Revista Colombiana de Estadística
Junio 2011, volumen 34, no. 1, pp. 133 a 146

Nonparametric Cutoff Point Estimation for Diagnostic Decisions with Weighted Errors

Estimación no paramétrica del punto de corte asociado a una decisión diagnóstica con errores ponderados

Pablo Martínez-Camblor[1,2,a]

[1]CAIBER, Oficina de Investigación Biosanitaria, Oviedo, Spain
[2]Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain

Abstract

The study of diagnostic tests is a hot topic which has direct applications in biomedical sciences. Despite of the relevance, in a diagnostic process, of the threshold (or cutoff point) employed on the decision taken by the physician, the study and comparison of the accuracy among different diagnostic criterions has been the main field of study. In this paper, the authors are interested in the study of the involved cutoff point estimation in diagnostic tests with weighted errors. With this goal, a nonparametric smoothed utility function estimator is considered. The bootstrap and the asymptotic distributions for the related M-estimator are derived. Finally, the obtained results are applied to study the Procalcitonin level which determines whether a child within the Pediatric Intensive Care Unit (UCIP) has a virical sepsis.

***Key words*:** Kernel density estimator, Sensitivity, Specificity, Threshold, Utility function.

Resumen

El estudio de tests diagnósticos es un tema candente con aplicaciones directas en las ciencias biomédicas. Aunque en la práctica, a la hora de tomar una decisión, los clínicos deben fijar un valor umbral (o punto de corte) a pesar de la relevancia que este valor tiene, el estudio y la comparación de la calidad entre diferentes criterios diagnósticos ha sido el principal campo de estudio. En este trabajo, los autores están interesados en el estudio de la estimación del punto de corte involucrado en un test diagnóstico con errores ponderados. Con este objetivo, se considera un estimador suavizado para una función de utilidad. Se estudian las distribuciones *bootstrap* y asintóticas del M-estimador resultante. Finalmente, los resultados obtenidos son aplicados al estudio de los niveles de Procalcitonina que determinan si un niño ingresado en la Unidad de Cuidados Intensivos Pediátricos (UCIP) tiene infección vírica.

***Palabras clave*:** especificadas, estimador núcleo para la densidad, función de utilidad, sensibilidad, umbral.

[a]Biostatistics and Associate professor. E-mail: pablomc@ficyt.es

1. Introduction

Diagnostic methods play an important role in the medical attention. The estimation and comparison of the accuracy among different methods are the focus of a wide variety of studies (see, for example, Zhou, Obuchowski & McClish (2002) and references therein). The main goal in a diagnostic test is to determine whether one individual is ill (positive). With this purpose, usually, some physiologic measure, T, is taken (as a marker) on a patient; the patient is classified as positive (with the illness) if this measure is upper (or lower) than a previously fixed threshold. This classification process has associated two possible mistakes –to classify a healthy individual in the positive group and to classify an unhealthy individual in the negative group. Of course, to determine the diagnostic test accuracy, these errors are basic. The proportion of positives which are correctly identified is known as *sensitivity* (S_E) and the proportion of negatives which are correctly identified is known as *specificity* (S_P).

The *Receiver Operating Characteristic* (ROC) curve (Green & Swets 1966) is a popular graphical method of displaying the discriminatory accuracy of a diagnostic test (based on a marker) for distinguishing between two populations. It is a plot of true-positive fraction (S_E) against the false-positive fraction ($1 - S_P$) over all possible threshold values of the considered marker. Although alternative indices have been discussed (see, for example, Lee & Hsiao (1998) or more recently Hand (2009)), the area under ROC curve (AUC) is, probably, the most commonly used index for diagnostic global accuracy. The ROC curve and the AUC index have been studied from different approaches (see Rodríguez-Álvarez, Tahoces, Cadarso-Suárez & Lado (2011) and Airola, Pahikkala, Waegeman, De Baets & Salakoski (2011) for some recent references). They have also been involved in the solution of different practical problems; for instance, recently, López-de Ulibarri, Cao, Cadarso-Suárez & Lado (2008) used a smooth estimation of the conditional ROC curve and the AUC on task discriminations and Martínez-Camblor & Yáñez-Juan (2009) developed a test to compare the equality of the diagnostic effectiveness of one measure with respect differents features based on the respective AUC values.

The Youden Index (Youden 1950) is also frequently used as accuracy measure. It is defined as $J = \max_{t\in\mathbb{R}}\{S_E(t)+S_P(t)-1\}$ and ranges between 0 and 1. Chin-Ying, Tian & Schisterman (2011) derived a procedure to build exact confidence interval estimations for the Youden index and its corresponding optimal cut-point. A vast study about the Youden Index and its associated cut-point estimations have been conducted by Fluss, Faraggi & Reiser (2005). They concluded that, in the estimation of the Youden Index the kernel is generally the best (among four considered estimators) unless the data can be well transformed to achieve normality whereas in estimation of the optimal threshold value results are more variable.

