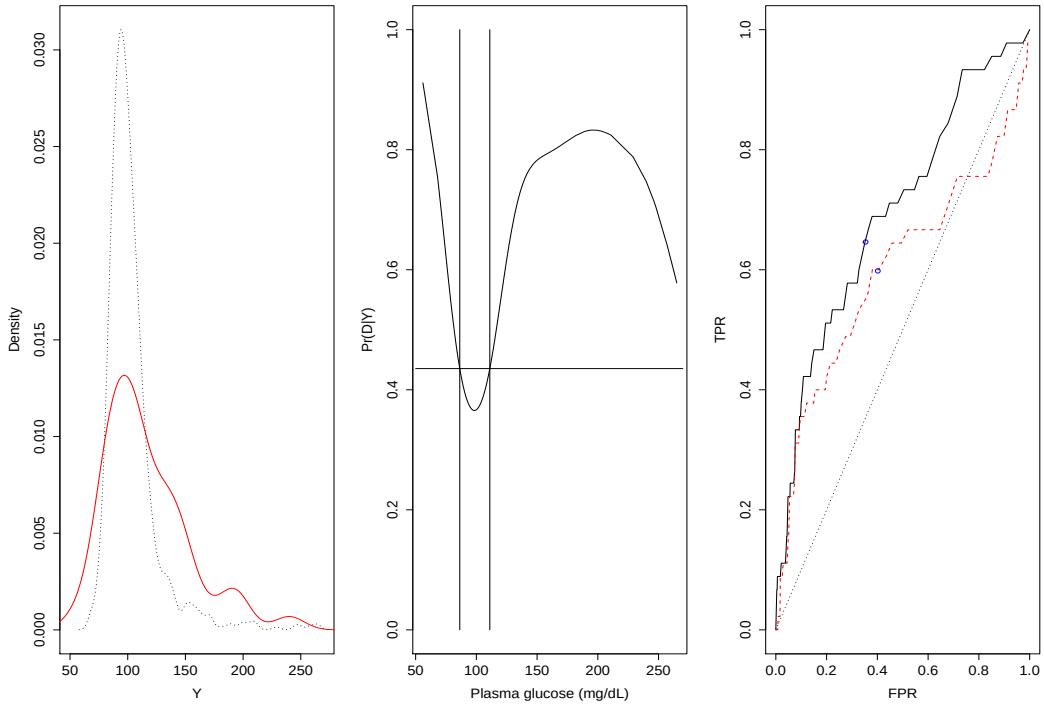


Figure 6.4: Analysis of $Glucose$ as a potential POI marker. *Left panel*: Smooth density estimates of $Glucose$ in the presence (*red solid line*) and absence (*black dashed line*) of POI. *Central panel*: Logistic GAM based estimation of the conditional probability of $POI = 1$ given $Glucose$ ($\widetilde{Glucose} \equiv f(Glucose) = \Pr(POI = 1|Glucose)$), and optimal cut-off point derived from the Symmetry point (*horizontal solid line*) based on the estimated transformed marker $\widetilde{Glucose}$. *Vertical lines* depict the two corresponding cut-off points in the original scale of $Glucose$. *Right panel*: ROC curves for $Glucose$ (*dashed red line*) and for the estimated transformed marker $\widetilde{Glucose}$ (*solid black line*), and operating points, $(1 - p(c), q(c))$, associated to the optimal cut-off points given by the Symmetry point (*blue*). The estimated transformed marker $\widetilde{Glucose}$ yields an ROC curve which is always above the ROC curve of $Glucose$.

With respect to the optimal cutpoints, when the optimality criterion was applied to the original marker values, the resulting glucose cut-off value was 101.787 mg/dL (Se = Sp = 0.598). When this same optimality condition was applied to the transformed glucose marker, an optimal probability value of 0.435 (Se = Sp = 0.646) was obtained, which corresponds to original glucose values in the interval between 86.324 mg/dL and 111.059 mg/dL. As in theoretical Scenarios 2 and 3, the choice of the optimal cut-off point varies according to whether the optimality criterion is based on the glucose values or on the risk associated with each such value. In this case, the transformed marker yields a 5% gain in sensitivity, a similar value obtained under Scenario 2.

Lastly, Figure 6.5 depicts the $lnOR$ function for the reference value y_{ref} = 95 mg/dL, that has been obtained as a byproduct from the fit of model (6.4), using the expression given in (6.3). As can be seen, rather than being linear, the relationship between $Glucose$ and POI shows a "spoon" shape (Figueiras and Cadarso-Suárez, 2001; Sáez et al., 2003). In other words, situations of hypo- and hyperglycaemia pose a higher risk of suffering from POI than do intermediate values.

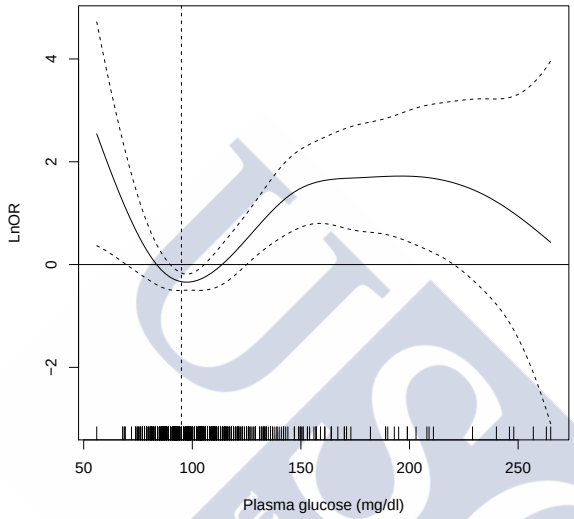


Figure 6.5: Estimation of Ln $OR(y; y_{ref})$, with $y_{ref} = 95$ mg/dL.

In literature and in life we ultimately pursue, not conclusions, but beginnings.

- Sam Tanenhaus

Chapter 7

Discussion and future research

In this chapter we present a summary of the main results presented in Chapters 3–6, together with some conclusions and interesting lines for future research.

7.1 Chapter 3: "Criteria to select the optimal cutoff point in diagnostic tests"

Several criteria have been proposed in the literature for selecting the optimal cutpoint of a continuous diagnostic test or biomarker in order to classify the individuals in two groups of interest. From the review that we have presented in Chapter 3, we have detected that some of the existing dichotomization criteria are equivalent from a theoretical perspective, that is, they define the same optimal cutpoint. However, it should be noted that they may yield different cut-off values from a practical viewpoint, depending on the estimation method or implementation process that is used. Therefore, an interesting new research line would be to perform a simulation study to evaluate which method among those that have been proposed to estimate equivalent optimal criteria is the most efficient one in terms of bias, standardized bias and mean squared error .

Moreover, some methods, although different from a theoretical point of view, can provide the same optimal cutpoint under specific situations. For example, under the assumption that the marker in healthy and diseased populations follows a normal distribution with the same standard deviation, the North-West corner (Metz, 1978; Vermont et al., 1991; Perkins and Schisterman, 2006), the Youden index (Youden, 1950; Aoki et al., 1997; Greiner et al., 2000), the Concordance probability (Lewis et al., 2008; Liu, 2012) and the Symmetry point (Greiner, 1995; Defreitas et al., 2004; Adlhoch et al., 2011) yield the same optimal cutpoint. Consequently, a comparative study between all the different (from a theoretical point of view) selection criteria is also an area for future research. In fact, we are improving the simulation study that we have conducted in Alvarenga Americano do Brasil et al. (2015) to check similarities of optimal cutpoints obtained by the variety of methods here studied, under different distributional scenarios taken from the literature (Fluss et al., 2005; Schisterman et al., 2008). Specifically, nine different population shapes, with varying conditions of disease prevalence, $\pi = 0.05, 0.1, 0.2, 0.3, 0.5, 0.75,$ and test accuracy measured in terms of AUC, AUC $= 0.6, 0.7, 0.8, 0.9,$ were simulated.

Such populations cover the following distributional models: normal (with and without equal variances in both healthy and diseased groups), lognormal, gamma, and several normal mixtures (where either both groups follow a bimodal normal distribution or only one of them does).

In this thesis, we have focused on the dichotomization methods to discriminate between two groups. However, there exist several extensions to the three group case: the thrichotomization methods and the three-class methods. On one hand, in the thrichotomization methods, the objective is to split the scale of the marker into three groups by means of two thresholds, even when there are only two groups of interest (healthy and diseased). These methods create an extra third group of "uncertainty", usually an inconclusive, intermediate or indeterminate group, represented by a range of test values where the test does not perform well to classify in any of the two groups of interest. On the other hand, the three class problems are those aiming at discriminating among three different groups such as: "condition A present", "condition B present" and "neither condition A nor B is present", or "without the condition", "with moderate condition" and "with severe condition". These strategies require that the three classes of interest are ordered, and thus suffer from the same limitations as most of the dichotomization methods do. Similarly to the dichotomization methods, the thrichotomization methods can be also classified into *completely data driven methods* and *methods with user requirements*. Some examples are the thrichotomization method based on the likelihood ratios gray (inconclusive) zone (Coste and Pouchot, 2003; Coste et al., 2006), or the thrichotomization with required sensitivity and specificity indexes (Greiner et al., 1995; Gallop et al., 2003). Among the three-class methods, we can find the three class Youden index (Nakas et al., 2010, 2013), the maximum expected decision utility (He and Frey, 2006) or methods based on ordinal logistic regression (Boland and Lehmann, 2010) and MANOVA (multivariate analysis of variance) (Zelman et al., 2005; Li et al., 2007; Hirschfeld and Zernikow, 2013a,b).

7.2 Chapter 4: "The Symmetry point and its cost-based generalization"

In Chapter 4 we have introduced two methods aimed at calculating point estimates and confidence intervals of the optimal cut-off value defined by the cost-based generalization of the Symmetry point, the Generalized Symmetry point, c_{GS}, that is, the cut-off value that maximizes simultaneously the two types of correct classifications or, equivalently, that satisfies the following condition, $\rho(1 - p(c_{GS})) = (1 - q(c_{GS}))$, where $\rho = c_{F+}/c_{F-}$ is the relative cost (or loss) of a false-positive classification as compared with a false-negative classification. The first method is a parametric approach based on the Generalized Pivotal Quantity (GPQ) (Weerahandi, 1993, 1995) that requires the assumption of binormality. The second method is a nonparametric approach based on the Empirical Likelihood (EL) methodology (Thomas and Grunkemeier, 1975) and the bootstrap technique. As a general concluding remark derived from our simulation study, we have seen that the EL approach is competitive with the GPQ approach when the data follow the Box-Cox family and outperforms the GPQ approach when the data do not follow the Box-Cox family and the healthy and diseased populations follow different parametric

models. Therefore, although the implementation of the EL method is more time consuming than the GPQ method, we recommend the use of the EL method when the true distributions of healthy and diseased populations are unknown.

Since the choice of the optimal cutpoint selection criterion should be based on the researcher's specific goal and the diagnostic properties sought, which will clearly depend on the disease under study, it would be interesting to extend the GPQ and EL methodologies to other optimal cutpoint selection criteria studied in Chapter 3, and to the incorporation of covariates that can affect the discrimination ability of the biomarker under study. Besides, since we have focused only on complete data cases, it would be also interesting to generalize the GPQ and EL methodologies to be used in situations of partial disease verification, that is, when only a portion of individuals have their true disease status ultimately verified, see Alonzo (2014) and references therein. Regarding this situation, it is interesting to mention here the advances carried out recently by Wang and Qin, see Wang and Qin (2013), who proposed several bias-corrected joint empirical likelihood confidence regions for the sensitivity and specificity indexes associated to a given threshold value with partial disease verification data. It would be interesting as well to generalize our approach to the case of a censored failure time outcome. In this setting, Rota et al. (2015) have recently extended the optimal cutpoint selection criteria given by the Youden index, the Concordance probability and the North-West corner by means of non-parametric estimators of the sensitivity and specificity indexes that account for censoring.

7.3 Chapter 5: "Software development: R packages"

Regarding the importance of having software implemented in a user-friendly environment that facilitates to professionals such as epidemiologists and clinicians, the process of obtaining optimal cutpoints in practice, we have described in Chapter 5 the two `R` packages that we have developed to estimate optimal cutpoints using most of the criteria and methodologies outlined in Chapters 3 and 4 for continuous diagnostic tests.

Section 5.1 of Chapter 5 is devoted to describe and illustrate `OptimalCutpoints` (López-Ratón and Rodríguez-Álvarez, 2014; López-Ratón et al., 2014), a user-friendly `R` package that allows users to choose among several popular methods (see Chapter 3). Unlike other packages (Freeman and Moisen, 2008; Brasil, 2010), `OptimalCutpoints` enables optimal thresholds to be calculated according to each of the levels of given categorical covariates. This is of great interest because the discriminatory capacity of a biomarker may often be different depending on certain characteristics, such as a particular patient's age group, gender or severity of disease (Pepe, 2004), and so when it comes to selecting the optimal cutpoint, this must be borne in mind in order to avoid drawing erroneous conclusions. Moreover, some packages only allow for the diagnostic test to take values from 0 to 1, since they are specifically designed for predictive diagnostic models, but in `OptimalCutpoints`, no restriction has been imposed with respect to the range of values of the diagnostic test. Thus, it can take values in a continuous range, including the unit interval, in case of a risk score obtained from a predictive diagnostic model. In addition, `OptimalCutpoints` includes criteria not included previously in other existing packages, such as criteria based on predictive values or likelihood ratios, and allows the incorporation of costs and/or prevalence in some of these criteria. Moreover, with

our package, users can easily obtain numerical (point and confidence interval estimates) and graphical output for all the implemented methods with just one input command, and make decisions accordingly. We trust that a program displaying these features will prove useful to the biomedical community; a program which we are thinking of continually improve, by enabling, say, the incorporation of covariates of a continuous nature and implementing new and more efficient methods for estimating the optimal cutpoints defined by the criteria outlined. In the current version of the package only empirical estimators are implemented, and it would be interested to implement more sophisticated methodologies for the available optimal criteria such as the GPQ and EL methodologies studied in Chapter 4 for the Generalized Symmetry point. According to several future research lines previously mentioned, it would be also interesting to include new functions in `OptimalCutpoints` that implement optimal criteria suitable for situations of partial disease verification, see Alonzo (2014) and references therein.

In Section 5.2 of Chapter 5, we have developed `GsymPoint` (López-Ratón et al., 2015b), a user-friendly `R` package that allows users to estimate the Generalized Symmetry point and its sensitivity and specificity indexes using recent methodology introduced in ROC analysis (Molanes-López and Letón, 2011; Lai et al., 2012), such as the GPQ and EL techniques proposed in Chapter 4, which are more efficient than the empirical approach. As in the case of `OptimalCutpoints`, the estimations can be obtained straightforwardly by levels of certain categorical covariates in `GsymPoint`. Possible interesting extensions of the `GsymPoint` package could be taken into account, so we are contemplating the possibility of applying this same methodology to estimate the other accuracy measures, such as the predictive values or the diagnostic likelihood ratios, the possibility of incorporating more efficient methods for estimating the Generalized Symmetry point and its accuracy measures, and also the possibility of extending this new methodology to estimations corrected for the measurement of error, to partial disease verification (see Alonzo (2014) and references therein), to the case of a censored failure time outcome (Rota et al., 2015) and to the incorporation of continuous covariates.

This study has been centered on the field of diagnostic tests, but the two `R` packages developed in this framework, `OptimalCutpoints` and `GsymPoint`, may also be applied in any field where signal-to-noise analysis is performed, such as screening, radiodiagnostic techniques or biology, among others.

The implementation of a new `R` package that incorporates the existing trichotomization and three class methods may constitute another important issue to cover in the future. Up to our knowledge, `DiagTest3Grp` is nowadays the only existing `R` package for analyzing diagnostic tests with three ordinal groups that implements a generalization of the Youden index to this situation (Luo and Xiong, 2012).

7.4 Chapter 6: "Improvement of the discriminatory capacity via generalized additive models"

The study performed in Chapter 6 has shown firstly, how the use of the logistic GAM regression model improves the classificatory capacity of a potential continuous marker,

particularly in situations where there is a high level of interspersion in the marker's distribution among the two populations of healthy and diseased subjects; and secondly, how this type of modeling may modify the choice of suitable cut-off points.

Situations of nonlinearity in the marker/outcome relationship tend to arise relatively frequently in the sphere of clinical and epidemiological research. For instance, it is frequent for U- or J-shaped relationships to be found when analyzing the appearance of cardiovascular events and plasma glucose (Cid-Álvarez et al., 2009) or hormone levels (Bertone-Johnson et al., 2009). In the different theoretical models analyzed, as well as in the real data application on postoperative infection risk, the classificatory capacity seems to improve in terms of ROC and AUC when the transformed marker is used in situations that reflect a nonlinear risk marker/disease relationship. An additional advantage of the method proposed here is that it provides a tool for calculating more appropriate cut-off points, because it takes the relationship between marker values and clinical outcomes into account. Hence, in such U- or J-shaped relationships, two values will be selected to delimit an optimal interval where there is a lower probability of being ill. Under these situations, the classification rule consists on labeling an individual as healthy if its corresponding value on the original marker scale lies in that optimal interval and as diseased otherwise.

For the sake of illustration, we have considered three optimal criteria to select the cut-off points (the criteria based on a designated FPF value of 0.2 to be set, the Youden index and the Symmetry point). From a theoretical perspective, these criteria have allowed us to evaluate the improvement obtained by using the transformed marker over the original one in terms of the gain obtained in three different accuracy measures, sensitivity (given a fixed value of specificity), Youden index (maximization of sensitivity plus specificity) and simultaneous maximization of specificity and sensitivity. Similarly, other optimal cut-off selection criteria could have been used. Since in practice the transformed marker is unknown and has to be estimated, we have carried out a simulation study to evaluate the behaviour of our estimated transformed marker based on the logistic GAM regression model. However, rather than considering the three above-mentioned optimal criteria, we have only considered the Symmetry point in both the simulation study and the real data application because it has been the optimal criterion on which we have focused more and, besides, similar gains are obtained using the other two optimal criteria. Similarly, other optimal cut-off selection criteria could have been used. Using the Symmetry point, we have obtained a plasma glucose interval of lower POI risk that ranges from 86 to 111 mg/dL. These cut-off points are in line with the results of published studies on the relationship between infection risk and glucose values (Van den Berghe et al., 2001).

To our knowledge, this is the first time an optimal interval -as opposed to a single point- has been proposed for enhancing the accuracy of a continuous marker. However, caution must be exercised when choosing the threshold or range of thresholds. The threshold selection criterion should be based on the clinical context, and may involve weighing the expected costs against the benefits associated with a high-risk designation. Other clinical settings may involve an entirely different balance between disease severity, treatment efficacy and side-effects, leading to different choices according to the importance of sensitivity versus specificity. One of the strengths of this proposal is that the proposed methodology can be easily used by clinicians and epidemiologists. We used spline smoothing methods to estimate the logistic GAM regression model and empirical

estimates to approximate the ROC curves and their related accuracy measures but other available tools, such as kernel or P-spline smoothing techniques (Du and Tang, 2009) could be used similarly.