Most considered indices assume that the sensitivity and the specificity have the same relevance. However, to understand that there exist situations in which the impact of the two possible mistakes is quite different is easy. Taking into acount these differences and, for each $\lambda \in (0, 1)$, we introduce the following linear *utility* function (although in other context, it has been previously considered by

Krzanowski & Hand 2009)

$$U_\lambda(t) = \lambda S_E(t) + (1-\lambda) S_P(t) \tag{1}$$

Because the λ value (weight) determines the final impact of the sensitivity and specificity, its election is really important and, usually, depends on the costs of the different decisions and the prevalence of the illness which is being studied. Obviously, for each particular problem, its real value will be previously fixed by the specialist who must taking into count the different misclassification effects. Note that, if for $0 \leq \lambda \leq 1$ it is considered the optimum reachable utility, i.e.

$$J_\lambda = \max_{t \in \mathbb{R}} \{U_\lambda(t)\} \qquad \lambda \in (0,1) \tag{2}$$

then $J = 2(J_{1/2} - 1/2)$. Therefore, J_λ generalizes J when the mistakes in the classification process have different weights.

In this paper, smoothed estimators for the coefficient J_λ and its associated threshold are studied. In Section 2, the asymptotic and the bootstrap approximations for the cutoff point smoothed estimator are derived. Finally, in Section 3, we apply the proposed methods on the data set which motivated this research. On this data set, we study the procalcitonin (PCT) level which determines whether a child into the Pediatric Intensive Care Unit (UCIP) has a virical sepsis.

2. Nonparametric Cutoff Point Estimation

Let T be a continuous marker, we can assume (without loss of generality) that an individual is classified within group E (positives) if $T > t$ and within group $\overline{E}$ ($\overline{E}$ denotes the complementary set of E) if $T \leq t$. Let F_N and f_N be the distribution and the density functions, respectively, of N_T (T in the negative population; without the characteristic), and let F_P and f_P be the distribution and the density functions, respectively, of P_T (T in the positive population; with the characteristic), we have the equalities

$$S_E(t) = \mathcal{P}\{T > t \mid \mathrm{E}\} = 1 - F_P(t) \tag{3}$$

$$S_P(t) = \mathcal{P}\left\{T \leq t \mid \overline{\mathrm{E}}\right\} = F_N(t) \tag{4}$$

As usual, to estimate S_E and S_P we must estimate the distribution functions involved in the above definitions. Following the conclusions obtained by Fluss et al. (2005), we employ the kernel estimator and put the respective Smoothed Empirical Cummulative Distribution Functions (SECDF) instead of the theoretical ones to estimate the sensitivity and the specificity. Let $X = \{x_1, \ldots, x_n\}$ be a random sample from a continuous distribution F, the SECDF introduced by Nadaraya (1962) is defined as

$$\widetilde{F}_n(X,t) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{K}\left(\frac{t - x_i}{h_n}\right)$$

where $\widetilde{K}$ is a kernel function, usually taken to be a continuous probability function, with continuous and symmetrical about zero first derivative and $\{h_n\}_{n\in\mathbb{N}}$ is a sequence of deterministic bandwidths. The properties of the kernel estimator and its related curves have been widely studied and there exists a vast literature about this topic (see, for example, Mugdadi & Ghebregiorgis (2005) or Liu & Yang (2008) and references therein). Under some regularity conditions over the theoretical distribution and the used kernel function (it is enough, although not necessary, that both functions have three bounded and continuous derivatives), the mean ($\mathbb{E}$) and the variance ($\mathbb{V}$) for the SECDF are

$$\mathbb{E}[\widetilde{F}_n(X,t)] = F(t) + (1/2)f'(t)h_n^2 + \mathcal{O}(h_n^3) \tag{5}$$

$$n\mathbb{V}[\widetilde{F}_n(X,t)] = F(t)(1-F(t)) - 2h_n f(t)\int v\widetilde{K}'(v)\widetilde{K}(v)dv + \mathcal{O}(h_n^2) \tag{6}$$

Let $X_P = \{x_{P_1},\ldots,x_{P_n}\}$ and $X_N = \{x_{N_1},\ldots,x_{N_m}\}$ be two random samples from the positive and the negative populations, respectively, the natural *smoothed estimators* for S_P and S_N are

$$\widetilde{S}_E(t) = 1 - \widetilde{F}_n(X_P,t) \tag{7}$$

$$\widetilde{S}_P(t) = \widetilde{F}_m(X_N,t) \tag{8}$$

In the same way, replacing the sensitivity and the specificity by the above estimators, it is obtained the smoothed estimator for the utility function defined in (1),

$$\widetilde{U}_\lambda(t) = \lambda\widetilde{S}_E(t) + (1-\lambda)\widetilde{S}_P(t) \qquad \lambda\in(0,1) \tag{9}$$

Finally, the estimator for the associated cutoff point which is one of focus of this research, is the M-statistic

$$\widetilde{\theta}_\lambda = \min\{\text{argmax}_{t\in\mathbb{R}}\{\widetilde{U}_\lambda(t)\}\} \qquad \lambda\in(0,1) \tag{10}$$

The following result proves the asymptotic normality for the statistic $\widetilde{\theta}_\lambda$ under quite general conditions on the theoretical underlying distribution and on the parameters involved in the estimator definition (kernel function and used bandwidth).

Theorem 1. *Let X_N and X_P be two independent random samples (both independient and identically distributed, iid) with size n and m, respectively. Let $\widetilde{F}_n(X_N,t)$ be and $\widetilde{F}_m(X_P,t)$ the respective Smoothed Empirical Cummulative Distribution Functions (SECDF). Under the following assumptions*

A_1. *The real distribution function have three bounded and continuous derivatives.*

A_2. *Used kernel, $\widetilde{K}$, is a symmetrical about zero function with three bounded and continuous derivatives and $\int x^2 d\widetilde{K}(x) = 1$.*

A_3. $\exists \lim_n \sqrt{nh_n/mh_m} = \lim_n \alpha_n = \alpha < \infty$.