As a general conclusion, we can say that it is important to stress the need to question linearity in marker-outcome relationships: firstly, because failure to do so may lead to erroneous conclusions; and secondly, because the use of statistical tools that allow for greater flexibility (e.g., GAMs) can optimize the classificatory capacity of a potential marker from the standpoint of ROC analysis.

The results obtained in this study can be extended to other biomedical scenarios, such as survival analysis. In this context, the outcome variable is deemed to be time dependent, and the concepts of ROC curve and AUC have been adapted to this situation (Heagerty et al., 2000; Heagerty and Zheng, 2005). The fitting of Cox regression models (Cox, 1972) used for survival analysis and the modeling of hazard ratios may likewise be performed by flexible methods analogous to those described for logistic regression (Cid-Álvarez et al., 2009; Cadarso-Suárez et al., 2010), so that nonparametric modeling can also be used to obtain more suitable markers in this context.

A further interesting task linked to the methodology presented here is its extension to a multivariate context. ROC analysis can also be used in conjunction with logistic GAM regression models in the context of using combined markers for diagnosis purposes (Lado et al., 2006).

Bibliography

Abdolell, M., LeBlanc, M., Stephens, D., and Harrison, RV. (2002). Binary partitioning for continuous longitudinal data: Categorizing a prognostic variable. *Statistics in Medicine*, 21:3395–3409.

Adlhoch, C., Kaiser, M., Hoehne, M., Mas Marques, A., Stefas, I., Veas, F., and Ellerbrok, H. (2011). Highly sensitive detection of the group A Rotavirus using Apolipoprotein H-coated ELISA plates compared to quantitative real-time PCR. *Virology Journal*, 8:63.

Agresti, A. and Coull, BA. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52:119–126.

Akobeng, AK. (2007). Understanding diagnostic tests 1: Sensitivity, specificity and predictive values. *Acta Paediatrica*, 96(3):338–341.

Al-Mawali, A., Gillis, D., and Lewis, I. (2009). The use of Receiver Operating Characteristic analysis for detection of minimal residual disease using five-color multiparameter flow cytometry in acute myeloid leukemia identifies patients with high risk of relapse. *Clinical Cytometry*, 76B:91–101.

Albert, A. and Harris, EK. (1987). *Multivariate interpretation of clinical laboratory data*. Marcel Dekker, New York.

Alonzo, TA. (2014). Verification bias-impact and methods for correction when assessing accuracy of diagnostic tests. *REVSTAT-Statistical Journal*, 12:67–83.

Altemeier, W. (1979). *Surgical infections: Incisional wounds*. In: Bennet JV, Brachman P, editors. Hospital infections. Boston: Little, Brown and Company.

Altman, DG. (1991). Categorizing continuous-variables. *British Journal of Cancer*, 64:975.

Altman, DG. and Bland, JM. (1994a). Statistics notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ (Clinical Research ed.)*, 308:1552.

Altman, DG. and Bland, JM. (1994b). Statistics notes: Diagnostic tests 2: Predictive values. *BMJ (Clinical Research ed.)*, 309:102.

Altman, DG. and Bland, JM. (1994c). Statistics notes: Diagnostic tests 3: Receiver Operating Characteristic plots. *BMJ (Clinical Research ed.)*, 309:188.

Altman, DG., Lausen, B., Sauerbrei, W., and Schumacher, M. (1994). Dangers of using "Optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute*, 86(11):829–835.

Alvarenga Americano do Brasil, PE., Hirschfeld, G., de Souza Borges Quintana, M., López Ratón, M., and Rodríguez Álvarez, MX. (2015). Methods to estimate decision thresholds for diagnostic tests in medicine: Comparing similarities of 55 methods. *Submitted*.

Álvarez García, G., Collantes-Fernández, E., Costas, E., Rebordosa, X., and Ortega-Mora, LM. (2003). Influence of age and purpose for testing on the cut-off selection of serological methods in bovine neosporosis. *Veterinary Research*, 34:341–352.

Amaro, A., Gude, F., González-Juanatey, R., Iglesias, F., Fernández-Vázquez, F., García-Acuña, J., and Gil, M. (1995). Plasma leukocyte elastase concentration in angiographically diagnosed coronary artery disease. *European Heart Journal*, 16:615–622.

Anderson, JA. (1982). *Logistic regression*. In: Krishnaiah P.R, Kanal LN (Eds.), Handbook of Statistics. North-Holland, New York.

Aoki, K., Misumi, J., Kimura, T., Zhao, W., and Xie, T. (1997). Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogens I, II and of PG I/PG II ratios in a gastric cancer case-control study. *Journal of Epidemiology*, 7:143–151.

Baker, SG. (2009). Putting risk prediction in perspective: Relative utility curves. *Journal of the National Cancer Institute*, 101(22):1538–1542.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology*, 12:387–415.

Barajas-Rojas, JA., Riemann, HP., and Franti, CE. (1993). Notes about determining the cut-off value in enzyme-linked immunosorbent assay (ELISA). *Preventive Veterinary Medicine*, 15:231–233.

Barrat, A., Irwig, LM., Glasziou, PP., Cumming, RG., Raffle, A., Hicks, N., Gray, JA., and Guyatt, GH. (1999). Users' guides to the medical literature: XVII. How to use guidelines and recommendations about screening. Evidence-based medicine working group. *Journal of the American Medical Association*, 281:2029–2034.

Bartlett, JW., Frost, C., Mattsson, N., Skillbäck, T., Blennow, K., Zetterberg, H., and Schott, JM. (2012). Determining cut-points for Alzheimer as disease biomarkers: Statistical issues, methods and challenges. *Biomarkers in Medicine*, 6(4):391–400.

Begg, CB., Cramer, LD., Venkatraman, ES., and Rosai, J. (2000). Comparing tumour staging and grading systems: A case study and a review of the issues, using thymoma as a model. *Statistics in Medicine*, 19:1997–2014.

Begg, CB. and Greenes, RA. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, 39:207–215.

Benowitz, NL., Bernert, JT., Caraballo, RS., Holiday, DB., and Wang, J. (2009). Optimal serum cotinine levels for distinguish cigarette smokers and non smokers within different racial/ethnic groups in the United States between 1999 and 2004. *Americal Journal of Epidemiology*, 169:236–248.

Bergus, GR. (1993). When is a test positive? The use of decision analysis to optimize test interpretation. *Family Medicine*, 25:656–660.

Berkson, J. (1947). "Cost-Utility" as a measure of the efficiency of a test. *Journal of the American Statistical Association*, 42(238):246–255.

Bertone-Johnson, ER., Tworoger, SS., and Hankinson, SE. (2009). Recreational physical activity and steroid hormone levels in postmenopausal women. *American Journal of Epidemiology*, 170:1095–1104.

Bickel, PJ. and Doksum, KA. (1977). *Mathematical statistics: Basic ideas and selected topics*. Oakland, CA: Holden- Day Inc.

Björk, T., Piironen, T., Pettersson, K., Lövgren, T., Stenman, U-H., Oesterling, JE., Abrahamsson, P-A., and Lilja, H. (1996). Comparison of analysis of the different prostate-specific antigen forms in serum for detection of clinically localized prostate cancer. *Urology*, 48:882–888.

Bland, M. (1987). *An introduction to medical statistics*. Oxford: Oxford University Press.

Böhning, D., Holling, H., and Patilea, V. (2011). A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. *Statistical Methods in Medical Research*, 20(5):541–550.

Boland, MV. and Lehmann, HP. (2010). A new method for determining physician decision thresholds using empiric, uncertain recommendations. *BMC Medical Informatics and Decision Making*, 10:20.

Bortheiry, AL., Malerbi, DA., and Franco, LJ. (1994). The ROC curve in the evaluation of fasting capillary blood glucose as a screening test for diabetes and IGT. *Diabetes Care*, 17:1269–1272.

Botana-López, MA., López-Ratón, M., Tomé, MA., Fernández-Mariño, A., Mato-Mato, JA., Rego-Iraeta, A., Pérez-Fernández, R., and Cadarso-Suárez, C. (2012). Relationship between glycated hemoglobin and glucose concentrations in the adult Galician population: Selection of optimal glycated hemoglobin cut-off points as a diagnostic tool of diabetes mellitus. *Endocrinología y Nutrición*, 59(8):496–504.

Boyd, JC. (1997). Mathematical tools for demonstrating the clinical usefulness of biochemical markers. *Scandinavian Journal of Clinical and Laboratory Investigation*, 227:46–63.

Boyko, EJ. (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test: Short cut or wrong turn? *Medical Decision Making*, 14:175–179.

Bradley, AP. (2014). Half-AUC for the evaluation of sensitive or specific classifiers. *Pattern Recognition Letters*, 38:93–98.

Brandao, SCS., Nishioka, SAD., Giorgi, MCP., Chen, J., Abe, R., Filho, MM., Hotta, VT., Vieira, ML., Garcia, EV., and Meneghetti, JC. (2009). Cardiac resynchronization therapy evaluated by myocardial scintigraphy with 99mTc-MIBI: Changes in left ventricular uptake, dyssynchrony, and function. *European Journal of Nuclear Medicine and Molecular Imaging*, 36:986–996.

Brasil, P. (2010). ***DiagnosisMed**: Diagnostic test accuracy evaluation for medical professionals*. R package version 0.2.3.

Briggs, WM. and Zaretzki, R. (2008). The Skill Plot: A graphical technique for evaluating continuous diagnostic tests. *Biometrics*, 64(1):250–256.

Brown, CD. and Davis, HT. (2006). Receiver Operating Characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24–38.