A_4. $U''_\lambda(\theta_\lambda) \neq 0$.

then,

$$\sqrt{nh_n}\frac{\widetilde{\theta}_\lambda - \theta_\lambda}{V_\lambda} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0,1) \tag{11}$$

with

$$V_\lambda^2 = \frac{R(K)\left(\lambda^2 f_P(\theta_\lambda) + (1-\lambda)^2\alpha^2 f_N(\theta_\lambda)\right)}{\left(\lambda f_P'(\theta_\lambda) + (1-\lambda)f_N'(\theta_\lambda)\right)^2} \tag{12}$$

where for each real function, g, $R(g) = \int g^2(x)\,dx$.

As usual, the variance of the statistic $\widetilde{\theta}_\lambda$ depends on several theoretical and unknown parameters, in particular, on the density functions (and its first derivative) in the positive and negative populations evaluated at the real optimal cutoff point, θ_λ. These theoretical (unknown) parameters are replaced by their *natural estimators* (the smoothed ones in the present study) to compute confidence intervals for $\widetilde{\theta}_\lambda$ (plug-in method) or for conducting inference on the parameter.

The Kernel density estimator, introduced by Rosenblatt (1956), is the most popular and commonly used density function estimator. Let $X = \{x_1, \ldots, x_n\}$ be a random sample (iid), it is defined as

$$\widetilde{f}_n(X,t) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h_n}\right) \tag{13}$$

where $K = \widetilde{K}' = (\partial\widetilde{K}(t)/\partial t)$ is a kernel function and $\{h_n\}_{n\in\mathbb{N}}$ is a sequence of deterministic bandwidths. In this setting, the *natural estimator* for the first density function derivative is

$$\widetilde{f}_n'(X,t) = \frac{1}{nh_n^2}\sum_{i=1}^{n} K'\left(\frac{t-x_i}{h_n}\right) \tag{14}$$

The *bandwidth* selection for the kernel estimators was a very hot topic in the 80s and early 90s (and it is still the focus of several recent papers). Their optimal convergence rates were widely studied. Cao (1990), looking for the bandwidth which minimizes the mean integrated square error (MISE), proved that the optimum convergence ratio for the SECDF is $\mathcal{O}(n^{-1/3})$, $\mathcal{O}(n^{-1/5})$ for kernel density function estimator and $\mathcal{O}(n^{-1/7})$ for its first derivative.

Silverman (1978) proved that if the real density function, f, is continuous, the used kernel is a variation bounded function and the bandwidth, h_n, is such that $nh_n \to_n \infty$ and $h_n \to_n 0$, the kernel density estimator, $\widetilde{f}_n$ converges uniformly almost surely to the real density function, i.e. $\sup_{t\in\mathbb{R}} |\widetilde{f}_n(X,t) - f(t)| \to 0$ a.s. (almost surely). This result allows deriving the following theorem

Theorem 2. *Under the assumptions in Theorem 1 and if it is also satisfied that*

A_5. $\widetilde{U}_\lambda''(\widetilde{\theta}_\lambda) \neq 0$.

A_6. *All the used bandwidth have the previously written optimal convergence rates (i.e. $\mathcal{O}(n^{-1/3})$ for SECDF; $\mathcal{O}(n^{-1/5})$ for density estimator and $\mathcal{O}(n^{-1/7})$ for its first derivative).*

then

$$\sqrt{nh_n}\ \frac{\widetilde{\theta}_\lambda - \theta_\lambda}{V_{n,\lambda}} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0,1) \tag{15}$$

with

$$V^2_{n,\lambda} = \frac{R(K)\left(\lambda^2 \widetilde{f}_n(X_P, \widetilde{\theta}_\lambda) + (1-\lambda)^2\alpha^2 \widetilde{f}_m(X_N, \widetilde{\theta}_\lambda)\right)}{\left(\lambda \widetilde{f}'_n(X_P, \widetilde{\theta}_\lambda) + (1-\lambda)\widetilde{f}'_m(X_N, \widetilde{\theta}_\lambda)\right)^2} \tag{16}$$

The main disadvantage of the previous result lies on the variance denominator estimator. Kernel estimators depend on the bandwidth selection which must be made by the investigator. There exist several automatic methods with this goal but does not exist an optimal solution (in adittion, the optimal bandwidth usual changes with each particular problem: density estimation, inference, etc.) For some discussion about this topic see Martínez-Camblor & De Uña-Álvarez (2009). The involved parameters on the denominator of the variance can be close to zero and, hence, small changes in their estimations can produce big changes on the final result. Trying to avoid these problems, as usual, we propose to use a resampling plan. Because the studied marker, T, is continuous and the expressions of the studied estimators depend on local properties (derivability), the *Smoothed Bootstrap* procedure (Hall, DiCiccio & Romano 1989) seems the most appropriate. The proposed algorithm is:

$\boldsymbol{B_1}$. From positive (X_P) and negative (X_N) samples, and for a fixed, or a grid of λ values ($\lambda \in (0,1)$), compute the SECDF and estimate: *sensitivity, specificity* and *utility functions*. Also compute the optimal cutoff point (threshold), $\widetilde{\theta}_\lambda(X_P, X_N) = \widetilde{\theta}_\lambda$.

$\boldsymbol{B_2}$. Run B pairs of bootstrap samples (X^b_P, X^b_N for $1 \leq b \leq B$) with the same sample sizes than the original ones from the respective SECDFs. On each bootstrap sample, compute and estimate functions which appear in $\boldsymbol{B}_1$. Also obtain the values for $\widetilde{\theta}^b_\lambda = \widetilde{\theta}^b_\lambda(X^b_P, X^b_N)$ with $1 \leq b \leq B$.

$\boldsymbol{B_3}$. The distribution of $\widetilde{\theta}_\lambda$ (and the other involved statistics) is approximated by $\{\widetilde{\theta}^1_\lambda, \ldots, \widetilde{\theta}^B_\lambda\}$.

Since the differences among the different resampling methods to make confidence intervals are, generally, negligible, we used the simplest and, probably, the most often used one; the percentile method (Efron & Tibshirani 1993). This method assumes that for a unknown monotone increasing transformation for the studied parameter, $h(\theta_\lambda)$ (in the present case, $\lambda \in (0,1)$), it is hold that

$$h(\widetilde{\theta}_\lambda) - h(\theta_\lambda) \sim \mathcal{N}(0, \sigma^2_{h(\widetilde{\theta}_\lambda)})$$

From this approach, a simple approximation for a $(1-\alpha)$ confidence interval can be found as $(\widetilde{\theta}^{(\alpha/2)}_\lambda, \widetilde{\theta}^{(1-\alpha/2)}_\lambda)$, where $\widetilde{\theta}^b_\lambda$ ($b \in 1, \ldots, B$) is obtained from the algorithm above.

The main goal of this algorithm is to approximate the $\widetilde{\theta}_\lambda$ distribution but, analogously, it can also be used to approximate the distribution for the other involved parameters (sensitivity, specificity and utility functions).

Despite of the AUC is widely used to summarize the global classification accuracy of a diagnostic rule, it is fundamentally incoherent in terms of misclassification costs (Hand 2009). The J_λ index defined in (2) provides a oportunity to define a global index for the diagnostic test accuracy which takes into count the different misclassification cost. With this goal, for each measure μ, it is define, by

$$\text{AUJ} = \int_0^1 J_\lambda d\mu(\lambda) \tag{17}$$

Note that AUJ index ranges between 0 and $\mu([0,1])$. If the chosen weight, μ, is the traditional Lebesgue measure, the AUJ stands for the area under J_λ curve and it means that all possible values of the weights are considered to be equally plausible.

3. Real Data Analysis

Bacterial sepsis is an important cause of mortality and morbidity in critically ill child. A delayed diagnosis of this condition is associated with worse prognosis. However, early detection of bacterial sepsis is difficult because the first signs of this disease may be minimal or non specific. Moreover, critically ill children present signs of sepsis such as fever, tachycardia, hyperventilation, and leukocytosis even in the absence of infection.

The availability of a laboratory test to accurately and rapidly identify critically ill children with sepsis would be of great value to improve the outcome of these patients. Early detection of the absence of infection would decrease the number of children started on antibiotics, shorten the length of hospital stay, and lessen the potential for emergent of resistant bacteria.

Body response to bacterial sepsis involves the release of several mediators. Recently, PCT, one of these mediators, has been proposed as an earlier marker of bacterial sepsis in children. Moreover, PCT levels are related to the severity of infection, presenting higher levels among patients with more severe sepsis. However, there is still some debate about the best cutoff levels to differentiate a patient with sepsis from a patient without sepsis (Rey, Los Arcos, Concha, Medina, Prieto, Martínez-Camblor & Prieto 2007). To define PCT cutoff levels with the optimum sensitivity and specifity to diagnose a critically ill child with sepsis can be very useful. Our goal is to study the cutoff point for the procalcitonin levels which determine that a child has a sepsis. With this objective we used the previous results for different λ values. We used information from patients admitted to the Pediatric Intensive Care Unit at the Hospital Universitario Central de Asturias (HUCA) from August 2002 until September 2004.

The descriptive statistics showed in Table 1 suggests a strongly asymmetry in the distribution of the PCT levels in both positive and negative considerated

groups. We know (see, for example, Silverman 1986) that the performance of the smoothed estimators is better for symmetrical distributions. To improve the estimations, we make a logarithmic transformation on the PCT levels. Kernel density estimations for the logarithmic of the PCT levels and for the function $\max_{t\in\mathbb{R}}\{\widetilde{U}_\lambda(t)\}$ ($\lambda \in (0,1)$) are shown in Figure 1.

TABLE 1: Descriptive statistics (mean, standard deviation (SD), minimum (Min), percentiles 25 (P_{25}), 50 (P_{50}), 75 (P_{75}), maximum (Max) and sample size (N)) for the Procalcitonin levels in the different considerated groups.