Bruno, A., Akinwuntan, AE., Lin, C., Close, B., Davis, K., Baute, V., Aryal, T., Brooks, D., Hess, DC., Switzer, JA., and Nichols, FT. (2011). Simplified modified rankin scale questionnaire: Reproducibility over the telephone and validation with quality of life. *Stroke*, 42(8):2276–2279.

Burgueño, MJ., García-Bastos, JL., and González-Buitrago, JM. (1995). Las curvas ROC en la evaluación de pruebas diagnósticas. *Medicina Clínica*, 104:661–670.

Cadarso-Suárez, C., Meira-Machado, L., Kneib, T., and Gude, F. (2010). Flexible hazard ratio curves for continuous predictors in multi-state models: An application to breast cancer data. *Statistical Modelling*, 10:291–314.

Cai, T. and Dodd, LE. (2008). Regression analysis for the partial area under the ROC curve. *Statistica Sinica*, 18:817–836.

Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*, 13:499–508.

Cantor, SB., Sun, CC., Tortolero-Luna, G., Richards-Kortum, R., and Follen, M. (1999). A comparison of C/B ratios from studies using Receiver Operating Characteristic curve analysis. *Journal of Clinical Epidemiology*, 52:885–892.

Caraguel, CGB., Stryhn, H., Gagné, N., Dohoo, IR., and Hammell, KL. (2011). Selection of a cutoff value for real-time polymerase chain reaction results to fit a diagnostic purpose: Analytical and epidemiologic approaches. *Journal of Veterinary Diagnostic Investigation*, 23(1):2–15.

Carstensen, B., Plummer, M., Laara, E., and Hills, M. (2013). ***Epi**: A package for statistical analysis in epidemiology*. R package version 1.1.49.

Chen, JJ., Tsai, C-A., Moon, H., Ahn, H., Young, JJ., and Chen, CH. (2006). Decision threshold adjustment in class prediction. *SAR and QSAR in Environmental Research*, 17(3):337–352.

Chien, PF. and Khan, KS. (2001). Evaluation of a clinical test. II: Assessment of validity. *BJOG: An International Journal of Obstetrics and Gynaecology*, 108(6):568–572.

Choi, BCK. (1998). Slopes of a Receiver Operating Characteristic curve and likelihood ratios for a diagnostic test. *American Journal of Epidemiology*, 148:1127–1132.

Cid-Álvarez, B., Gude, F., Cadarso-Suárez, C., González-Babarro, E., Rodríguez-Álvarez, MX., JM, G.-A., and González-Juanatey, JR. (2009). Admission and fasting plasma glucose for estimating risk of death of diabetic and nondiabetic patients with acute coronary syndrome: Nonlinearity of hazard ratios and time-dependent comparison. *American Heart Journal*, 58:989–997.

Clopper, C. and Pearson, ES. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26:404–413.

Coffin, M. and Sukhatme, S. (1997). Receiver Operating Characteristic studies and measurement errors. *Biometrics*, 53:823–837.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Coste, J., Jourdain, P., and Pouchot, J. (2006). A gray zone assigned to inconclusive results of quantitative diagnostic tests: Application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients. *Clinical Chemistry*, 52(12):2229–2235.

Coste, J. and Pouchot, J. (2003). A grey zone for quantitative diagnostic and screening tests. *International Journal of Epidemiology*, 32(2):304–313.

Costello-Boerrigter, LC., Boerrigter, G., Redfield, MM., Rodeheffer, RJ., Urban, LH., Mahoney, DW., Jacobsen, SJ., Heublein, DM., and JC, B. (2006). Aminoterminal pro-B-type natriuretic peptide and B-type natriuretic peptide in the general community: Determinants and detection of left ventricular dysfunction. *Journal of the American Collegue of Cardiology*, 47(2):345–353.

Courdi, A., Hery, M., and Chauvel, P. (1988). Prognostic value of continuous variables in breast cancer and head and neck cancer. Dependence on the cut-off level. *British Journal of Cancer*, 58:88–90.

Cox, DR. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society, Series B Statistical Methodology*, 34:187–220.

Cramer, JS. (2003). *Logit models: From economics and other fields*. Cambridge Univ. Press.

Crits-Christoph, P., Connolly, MB., Gallop, R., Barber, JP., Tu, X., Gladis, M., and Siqueland, L. (2001). Early improvement during manual-guided cognitive and dynamic psychotherapies predicts 16 week remission status. *The Journal of Psychotherapy Practice and Research*, 10:145–154.

Cummings, SR. and Richard, RJ. (1988). Optimum cutoff points for biochemical validation of smoking status. *American Journal of Public Health*, 78(5):574–575.

David, HA. (1970). *Order statistics*. Wiley, New York.

Davidson, M. (2002). The interpretation of diagnostic test: A primer for physiotherapists. *The Australian Journal of Physiotherapy*, 48(3):227–32.

de Boor, C. (1978). *A practical guide to splines*. Springer-Verlag, New York.

DeBaun, MR. and Sox, HC. (1991). Setting the optimal erythrocyte protoporphyrin screening decision threshold for lead poisoning: A decision analytic approach. *Pediatrics*, 8:121–131.

Deeks, JJ. and Altman, DG. (2004). Diagnostic tests 4: Likelihood ratios. *BMJ (Clinical Research ed.)*, 329:168–169.

Defreitas, GA., Zimmern, PE., Lemack, GE., and Shariat, SF. (2004). Refining diagnosis of anatomic female bladder outlet obstruction: comparison of pressure-flow study parameters in clinically obstructed women with those of normal controls. *Urology*, 64:675–679.

Delong, ER., Delong, DM., and Clarke-Pearson, DL. (1988). Comparing the areas under two or more correlated Receiver Operating Characteristic curves: A nonparametric approach. *Biometrics*, 44:837–845.

DeNeef, P. and Kent, DL. (1993). Using treatment-tradeoff preferences to select diagnostic strategies: Linking the ROC curve to threshold analysis. *Medical Decision Making*, 13:126–132.

Du, P. and Tang, L. (2009). Transformation-invariant and nonparametric monotone smooth estimation of ROC curves. *Statistics in Medicine*, 28:349–359.

Dujardin, B., Van der Ende, J., Van Gompel, A., Unger, JP., and Van der Stuyft, P. (1994). Likelihood ratios: A real improvement for clinical decision making? *European Journal of Epidemiology*, 10:29–36.

Dwyer, AJ. (1996). In pursuit of a piece of the ROC. *Radiology*, 201(3):621–625.

Edwards, W., Guttentag, M., and Snapper, K. (1975). *A decision theoretic approach to evaluation research*. In: Struening EL, Guttentag M(eds): Handbook of Evaluation Research, vol1. Beverly Hills Calif, Sage.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26.

Egan, JP. (1975). *Signal detection theory and ROC Analysis*. New York: Academic Press.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121.

Eilers, PHC. and Marx, BD. (1992). Generalized linear models with P-splines. In *Proceedings of GLIM 92 and 7th International Workshop on Statistical Modelling, Munich, Germany*, pages 72–77. Lecture Notes in Statistics, Vol 78, Advances in GLIM and Statistical Modelling, Eds. L. Fahrmeir, B.Francis, R. Gilehrist, G.Tutz. Springer-Verlag, New York.

England, WL. (1988). An exponential model used for optimal threshold selection on ROC curves. *Medical Decision Making*, 8:120–131.

Epanechnikov, V. (1969). Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications*, 14:153–158.

Espallardo, NL. (2003). Decisions on diagnosis in family practice: Use of sensitivity, specificity, predictive values and likelihood ratios. *Asia Pacific Family Medicine*, 2:229–232.

Feinstein, AR. (1985). *Clinical epidemiology: The architecture of clinical research*. W.B. Saunders Company.

Feinstein, AR (1975). Clinical biostatistics. XXXI. On the sensitivity, specificity and discrimination of diagnostic tests. *Clinical Pharmacology and Therapeutics*, 17:104–116.

Feinstein, SH (1975). The accuracy of diver sound localization by pointing. *Understanding Biomedical Research*, 2(3):173–184.

Fescina, RH., Simini, F., and Belitzky, R. (1985). Evaluación de los procedimientos diagnósticos. Aspectos metodológicos. *Salud Perinatal*, 2:39–43.

Figueiras, A. and Cadarso-Suárez, C. (2001). Application of nonparametric models for calculating odds ratios and their confidence intervals for continuous exposures. *American Journal of Epidemiology*, 154:264–275.

Filella, X., Alcover, J., Molina, R., Giménez, N., Rodríguez, A., Jo, J., Carretero, P., and Ballesta, AM. (1995). Clinical usefulness of free PSA fraction as an indicator of prostate cancer. *International Journal of Cancer*, 63:780–784.

Fleiss, JL. (1981). *Statistical methods for rates and proportions*. John Wiley & Sons, New York.

Fluss, R., Faraggi, D., and Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. *Biometrical Journal*, 47:458–472.

Fombonne, E. (1991). The use of questionnaires in child psychiatry research: Measuring their performance and choosing an optimal cut-off. *Journal of Child Psychology and Psychiatry*, 32:677–693.

Freeman, E. (2007). ***PresenceAbsence**: An R package for presence-absence model evaluation*. USDA Forest Service, Rocky Mountain Research Station, 507 25th street, Ogden, UT, USA.

Freeman, EA. and Moisen, G. (2008). **PresenceAbsence**: An R package for presence-absence analysis. *Journal of Statistical Software*, 23(1):1–31.

Fujiyama, S., Izuno, K., Yamasaki, K., Sato, T., and Taketa, K. (1992). Determination of optimum cutoff levels of plasma des-gamma-carboxy prothrombin and serum alpha-fetoprotein for the diagnosis of hepatocellular carcinoma using Receiver Operating Characteristic curves. *Tumour Biology*, 13:316–323.

Gail, MH. and Pfeiffer, RM. (2005). On criteria for evaluating models for absolute risk. *Biostatistics*, 6:227–239.

Galen, RS. (1986). Use of predictive value theory in clinical immunology. In Rose, N. R., Friedmann, H., and Fahey, J. L., editors, *Manual of Clinical Laboratory Immunology*, pages 966–970. American Society of Microbiology. Washington, DC.

Gallas-Torreira, MM., Valladares-Durán, M., and López-Ratón, M. (2014). Comparison between two radiographic methods used for the prediction of mandibular third molar impaction. *Revista Portuguesa de Estomatologia, Medicina Dentária e Cirurgia Maxilofacial*, 55(4):207–213.

Gallop, RJ., Crits-Christoph, P., Muenz, LR., and Tu, XM. (2003). Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. *Understanding Statistics*, 2(4):219–242.

Gart, JJ. and Nam, J. (1998). Approximate interval estimation of the ratio of binomial parameters: A review and corrections for skewness. *Biometrics*, 44:323–338.

Geisser, S. (1998). Comparing two tests used for diagnostic or screening processes. *Statistics Probability Letters*, 40:113–119.

Giard, RW. and Hermans, J. (1993). The evaluation and interpretation of cervical cytology: Application of the likelihood ratio concept. *Cytopathology*, 4:131–137.

Gönen, M. and Sima, C. (2013). Optimal cutpoint estimation with censored data. *Journal of Statistical Theory and Practice*, 7:345–359.

Green, DM. and Swets, JA. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons, Inc. New York.

Greenhouse, W. and Mantel, N. (1950). The evaluation of diagnostic tests. *Biometrics*, 6:399–412.

Greenland, S. (1995). Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology*, 6:450–454.

Greenland, S. (2008). The need for reorientation toward cost-effective prediction: Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond" by M. J. Pencina et al. *Statistics in Medicine*, 27:199–206.

Greiner, M. (1995). Two-Graph Receiver Operating Characteristic (TG-ROC): A Microsoft-EXCEL template for the selection of cut-off values in diagnostic tests. *Journal of Immunological Methods*, 185(1):145–146.

Greiner, M. (1996). Two-Graph Receiver Operating Characteristic (TG-ROC): Update version supports optimisation of cut-off values that minimise overall misclassification costs. *Journal of Immunological Methods*, 191:93–94.

Greiner, M. and Böhning, D. (1994). Notes about determining the cut-off value in enzyme-linked immunosorbent assay (ELISA). Reply. *Preventive Veterinary Medicine*, 20:307–310.

Greiner, M., Pfeiffer, D., and Smith, RD. (2000). Principals and practical application of the Receiver Operating Characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, 45:23–41.

Greiner, M., Sohr, D., and Göbel, P. (1995). A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *Journal of Immunological Methods*, 185:123–132.

Griner, PF., Mayewski, RJ., Mushlin, AI., and Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine*, 94:557–563.

Grmec, S. and Gasparovic, V. (2001). Comparison of APACHE II, MEES and Glasgow Coma Scale in patients with nontraumatic coma for prediction of mortality. *Critical Care*, 5:19–23.

Hadzi-Pavlovic, D. (2008). Diagnostic tests: III. Optimal cut-offs. *Acta Neuropsychiatrica*, 20(5):273–276.

Hagen, MD. (1995). Test characteristics: How good is that test? *Primary Care*, 22:213–233.

Hajian-Tilaki, KO., Hanley, JA., Joseph, L., and Collect, JP. (1997). Extension of Receiver Operating Characteristic analysis to data concerning multiple signal detection tasks. *Academic Radiology*, 4:222–229.

Haldane, J. (1955). The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics*, 20:309–314.

Halkin, A., Reichman, J., Schwaber, M., Paltiel, O., and Brezis, M. (1998). Likelihood ratios: Getting diagnostic testing into perspective. *The Quartely Journal of Medicine*, 91:247–258.

Hall, PG. and Hyndman, RJ. (2003). Improved methods for bandwidth selection when estimating ROC curves. *Statistics & Probability Letters*, 64:181–189.

Halpern, EJ., Albert, M., Krieger, AM., Metz, CE., and Maidment, AD. (1996). Comparison of Receiver Operating Characteristic curves on the basis of optimal operating point. *Academic Radiology*, 3:245–253.

Hand, D. (1987). Screening versus prevalence estimation. *Applied Statistics*, 36:1–7.

Hanley, JA. (1988). The robustness of the binormal assumptions used in fitting ROC curves. *Medical Decision Making*, 8:197–203.

Hanley, JA. (1996). The use of binormal model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine*, 15:1575–1585.

Hanley, JA. and McNeil, BJ. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143:29–36.

Hanning, J., Iyer, HK., and Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101:254–269.

Harrell, FE. (2001). *Regression modelling strategies with applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.

Hastie, T., Tibshirani, R., and Friedman, J. (2000). *Diagnostic tests*. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. Statistics with confidence 2nd Ed. Bristol: BMJ Books, 105-119.

Hastie, TJ. and Tibshirani, RJ. (1990). *Generalized additive models*. London: Chapman and Hall.

Hdez-Armas, J., Allende, A., Alarco, R., Panero, J., and Oton, C. (1982). ROC analysis in gammagraphic diagnosis of hepatic cirrhosis. *European Journal of Nuclear Medicine*, 7:417–418.

He, X. and Frey, EC. (2006). Three-class ROC analysis–the equal error utility assumption and the optimality of three-class ROC surface using the ideal observer. *IEEE Transactions on Medical Imaging*, 25(8):979–986.

Heagerty, PJ., Lumley, T., and Pepe, MS. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56:337–344.

Heagerty, PJ. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61:92–105.

Hillis, SL. and Metz, CE. (2012). An analytic expression for the binormal partial area under the ROC curve. *Academic Radiology*, 19:1491–1498.

Hilsenbeck, SG. and Clark, GM. (1996). Practical p-value adjustment for optimally selected cutpoints. *Statistics in Medicine*, 15:103–112.

Hirschfeld, G. and Zernikow, B. (2013a). Cut points for mild, moderate, and severe pain on the VAS for children and adolescents: What can be learned from 10 million ANOVAs? *Pain*, 154(12):2626–2632.

Hirschfeld, G. and Zernikow, B. (2013b). Variability of "optimal" cut points for mild, moderate, and severe pain: Neglected problems when comparing groups. *Pain*, 154(1):154–159.

Hlatky, MA., Mark, DB., Harrell, FE., Lee, KL., Califf, RRM., and Pryor, DB. (1987). Rethinking sensitivity and specificity. *American Journal of Cardiology*, 59:1195–1198.

Hoffman, RM., Clanon, DL., Littenberg, B., Frank, JJ., and Peirce, JC. (2000). Using the free-to-total prostate-specific antigen ratio to detect prostate cancer in men with nonspecific elevations of prostate-specific antigen levels. *Journal of General Internal Medicine*, 15:739–748.

Hosmer, DW. and Lemeshow, S. (2000). *Applied logistic regression*. Wiley-Interscience, New York, USA.

Hsiao, JK., Bartko, JJ., and Potter, WZ. (1989). Diagnosing diagnoses. Receiver Operating Characteristic methods and psychiatry. *Archives of General Psychiatry*, 46(7):664–667.

Irwin, RJ. and Irwin, TC. (2011). A principled approach to setting optimal diagnostic thresholds: Where ROC and indifference curves meet. *European Journal of Internal Medicine*, 22(3):230–234.

Itoh, M., Takahashi, K., Nishida, H., Sakagami, K., and Okubo, T. (1996). Estimation of the optimal cut off point in a new immunological faecal blood test in a corporate colorectal cancer screening programme. *Journal of Medical Screening*, 3(2):66–71.

Jaescheke, R., Guyatt, G., and Lijmer, J. (2002). *Diagnostic tests*. In: Guyatt G, Rennie D, eds. User's guides to the medical literature. Chicago: AMA Press,121-140.

Jiang, Y., Metz, CE., and Nishikawa, RM. (1996). A Receiver Operating Characteristic partial area index for highly sensitive diagnostic tests. *Radiology*, 201:745–750.

Jiménez-Valverde, A. (2012). Insights into the area under the Receiver Operating Characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21:498–507.

Jiménez-Valverde, A. (2014). Threshold-dependence as a desirable attribute for discrimination assessment: Implications for the evaluation of species distribution models. *Biodiversity Conservation*, 23:369–385.

John, AS. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47(4):522–532.

Johnson, KM. (1998). The two by two diagram: A graphical truth table. *Journal of Clinical Epidemiology*, 52(11):1073–1082.

Jund, J., Rabilloud, M., Wallon, M., and Ecochard, R. (2005). Methods to estimate the optimal threshold for normally or log-normally distributed biological tests. *Medical Decision Making*, 25(4):406–415.

Jung, K., Stephan, C., Lein, M., Henke, W., Schnorr, D., Brux, B., Schürenkämper, P., and Loening, SA. (1996). Analytical performance and clinical validity of two free prostate-specific antigen assays compared. *Clinical Chemistry*, 42:1026–1033.

Kaivanto, K. (2008). Maximization of the sum of sensitivity and specificity as a diagnostic cutpoint criterion. *Journal of Clinical Epidemiology*, 61(5):517–518.

Kay, EJ. and Knill-Jones, R. (1992). Variation in restorative treatment decisions: Application of Receiver Operating Characteristic curve (ROC) analysis. *Community Dentistry and Oral Epidemiology*, 20:113–117.

Kelly, MJ., Dunstan, FD., Lloyd, K., and Fone, DL. (2008). Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: A comparison of five different methods. *BMC Psychiatry*, 8:10.