	Mean	**SD**	**Min**	$\mathbf{P}_{25}$	$\mathbf{P}_{50}$	$\mathbf{P}_{75}$	**Max**	**N**
Positive Group	22.89	39.83	0.11	2.81	10.64	27.53	347.10	125
Negative Group	1.48	3.98	0.01	0.12	0.30	1.00	39.01	232
Totals	8.98	25.83	0.01	0.18	0.95	5.73	347.10	357

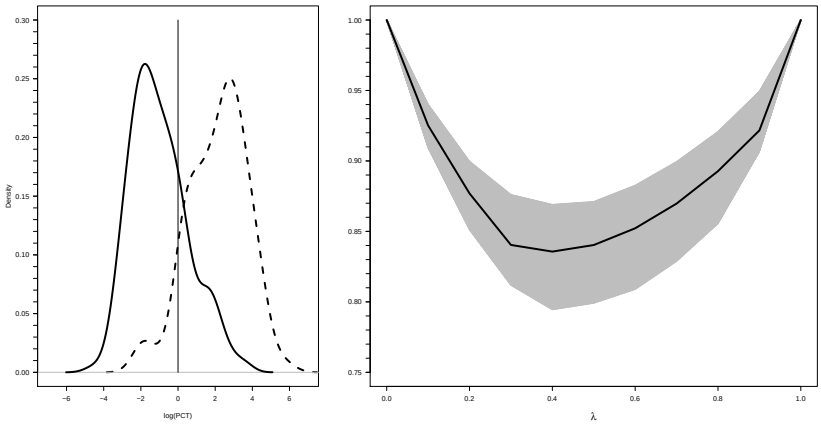

FIGURE 1: Kernel density estimations (left) for the logarithmic of the PCT levels in the positive (dotted line) and negative (continuous line) populations and the function $J_\lambda = \max_{t\in\mathbb{R}}\{\widetilde{U}_\lambda(t)\}$ with a 95% bootstrap confidence band (right).

Table 2 shows the obtained estimations for $\widetilde{\theta}_\lambda$, $\widetilde{U}_\lambda(\widetilde{\theta}_\lambda)$, $\widetilde{S}_E(\widetilde{\theta}_\lambda)$, $\widetilde{S}_P(\widetilde{\theta}_\lambda)$ and the square root for the asymptotic ($SD(\widetilde{\theta}_\lambda)$) and the bootstrap ($SD_B(\widetilde{\theta}_\lambda)$) (based on 10 000 Monte Carlo simulations) variance (SD) for $\widetilde{\theta}_\lambda$ for several λ values. For this data set, the asymptotic variance is, smaller than the bootstrap one. This fact suggests a slow speed for the asymptotic convergence. The value for the AUJ index when μ is the Lebesgue measure is 0.894 (really, the AUJ value represents the global utility of the particular diagnostic test when all the possible values of the weights are chosen to be equally plausible).

Figure 2 depicts 95% asymptotic and bootstrap confidence intervals (upper). In the lower plots, utility functions at the extremes of these confidence intervals are shown. The difference among the values is always quite small, which suggests robustness with respect to the chosen threshold.

TABLE 2: Values for $\widetilde{\theta}_\lambda$, $\widetilde{U}_\lambda(\widetilde{\theta}_\lambda)$, $\widetilde{S}_E(\widetilde{\theta}_\lambda)$, $\widetilde{S}_P(\widetilde{\theta}_\lambda)$, square root for asymptotic variance of $\widetilde{\theta}_\lambda$ ($SD(\widetilde{\theta}_\lambda)$) and bootstrap variance of $\widetilde{\theta}_\lambda$ ($SD_B(\widetilde{\theta}_\lambda)$) based on 10 000 Monte Carlo simulations for several λ values.

λ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\widetilde{\theta}_\lambda$	12.67	9.02	2.34	1.65	1.25	1.03	0.85	0.69	0.52
$\widetilde{U}_\lambda(\widetilde{\theta}_\lambda)$	0.92	0.88	0.84	0.83	0.84	0.85	0.87	0.89	0.92
$\widetilde{S}_E(\widetilde{\theta}_\lambda)$	0.44	0.53	0.77	0.34	0.88	0.91	0.93	0.94	0.95
$\widetilde{S}_P(\widetilde{\theta}_\lambda)$	0.98	0.96	0.87	0.83	0.80	0.76	0.72	0.69	0.63
$SD(\widetilde{\theta}_\lambda)$	8.69	3.37	0.73	0.26	0.12	0.09	0.08	0.07	0.05
$SD_B(\widetilde{\theta}_\lambda)$	9.39	2.28	2.11	0.66	0.26	0.18	0.13	0.11	0.19

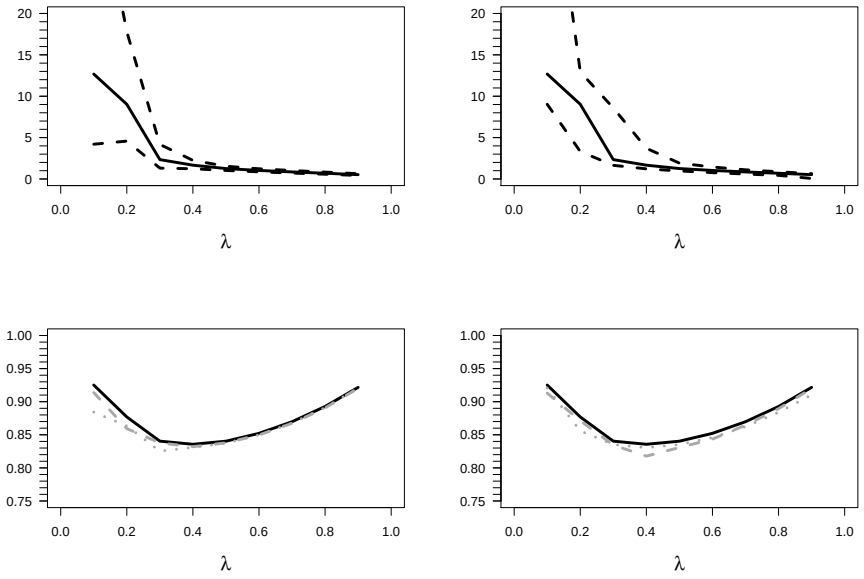

FIGURE 2: Upper, asymptotic (left) and bootstrap (right) 95% confidence intervals for the associated cutoff point estimation. Lower, utility function evaluated at the optimal estimated cutoff point (continuous lines) and at the extremes of the previous confidence intervals, asymptotic (left) and bootstrap (right) upper bound (grey dashed lines) and lower bound (grey dotted lines).

When sensitivity and specificity have the same relevance, both the AUC (0.913) and the Youden Index (0.680) suggest that the procalcitonin is a very good sepsis marker for the studied population. If different weights are assigned to S_E and S_P, in spite of the results are still quite goods, when we pay more attention to the sensitivity, the obtained utility is bigger. On the contrary the lower plots in the Figure 2 show that the final *gain* not changes when the cutoff points are within a reasonable interval.

4. Main Conclusions

Two important points of diagnostic medicine research, which are usually omitted, are the possible different impact of the two involved errors on the decision process and the effects on the final results in the variability of the associated cutoff point estimator. In this paper, we deal with the first problem introducing a linear *utility* function (obviously, most complex utility function could be considered with this goal) which allows study different weights for the sensitivity and the specificity. These weights must be previously chosen for the specialist which will take into count the different cost of the possible misclassification. The methods to cancer diagnostic tests are a special intersting field of application. There are continuous advance in this field with the aparition of new diagnostic markers (usually related with genes but which sensitivity and specificity are, generally, not large) and new (customized) drugs. The cost of the misclassification in this situation is usually different with great advantages for the early diagnostic.

We studied a nonparametric smoothed estimator for a linear utility function which allows to weight sensitivity and specificity and the corresponding associated cutoff point. We also derived its asymptotic distribution. In addition, the smoothed bootstrap procedure is considered. Because in the case of discrete markers all possible cutoff points could be studied and the researcher could chose among all the possibilities, we focus on continuos markers.

The obtained asymptotic variance for the threshold estimator is strongly depending on the first derivative of the density function. Because the convergence speed of the usual (kernel) estimator for this function is quite slow (see, for example, Silverman 1978), the use of the bootstrap approximation is advised when sample size is not large. To obtain adequate asymptotic conficence intervals, the required sample size depends on the variability and, in special, on the shape of the functions but, under simmetry, sizes around 100σ (σ^2 denotes the variance population) are advisables.

The effect that a little change on the used threshold produces on the final utility function is a specially interesting issue. In our analysis, this change seems to have a minor effect and the developed methods seem to be robust in this sense.

The AUC is a very widely used measure of performance for classification and diagnostic rules. It is mainly used in medicine and, recently, its use has been generalized to measure the accuracy in evaluating learning algorithms (see, for example, Huang & Ling (2005) and references therein). It has the appealing property of being objective, requiring no subjective input from the user but it is incoherent in terms of misclassification costs (Hand 2009). From the J_λ an coherent alternative (AUJ) to the AUC index (in cost terms) is also defined and studied.

Acknowledgements

The author is very grateful with Corsino Rey Galan and Marta Los Arcos Solas from the Hospital Universitario Central de Asturias (HUCA) for permission to use their data and for suggesting this research. The author is also grateful with the three anonymous referees whose suggestions and comments have really improved the paper.

[Recibido: junio de 2010 — Aceptado: diciembre de 2010]

References

Airola, A., Pahikkala, T., Waegeman, W., De Baets, B. & Salakoski, T. (2011), 'An experimental comparison of cross-validation techniques for estimating the area under the ROC curve', *Computational Statistics & Data Analysis* **55**(4), 1828–1844.

Cao, R. (1990), Aplicaciones y nuevos resultados del Método Bootstrap en la estimación no paramétrica de curvas, PhD thesis, University of Santiago de Compostela, Santiago de Compostela, Spain.

Chin-Ying, L., Tian, L. & Schisterman, E. F. (2011), 'Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point', *Computational Statistics & Data Analysis* . In Press, Corrected Proof. DOI: 10.1016/j.csda.2010.11.023.
*http://www.sciencedirect.com/science/article/B6V8V-51N223Y-1/2/9ac9b12dcddcf280e599a152375ca56c

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, London, United Kingdom.

Fluss, R., Faraggi, D. & Reiser, B. (2005), 'Estimation of the Youden index and its associated cutoff point', *Biometrical Journal* **47**(4), 458–472.

Green, D. M. & Swets, J. A. (1966), *Signal Detection Theory and Psychophysics*, Wiley, New York, United States.

Hall, P., DiCiccio, J. T. & Romano, J. P. (1989), 'On smoothing and the bootstrap', *Annals of Statistics* **17**, 692–702.

Hand, D. J. (2009), 'Measuring classifier performance: A coherent alternative to the area under the ROC curve', *Machine Learning - ML* **77**(1), 103–123.

Huang, J. & Ling, C. X. (2005), 'Using AUC and accuracy in evaluating learning algorithms', *IEEE Transactions on Knowledge and Data Engineering* **17**(3), 299–310.

Krzanowski, W. J. & Hand, D. J. (2009), *ROC Curves for Continuous Data*, Chapman and Hal, New York, United States.