Kerlikowske, K., Grady, D., Barclay, J., Sickles, EA., and Ernster, V. (1996). Likelihood ratios for moderns screening mammography. Risk of breast cancer based on age and mammographic interpretation. *Journal of the American Medical Association*, 276:39–43.

Khan, KS. and Chien, PF. (2001). Evaluation of a clinical test. I: Assessment of reliability. *BJOG: An International Journal of Obstetrics and Gynaecology*, 108(6):562–567.

Klotsche, J., Ferger, D., Pieper, L., Rehm, J., and Wittchen, HU. (2009). A novel nonparametric approach for estimating cut-offs in continuous risk indicators with application to diabetes epidemiology. *BMC Medical Research Methodology*, 9:63.

Knottnerus, JA. and Leffers, P. (1992). The influence of referral patterns on the characteristics of diagnostic tests. *Journal of Clinical Epidemiology*, 45:1143–1154.

Koepsell, TD. and Connell, FA. (2002). Measures of gain in certainty from a diagnostic test. *American Journal of Epidemiology*, 121(5):744–753.

Komori, O. and Eguchi, S. (2010). A boosting method for maximizing the partial area under the ROC curve. *BMC Bioinformatics*, 11:314.

Koopman, PAR. (1984). Confidence limits for the ratio of two binomial proportions. *Biometrics*, 40:513–517.

Kraemer, HC. (1992). Risk ratios, odds ratio, and the test QROC. In *Evaluating Medical Tests*, pages 103–113. Sage Publications, Beverly Hills, California.

Kraemer, HC., Periyakoil, VS., and Noda, A. (2002). Kappa coefficients in medical research. *Statistics in Medicine*, 21:2109–2129.

Kramar, A., Faraggi, D., Fortune, A., and Reiser, B. (2001). mROC: A computer program for combining tumour markers in predicting disease states. *Computer Methods and Programs in Biomedicine*, 166:199–207.

Krieg, AF., Abendroth, TW., and Bongiovanni, MB. (1986). When is a diagnostic test result positive? Decision tree models based on net utility and threshold. *Archives of Pathology & Laboratory Medicine*, 110(9):787–791.

Kristjansson, B., Hill, G., McDowell, I., and J., L. (1996). Optimal cut-points when screening for more than one disease state: An example from the Canadian study of health and aging. *Journal of Clinical Epidemiology*, 49(12):1423–1428.

Lado, MJ., Cadarso-Suárez, C., Roca-Pardiñas, J., and Tahoces, PG. (2006). Using generalized additive models for construction of nonlinear classifiers in computer-aided diagnosis systems. *IEEE Transactions on Information Technology in Biomedicine*, 10:246–253.

Lai, CY., Tian, L., and Schisterman, EF. (2012). Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. *Computational Statistics and Data Analysis*, 56:1103–1114.

Landis, JR. and Koch, GG. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13:183–212.

Lasko, TA., Bhagwat, JG., Zou, KH., and Ohno-Machado, L. (2005). The use of Receiver Operating Characteristics curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5):404–415.

Last, JM. (2001). *A dictionary of epidemiology*. New York: Oxford University Press.

Lausen, B. and Schumacher, M. (1996). Evaluating the effect of optimized cutoff values in the assessment of prognostic factors. *Computational Statistics & Data Analysis*, 21:307–326.

Le, CT. (2006). A solution for the most basic optimization problem associated with an ROC curve. *Statistical Methods in Medical Research*, 15:571–584.

Lewis, JD., Chuai, S., Nessel, L., Lichtenstein, GR., Aberra, FN., and Ellenberg, JH. (2008). Use of the noninvasive components of the Mayo score to assess clinical response in ulcerative colitis. *Inflammatory Bowel Diseases*, 14(12):1660–1666.

Li, KK., Harris, K., Hadi, S., and Chow, E. (2007). What should be the optimal cut points for mild, moderate, and severe pain? *Journal of Palliative Medicine*, 10(6):1338–1346.

Linn, S. and Grunau, PD. (2006). New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. *Epidemiologic Perspectives & Innovations*, 3(1):11.

Linnet, K. (1987). Comparison of quantitative diagnostic tests: Type I error, power, and sample size. *Statistics in Medicine*, 6:147–158.

Littenberg, B. and Moses, LE. (1993). Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. *Medical Decision Making*, 13(4):313–321.

Liu, C., Berry, PM., Dawson, TP., and Pearson, RG. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3):385–393.

Liu, X. (2012). Classification accuracy and cut point selection. *Statistics in Medicine*, 31(23):2676–2686.

Lloyd, CJ. (1998). Using smoothed Receiver Operating Characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, 93 (444):1356–1364.

Lloyd, CJ. and Yong, Z. (1999). Kernel estimators of the ROC curve are better than empirical. *Statistics & Probability Letters*, 44:221–228.

López de Ullibarri Galparsoro, I. and Pita Fernández, S. (1998). Curvas ROC. *Cadernos de Atención Primaria*, 5(4):229–235.

López-Ratón, M., Cadarso-Suárez, C., and Lado, MJ. (2012a). *Técnicas de estimación e inferencia de las curvas ROC*. Editorial Académica Española.

López-Ratón, M., Cadarso-Suárez, C., Molanes-López, EM., and Letón, E. (2015a). Confidence intervals for the symmetry point: An optimal cutpoint in continuous diagnostic tests. *Submitted*.

López-Ratón, M., Cadarso Suárez, C., Molanes-López, EM., and Letón, E. (2015b). ***GsymPoint**: Estimation of the generalized symmetry point, an optimal cutpoint in continuous diagnostic tests*. R package version 1.0.

López-Ratón, M., Carmen Cadarso-Suárez, C., Molanes-López, EM., and Letón, E. (2012b). Inference of the Symmetry point with different costs for the specificity and sensitivity. In *Proceedings of 27th International Workshop on Statistical Modelling*, pages 191–196. Ed.: Komárek, A. and Nagy, S.

López-Ratón, M., Carmen Cadarso-Suárez, C., Molanes-López, EM., and Letón, E. (2013). GsymPoint: An R package for estimating the Generalized Symmetry point as the optimal cutpoint in continuous diagnostic tests. In *Proceedings of 28th International Workshop on Statistical Modelling*, pages 663–667. Ed.: Muggeo, V.M.R., Capursi, V., Boscaino, G. and Lovison, G.

López-Ratón, M., Molanes-López, EM., Letón, E., and Cadarso-Suárez, C. (2015c). **GsymPoint**: An R package for constructing confidence intervals for the generalized symmetry point. *To be submitted*.

López-Ratón, M. and Rodríguez-Álvarez, MX. (2014). ***OptimalCutpoints****: Computing optimal cutpoints in diagnostic tests*. R package version 1.1-1.

López-Ratón, M., Rodríguez-Álvarez, MX., Cadarso-Suárez, C., and Gude-Sampedro, F. (2014). **OptimalCutpoints**: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8):1–36.

López-Ratón, M., Rodríguez-Girondo, M., Rodríguez-Álvarez, MX., Cadarso-Suárez, C., and Gude, F. (2015d). Application of generalized additive models to the evaluation of continuous markers for classification purposes. *International Journal of Statistics in Medical Research*, 4(3):296–305.

Luo, J. and Xiong, C. (2012). **DiagTest3Grp**: An R package for analyzing diagnostic tests with three ordinal groups. *Journal of Statistical Software*, 51(3):1–24.

Lusted, LB. (1960). Logical analysis in Roentgen diagnosis. *Radiology*, 74:178–193.

Lusted, LB. (1968). *Introduction to medical decision making*. Charles C. Thomas, Springfield, Illinois.

Lusted, LB. (1971). Decision-making studies in patient management. *The New England Journal of Medicine*, 171:1217–1219.

Lustres-Pérez, V., Rodríguez-Álvarez, MX., Pazos-Pata, M., Cadarso-Suárez, C., and Fernández-Pulpeiro, E. (2010). The application of Receiver Operating Characteristic (ROC) methodology in biological studies of marine resources: Sex determination of Paracentrotus lividus (Lamarck, 1816). *SORT*, 34:239–248.

Magder, LS. and Fix, AD. (2003). Optimal choice of a cut point for a quantitative diagnostic test performed for research purposes. *Journal of Clinical Epidemiology*, 56:956–962.

Maneesai, A. and Krittayaphong, R. (2007). N-terminal probrain natriuretic peptide (NT-proBNP) in the diagnosis of heart failure in patients with acute dyspnea. *Thai. Heart Journal*, 20(1).

Manel, S., Williams, H., and Ormerod, S. (2001). Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology*, 38:921–931.

Martín-Andrés, A. and Luna del Castillo, JD. (2004). *Bioestadística para las ciencias de la salud*. Norma-Capitel.

Martínez-Camblor, P. (2011). Nonparametric cutoff point estimation for diagnostic decisions with weighted errors. *Revista Colombiana de Estadística*, 34:133–146.

Mayer, D. (2004). *Essential evidence based medicine*. Cambridge: Cambridge University Press.

Mazumdar, M. and Glassman, JR. (2000). Categorizing a prognostic variable: Review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statistics in Medicine*, 19:113–132.

McClish, DK. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making*, 9:190–195.

McClish, DK. (2012). Evaluation of the accuracy of medical Tests in a region around the optimal point. *Academic Radiology*, 19(12):1484–1490.

McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. 2nd ed. London: Chapman and Hall.

McIntosh, MW. and Pepe, MS. (2002). Combining several screening tests: Optimality of the risk score. *Biometrics*, 58:657–664.

McLaughlin, T., Reaven, G., Abbasi, F., Lamendola, C., Saad, M., Waters, D., Simon, J., and Krauss, RM. (2005). Is there a simple way to identify insulin-resistant individuals at increased risk of cardiovascular disease? *The Americal Journal of Cardiology*, 96:399–404.