Lee, W. C. & Hsiao, C. K. (1998), 'Alternative summary indices for the receiver operating characteristic curve', *Epidemiology* **7**, 605–611.

Liu, R. & Yang, L. (2008), 'Kernel estimation of multivariate cumulative distribution function', *Journal of Nonparametric Statistics* **20**(8), 661–667.

López-de Ulibarri, I., Cao, R., Cadarso-Suárez, C. & Lado, M. J. (2008), 'Nonparametric estimation of conditional ROC curves: Application to discrimination tasks in computarized detection of early breast cancer', *Computational Statistics & Data Analysis* **52**(5), 2623–2631.

Martínez-Camblor, P. & De Uña-Álvarez, J. (2009), Studying the bandwidth in k-sample smooth tests, Technical Report 09, Universidad de Vigo, Vigo, Spain.

Martínez-Camblor, P. & Yáñez-Juan, A. (2009), 'Testing the equality of diagnostic effectiveness of one measure with respect to k different features', *Journal of Applied Statistics* **36**(4), 359–367.

Mugdadi, A. R. & Ghebregiorgis, G. S. (2005), 'The kernel distribution estimator of functions of random variables', *Journal of Nonparametric Statistics* **17**(7), 807–818.

Nadaraya, E. A. (1962), 'Some new estimates for distribution functions', *Theory Probability Application* **9**, 497–500.

Rey, C., Los Arcos, M., Concha, A., Medina, A., Prieto, S., Martínez-Camblor, P. & Prieto, B. (2007), 'Procalcitonin and C-reactive protein as markers of systemic inflammatory response syndrome severity in critically ill children', *Intensive Care Medicine* **33**(3), 477–484.

Rodríguez-Álvarez, M. X., Tahoces, P. G., Cadarso-Suárez, C. & Lado, M. J. (2011), 'Comparative study of ROC regression techniques-applications for the computer-aided diagnostic system in breast cancer detection', *Computational Statistics & Data Analysis* **55**(1), 888–902.

Rosenblatt, M. (1956), 'Remarks on some nonparametric estimates of a density function', *Annals of Mathematical Statistics* **17**, 832–837.

Silverman, B. W. (1978), 'Weak and strong uniform consistency of the density estimation and its derivatives', *Annals of Statistics* **6**, 1177–1184.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, United States.

Youden, W. J. (1950), 'Index for rating diagnostic test', *Cancer* **3**, 32–35.

Zhou, X. H., Obuchowski, N. A. & McClish, D. K. (2002), *Statistical Methods in Diagnostic Medicine*, Wiley & Sons, New York, United States.

Appendix
Proof of the Results

Following, we deal with the proofs for the Theorems 1 and 2. Both demonstrations, quite similar, are based on the smoothed estimators and M-statistic properties and on the regularity conditions asked to the involved functions.

Proof. (Theorem 1) Conditions A_1 and A_2 guarantee the uniformly almost surely convergence for the kernel density estimator and its two first derivatives (Silverman 1978), therefore we can derive that $(\widetilde{U}_\lambda(\theta_\lambda) - U_\lambda(\theta_\lambda)) \to_P 0$.

$\widetilde{\theta}_\lambda = \text{argmax}\{\widetilde{U}_\lambda(t)\}$ and $\theta_\lambda = \text{argmax}\{U_\lambda(t)\}$, hence $\widetilde{U}'_\lambda(\widetilde{\theta}_\lambda) = 0 = U'_\lambda(\theta_\lambda)$. From the Theorem of the Mean Value, there exists ξ_λ between $\widetilde{\theta}_\lambda$ and θ_λ such that

$$\widetilde{U}'_\lambda(\theta_\lambda) - U'_\lambda(\theta_\lambda) = \widetilde{U}'_\lambda(\theta_\lambda) - \widetilde{U}'_\lambda(\widetilde{\theta}_\lambda) = \widetilde{U}''_\lambda(\xi_\lambda)(\theta_\lambda - \widetilde{\theta}_\lambda) \quad \lambda \in (0,1)$$

therefore $(\theta_\lambda - \widetilde{\theta}_\lambda) \to_P 0$.

Applying a three-term Taylor expansion on the first derivative of the utility function at point $\widetilde{\theta}_\lambda$, there exists η_λ between $\widetilde{\theta}_\lambda$ and θ_λ such that

$$\begin{aligned} 0 =& \widetilde{U}'_\lambda(\widetilde{\theta}_\lambda) = \widetilde{U}'_\lambda(\theta_\lambda + \widetilde{\theta}_\lambda - \theta_\lambda) \\ =& \widetilde{U}'_\lambda(\theta_\lambda) + \widetilde{U}''_\lambda(\theta_\lambda)(\widetilde{\theta}_\lambda - \theta_\lambda) + (1/2)\widetilde{U}'''_\lambda(\eta_\lambda)(\widetilde{\theta}_\lambda - \theta_\lambda)^2 \quad \lambda \in (0,1) \end{aligned}$$

then

$$\sqrt{nh_n}\,(\widetilde{\theta}_\lambda - \theta_\lambda) = -\frac{\sqrt{nh_n}\,\widetilde{U}'_\lambda(\theta_\lambda)}{\widetilde{U}''_\lambda(\theta_\lambda) + \frac{1}{2}\widetilde{U}'''_\lambda(\eta_\lambda)(\widetilde{\theta}_\lambda - \theta_\lambda)}$$