McNeill, BJ., Keeler, E., and Adelstein, SJ. (1975). Primer on certain elements of medical decision making, with comments on analysis ROC. *The New England Journal of Medicine*, 293:211–215.

Metz, CE. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8:183–298.

Metz, CE. (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, 21:720–733.

Metz, CE. (1989). Some practical issues of experimental designs and data analysis in radiologic ROC studies. *Investigative Radiology*, 24:234–245.

Metz, CE., Goodenough, DJ., and Rossmann, K. (1973). Evaluation of Receiver Operating Characteristic curve data in terms of information theory, with applications in radiography. *Radiology*, 109:297–303.

Metz, CE., Herman, BA., and Shen, JH. (1998). Maximum likelihood estimation of Receiver Operating Characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17:1033–1053.

Metz, CE., Starr, SJ., Lusted, LB., and Rossmann, K. (1975). Progress in evaluation of human observer visual detection performance using the ROC curve approach. In Raynaud, C. and Todd-Pokropek, A. E., editors, *Information processing in scintigraphy*, pages 420–436. Orsay, France: Commissariat à l'Energie Atomique, Département de Biologie, Service Hospitalier Frédéric Joliot.

Miller, R. and Siegmund, D. (1982). Maximally selected Chi square statistics. *Biometrics*, 38:1011–1016.

Miller, RG., Efron, B., Brown, BW., and Moses, LE. (1980). *Biostatistics casebook*. Wiley-Interscience, 3-18.

Mogulkoc, N., Brutsche, MH., Bishop, PW., Greaves, SM., Horrocks, AW., and Egan, JJ. (2001). Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. *American Journal of Respiratory and Critical Care Medicine*, 164:103–108.

Molanes-López, EM. and Letón, E. (2011). Inference of the Youden index and associated threshold using empirical likelihood for quantiles. *Statistics in Medicine*, 30:2467–2480.

Molanes-López, EM., Van Keilegom, I., and Veraverbeke, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, 36:413–432.

Molodianovitch, K., Faraggi, D., and Reiser, B. (2006). Comparing the areas under two correlated ROC curves: Parametric and non-parametric approaches. *Biometrical Journal*, 48(5):745–757.

Morrison, AS. (1992). *Screening in chronic disease*. Second edition. New York: Oxford University Press.

Murphy, DR., Byfield, D., McCarthy, P., Humphreys, K., Gregory, AA., and Rochon, R. (2006). Interexaminer reliability of the hip extension test for suspected impaired motor control of the lumbar spine. *Journal of Manipulative and Physiological Therapeutics*, 29(5):374–377.

Nakas, CT., Alonzo, TA., and Yiannoutsos, CT. (2010). Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Statistics in Medicine*, 29(28):2946–2955.

Nakas, CT., Dalrymple-Alford, JC., Anderson, TJ., and Alonzo, TA. (2013). Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening. *Statistics in Medicine*, 32(6):995–1003.

Navarro, JB., Domenech, JM., de la Osa, N., and Ezpeleta, L. (1998). El análisis de curvas ROC en estudios epidemiológicos de psicopatología infantil: Aplicación al cuestionario CBCL. *Anuario de Psicología*, 29(1):3–15.

Neyman, J. and Pearson, ES. (1933). On the problem of the most efficient tests of statistical hypothesis. *Philos Trans R Soc Lond A.*, 231:289–337.

Obuchowski, NA. (2005). ROC analysis. *American Journal of Roentgenology*, 184(2):364–372.

Peirce, CS. (1884). The numerical measure of the success of predictions. *Science*, 4:453–454.

Peng, CYJ. and So, TSH. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, 1(1):31–70.

Pepe, MS. (2004). *The statistical evaluation of medical tests for classification and prediction*. 1st ed. Oxford University Press, USA.

Perkins, NJ. and Schisterman, EF. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the Receiver Operating Characteristics Curve. *American Journal of Epidemiology*, 163:670–675.

Peterson, WW., Birdsall, TG., and Fox, WC. (1954). The theory of signal detectability. *Information Theory, Transactions of the IRE Professional Group on*, pages 171–212.

Pita-Fernández, S. and Pértega-Díaz, S. (2003). Pruebas diagnósticas: Sensibilidad y especificidad. *Atención Primaria*, 10:110–125.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Remaley, AT., Sampson, ML., DeLeo, JM., Remaley, NA., Farsi, BD., and Zweig, MH. (1999). Prevalence-value-accuracy plots: A new method for comparing diagnostic tests based on misclassification costs. *Clinical Chemistry*, 45:934–941.

Richardson, MD., Turner, A., Warnock, DW., and Llewellyn, PA. (1983). Computer-assisted rapid enzyme-linked immunosorbent assay (ELISA) in the serological diagnosis of aspergillosis. *Journal of Immunological Methods*, 56:201–207.

Riddle, DL. and Stratford, PW. (1999). Interpreting validity indexes for diagnostic tests: An illustration using the Berg balance test. *Physical Therapy*, 79:939–950.

Riegelman, RK. and Hirsch, RP. (1991). Discriminación diagnóstica de las pruebas. *Boletín de la Oficina Sanitaria de Panamá*, 111:536–547.

Robertson, A., Zweig, MH., and Van Steirteghem, AC. (1983). Evaluating the clinical efficacy of laboratory tests. *American Journal of Clinical Pathology*, 79(1):78–86.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, JC., and Müller, M. (2011). **pROC**: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77.

Rota, M. and Antolini, L. (2014). Finding the optimal cut-point for Gaussian and Gamma distributed biomarkers. *Computational Statistics and Data Analysis*, 69:1–14.

Rota, M., Antolini, L., and Valsecchi, MG. (2015). Empirical likelihood confidence regions for the evaluation of continuous-scale diagnostic tests in the presence of verification bias. *BMC Medical Research Methodology*, 15:24.

Rubin, DB. and Schenker, N. (1987). Logit-based interval estimation for binomial data using the Jeffreys prior. *Sociological Methodology*, 17:131–144.

Rust, KF. and Rao, JN. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3):381–397.

Rutter, CM. and Miglioretti, DL. (2003). Estimating the accuracy of psychological scales using longitudinal data. *Biostatistics*, 4:97–107.

Sabatine, MS., Morrow, DA., De Lemos, JA., Jarolim, P., and Braunwald, E. (2009). Detection of acute changes in circulating troponin in the setting of transient stress test-induced myocardial ischaemia using an ultrasensitive assay: Results from TIMI 35. *European Heart Journal*, 30:162–169.

Sackett, DL., Brian Haynes, R., and Tugwell, P. (1989). *Clinical epidemology: A basic science for clinical medicine*. Madrid: Díaz de Santos.

Sackett, DL., Haynes, RB., Guyatt, GH., and Tugwell, P. (1994). *Epidemiología clínica. Ciencia básica para la medicina clínica*. 2ª ed. Madrid: Editorial médica panamericana.

Sackett, DL., Haynes, RB., and Tugwell, P. (1985). *Clinical epidemiology: a basic science for clinical medicine*. Little, Brown and Company, Boston.

Sackett, DL., Strauss, SE., Richardson, WS., Rosenberg, W., and Haynes, RB. (2000). *Evidence-based medicine: How to practice and teach EBM*. London: Churchill-Livingstone.

Sáez, M., Cadarso-Suárez, C., and Figueiras, A. (2003). **np.OR**: An S-Plus function for pointwise nonparametric estimation of odds-ratios of continuous predictors. *Computer Methods and Programs in Biomedicine*, 71:175–179.

Sandler, G. (1980). The importance of the history in the medical clinic and the cost of unnecessary test. *American Heart Journal*, 100:928.

Santana-Mora, U., López-Ratón, M., Mora, MJ., Cadarso-Suárez, C., López-Cedrún, J., and Santana-Penín, U. (2014). Surface raw electromyography has a moderate discriminatory capacity for differentiating between healthy individuals and those with TMD: A diagnostic study. *Journal of Electromyography and Kinesiology*, 24:332–340.

Schäfer, H. (1989). Constructing a cut-off point for a quantitative diagnostic test. *Statistics in Medicine*, 8:1381–1391.

Schapire, RE., Singer, Y., and Singhal, A. (1998). Boosting and Rocchio applied to text filtering. *In Proceedings of ACM SIGIR*, pages 215–223.

Schisterman, EF., Faraggi, D., Reiser, B., and Hu, J. (2008). Youden index and the optimal threshold for markers with mass at zero. *Statistics in Medicine*, 27(2):297–315.

Schisterman, EF. and Perkins, N. (2007). Confidence intervals for the Youden index and corresponding optimal cut-point. *Communications in Statistics-Simulation and Computation*, 36(3):549–563.

Schisterman, EF., Perkins, NJ., Liu, A., and Bondell, H. (2005). Optimal cutpoint and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology*, 16:73–81.

Schuchard, RA. and Massof, RW. (1990). Decision making with clinical diagnostic tests in terms of statistical decision theory and information theory. *Applied Optics*, 29(10):1436–1444.

Schulgen, G., Lausen, B., Olsen, JH., and Schumacher, M. (1994). Outcome-oriented cutpoints in analysis of quantitative exposures. *American Journal of Epidemiology*, 140:172–184.

Schurman, JV., Friesen, CA., Andre, L., Welchert, E., Lavenbarg, T., Danda, CE., Cocjin, JT., and Hyman, PE. (2007). Diagnostic utility of the water load test in children with chronic abdominal pain. *Journal of Pediatric Gastroenterology and Nutrition*, 44(1):51–57.

Schwarz, PEH., Li, J., Reimann, M., Schutte, AE., Bergmann, A., Hanefeld, M., Bornstein, SR., J., S., Tuomilehto, J., and J., L. (2009). The finnish diabetes risk score is associated with insulin resistance and progression towards type 2 diabetes. *Journal of Clinical Endocrinology & Metabolism*, 94(3):920–926.