The A_2 assumption also implies that $\widetilde{U}'''_\lambda(t)$ is a bounded function $\forall t \in \mathbb{R}$, therefore $\widetilde{U}'''_\lambda(\eta_\lambda)(\widetilde{\theta}_\lambda - \theta_\lambda) \to_P 0$ for $\lambda \in (0,1)$. Kernel estimator convergence properties (cited at the beginning of the proof) imply $(\widetilde{U}''_\lambda(\theta_\lambda) - U''_\lambda(\theta_\lambda)) \to_P 0$, and then

$$\left(\sqrt{nh_n}(\widetilde{\theta}_\lambda - \theta_\lambda) + \frac{\sqrt{nh_n}U'_\lambda(\theta_\lambda)}{U''_\lambda(\theta_\lambda)}\right) \xrightarrow{P} 0$$

The Central Limit Theorem leads us to the convergence

$$\sqrt{nh_n}\,\widetilde{U}'_\lambda(\theta_\lambda) = \sqrt{nh_n}(\widetilde{U}'_\lambda(\theta_\lambda) - U'_\lambda(\theta_\lambda)) \xrightarrow{\mathcal{L}}_n \mathcal{N}(0, \sigma_\lambda)$$

with $\sigma^2_\lambda = R(K)\left(\lambda^2 f_P(\theta_\lambda) + (1-\lambda)^2\alpha^2 f_N(\theta_\lambda)\right)$.

The Slutski Lemma allows deducing, inmediately, that $\sqrt{nh_n}\,(\widetilde{\theta}_\lambda - \theta_\lambda)$ is asymptotically normal distributed with mean zero and variance

$$\frac{R(K)\left(\lambda^2 f_P(\theta_\lambda) + (1-\lambda)^2\alpha^2 f_N(\theta_\lambda)\right)}{\left(U''_\lambda(\theta_\lambda)\right)^2} = V^2_\lambda$$

□

Proof. To prove the Theorem 2 we only need to check that $(V^2_{n,\lambda} - V^2_\lambda) \to_P 0$. From the regularity assumptions (conditions A_1, A_2 and A_5) and the convergence rates of the used bandwidth (A_6), the already known kernel estimator convergence properties, we can write for $t, s \in \mathbb{R}$,

$$\begin{aligned}\widetilde{f}_n(X,t) =& f(s) + \mathcal{O}(t-s) + \mathcal{O}_P(n^{-2/5})\\ \widetilde{f}'_n(X,t) =& f'(s) + \mathcal{O}(t-s) + \mathcal{O}_P(n^{-2/7})\end{aligned}$$

Therefore

$$\begin{aligned}V^2_{n,\lambda} &= \frac{R(K)\left(\lambda^2\widetilde{f}_n(X_P,\widetilde{\theta}_\lambda) + (1-\lambda)^2\alpha_n^2\widetilde{f}_m(X_N,\widetilde{\theta}_\lambda)\right)}{\left(\lambda\widetilde{f}'_n(X_P,\widetilde{\theta}_\lambda) + (1-\lambda)\widetilde{f}'_m(X_N,\widetilde{\theta}_\lambda)\right)^2}\\ &= \frac{R(K)\left(\lambda^2\widetilde{f}_n(X_P,\widetilde{\theta}_\lambda) + (1-\lambda)^2\alpha_n^2\widetilde{f}_m(X_N,\widetilde{\theta}_\lambda)\right)}{\left(\lambda f'_P(\theta_\lambda) + (1-\lambda)f'_N(\theta_\lambda)\right)^2}\\ &\qquad + \mathcal{O}(\widetilde{\theta}_\lambda - \theta_\lambda) + \mathcal{O}_P(n^{-2/7}) + \mathcal{O}_P(m^{-2/7})\end{aligned}$$

therefore

$$\begin{aligned}(V^2_{n,\lambda} - V^2_\lambda) =& \\ &\frac{R(K)[\lambda^2(\widetilde{f}(X_P,\widetilde{\theta}_\lambda) - f_P(\theta_\lambda)) + (1-\lambda)^2\alpha_n^2(\widetilde{f}(X_N,\widetilde{\theta}_\lambda) - f_N(\theta_\lambda))]}{\left(\lambda f'_P(\theta_\lambda) + (1-\lambda)f'_N(\theta_\lambda)\right)^2}\\ &\qquad + \mathcal{O}(\widetilde{\theta}_\lambda - \theta_\lambda) + \mathcal{O}_P(n^{-2/7}) + \mathcal{O}_P(m^{-2/7})\\ &= \frac{R(K)[\lambda^2(\mathcal{O}(\widetilde{\theta}_\lambda - \theta_\lambda) + \mathcal{O}_P(n^{-2/5}))]}{\left(\lambda f'_P(\theta_\lambda) + (1-\lambda)f'_N(\theta_\lambda)\right)^2}\\ &\quad + \frac{R(K)[(1-\lambda)^2\alpha_n^2(\mathcal{O}(\widetilde{\theta}_\lambda - \theta_\lambda) + \mathcal{O}_P(m^{-2/5}))]}{\left(\lambda f'_P(\theta_\lambda) + (1-\lambda)f'_N(\theta_\lambda)\right)^2}\\ &\qquad + \mathcal{O}(\widetilde{\theta}_\lambda - \theta_\lambda) + \mathcal{O}_P(n^{-2/7}) + \mathcal{O}_P(m^{-2/7}) \to_P 0\end{aligned}$$ □