Seffinger, MA., Najm, WI., Mishra, SI., Adams, A., Dickerson, VM., Murphy, LS., and Reinsch, S. (2004). Reliability of spinal palpation for diagnosis of back and neck pain. *Spine*, 29 (19):413–425.

Shapiro, DE. (1999). The interpretation of diagnostic tests. *Statistical Methods in Medical Research*, 8:113–134.

Silman, AJ. and Macfarlane, GJ. (2002). *Epidemiological studies: A practical guide*. Cambridge, UK: Cambridge University Press.

Silva, LC. (1987). *Métodos estadísticos para la investigación epidemiológica. Seminario internacional de estadísticas en Euskadi*. Instituto Vasco de Estadísticas.

Simel, DL., Samsa, GP., and Matchar, DB. (1991). Likelihood ratios with confidence: Sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*, 44(8):763–770.

Simel, DL., Samsa, GP., and Matchar, DB. (1993). Likelihood ratios for continuous test results - making the clinician's job easier or harder? *Journal of Clinical Epidemiology*, 46:85–93.

Skaltsa, K., Jover, L., and Carrasco, JL. (2010). Estimation of the diagnostic threshold accounting for decision costs and sampling uncertainty. *Biometrical Journal*, 52:676–697.

Smith, RD. (1991). Evaluation of diagnostic tests. In Smith, RD., editor, *Veterinary Clinical Epidemiology*, pages 29–43. Butterworth-Heinemann, Stoneham, 3rd edition.

Smits, N., Smit, F., Cuijpers, P., and De Graaf, R. (2007). Using decision theory to derive optimal cut-off scores of screening instruments: An illustration explicating costs and benefits of mental health screening. *International Journal of Methods in Psychiatric Research*, 16(4):219–229.

Sorribas, A., March, J., and Voit, EO. (2000). Estimating age-related trends in cross-sectional studies using S-distributions. *Statistics in Medicine*, 19:697–713.

Sox, HC. (1986). Probability theory in the use of diagnostic tests. *Annals of Internal Medicine*, 104:60–66.

Sox, HC., Blatt, MA., Higgins, MC., and KI, M. (1988). *Medical decision making*. Boston, MA: Butterworth-Heinemann.

Steurer, J., Fischer, JE., Bachmann, LM., Koller, M., and ter Riet, G. (2002). Communicating accuracy of tests to general practitioners: A controlled study. *BMJ (Clinical Research ed.)*, 324:824–826.

Swets, JA. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, 14:109–121.

Swets, JA. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240:1285–1293.

Swets, JA. and Pickett, RM. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. Academic Press, New York.

Swets, JA., Pickett, RM., Whitehead, SF., Getty, DJ., Schnur, JA., Swets, JB., and Freeman, BA. (1979). Assessment of diagnostic technologies: Advanced measurement methods are trated in a study of computed tomography of the brain. *Science*, 205:753–760.

Swets, JA. and Swets, JB. (1979). ROC approach to cost/benefit analysis. In *Proceedings of the sixth IEEE conference on computer applications in radiology*, pages 203–206. Reprinted in Ripley KL, Murray A, eds.

Swets, JA., Tanner, WPJ., and Birdsall, TG. (1961). Decision processes in perception. *Psychological Review*, 68:301–340.

Taube, A. (1986). Sensitivity, specificity and predictive values: A graphical approach. *Statistics in Medicine*, 5:585–591.

Thomas, DR. and Grunkemeier, GL. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70:865–871.

Tietz, NW. (1986). *Textbook on clinical chemistry*. W.B. Saunders, Philadelphia.

Torpy, JM., Burke, AE., and Glass, RM. (2010). Postoperative infections. *JAMA*, 303(24):2544.

Tosteson, AAN. and Begg, CB. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8:204–215.

Toubert, ME., Guillet, J., Chiron, M., Meria, P., Role, C., Schlageter, MH., Francois, H., Borschneck, C., Nivelon, F., Desgrandchamps, F., Rastel, D., Cussenot, O., Teillac, P., Le Duc, A., and Najean, Y. (1996). Percentage of free serum prostate-specific antigen: A new tool in the early diagnosis of prostatic cancer. *European Journal of Cancer*, 32:2088–2093.

Turner, DA. (1978). An intuitive approach to Receiver Operating Characteristic curve analysis. *The Journal of Nuclear Medicine*, 19:213–220.

Van den Berghe, G., Wouters, P., Weekers, F., Verwaest, C., Bruyninckx, F., Schetz, M., Vlasselaers, D., Ferdinande, P., Lauwers, P., and Bouillon, R. (2001). Intensive insulin therapy in the critically ill patients. *The New England Journal of Medicine*, 345:1359–1367.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., and Storlie, C. (2012). ***SDMTools*** *- Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises*. R package version 1.1-20.

Venkatraman, ES. and Begg, CB. (1996). A distribution-free procedure for comparing Receiver Operating Characteristic curves from a paired experiment. *Biometrika*, 83:835–848.

Vermont, J., Bosson, JL., François, P., Robert, C., Rueff, A., and Demongeot, J. (1991). Strategies for graphical threshold determination. *Computer Methods and Programs in Biomedicine*, 35:141–150.

Vickers, AJ. and Elkin, EB. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26:565–574.

Vizard, AL., Anderson, GA., and Gasser, RB. (1990). Determination of the optimum cut-off value of a diagnostic test. *Preventive Veterinary Medicine*, 10(1-2):137–143.

Wald, A. and Wolfowitz, J. (1939). Confidence limits for continuous distribution functions. *The Annals of Mathematical Statistics*, 10:105–118.

Walsh, JS. (1997). Limitations to the robustness of binormal ROC curves: Effects of model misspecification and location of decision thresholds on bias, precision, size and power. *Statistics in Medicine*, 16 (6):669–679.

Walter, SD. (2005). The partial area under the summary ROC curve. *Statistics in Medicine*, 24(13):2025–2040.

Wand, MP. and Jones, MC. (1995). *Kernel smoothing*. London: Chapman and Hall.

Wang, B. and Qin, G. (2013). Empirical likelihood confidence regions for the evaluation of continuous-scale diagnostic tests in the presence of verification bias. *The Canadian Journal of Statistics*, 41:398–420.

Wang, MD. and Geisser, S. (2005). Optimal dichotomization of screening test variables. *Journal of Statistical Planning and Inference*, 131(1):191–206.

Webb, AR. (2002). *Statistical pattern recognition*. Second Edition, John Wiley & Sons, QinetiQ Ltd., Malvern, UK.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88:899–905.

Weerahandi, S. (1995). *Exact statistical methods for data analysis*. Springer-Verlag: New York.

Weinstein, MC. and Fineberg, HV. (1980). *Clinical decision analysis*. Philadelphia: W.B. Saunders Company.

Wheeler, TL., Blackhurst, DW., Dellinger, EH., and Ramsey, PS. (2007). Usage of spot urine protein-to-creatinine ratios in the evaluation of preeclampsia. *American Journal of Obstetrics and Gynecology*, 196:465.e1–465e4.

Williams, BA., Mandrekar, JN., Mandrekar, SJ., Cha, SS., and Furth, AF. (2006). Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes. , Technical Report Series 79. Department of Health Sciences Research Mayo Clinic, Rochester, Minnesota.

Wood, SN. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B Statistical Methodology*, 65:95–114.

Wood, SN. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99:673–686.

Wood, SN. (2006). *Generalized additive models, an introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC.

Woodward, M. (1999). *Epidemiology study design and data analysis*. Boca Raton, Florida, Chapman & Hall/CRC Press.

Yang, Li., Maie, A., Xiaofeng, Z., Yong, K., Uttan, S., Richard, C., David, E., Elliot, A., David, E., No-Hee, P., and David, T. W. (2004). Salivary transcriptome diagnostics for oral cancer detection. *Clinical Cancer Research*, 10:8442–8450.

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports*, 62(40):1432–1449.

Youden, WJ. (1950). Index for rating diagnostic tests. *Cancer*, 3:32–35.

Zelman, DC., Dukes, E., Brandenburg, N., Bostrom, A., and Gore, M. (2005). Identification of cut-points for mild, moderate and severe pain due to diabetic peripheral neuropathy. *Pain*, 115(1-2):29–36.

Zetterberg, H. (2006). Re: The inconsistency of "optimal" cutpoints obtained using two criteria based on the Receiver Operating Characteristic curve. *Americal Journal of Epidemiology*, 164(7):707–708.

Zhao, LP., Kristal, AR., and E., W. (2006). Estimating relative risk functions in case-control studies using a nonparametric logistic regression. *American Journal of Epidemiology*, 144:598–609.

Zhou, H. and Qin, G. (2013). Confidence intervals for the difference in paired Youden indices. *Pharmaceutical Statistics*, 12:17–27.

Zhou, W. and Jing, BY. (2003). Adjusted empirical likelihood method for quantiles. *Annals of the Institute of Statistical Mathematics*, 55:689–703.

Zhou, X-H., McClish, DK., and Obuchowski, NA. (2002). *Statistical methods in diagnostic medicine*. Wiley-Interscience.

Zou, KH., Hall, WJ., and Shapiro, DE. (1997). Smooth non-parametric Receiver Operating Characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, 16:2143–2156.

Zou, KH., Yu, CR., Liu, K., Carlsson, MO., and Cabrera, J. (2013). Optimal thresholds by maximizing or minimizing various metrics via ROC-type analysis. *Academic Radiology*, 20(7):807–815.

Zweig, MH. and Campbell, G. (1993). Receiver-Operating Characteristics (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39:561–577.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA