Three-group ROC predictive analysis for ordinal outcomes

Tahani Coolen-Maturi

Durham University Business School

Durham University, UK

tahani.maturi@durham.ac.uk

June 26, 2016

Abstract

Measuring the accuracy of diagnostic tests is crucial in many application areas including medicine, machine learning and credit scoring. The receiver operating characteristic (ROC) surface is a useful tool to assess the ability of a diagnostic test to discriminate among three ordered classes or groups. In this paper, nonparametric predictive inference (NPI) for three-group ROC analysis for ordinal outcomes is presented. NPI is a frequentist statistical method that is explicitly aimed at using few modelling assumptions, enabled through the use of lower and upper probabilities to quantify uncertainty. This paper also includes results on the volumes under the ROC surfaces and consideration of the choice of decision thresholds for the diagnosis. Two examples are provided to illustrate our method.

AMS Subject Classification: 60A99; 62G99; 62P10

Keywords: Accuracy of diagnostic tests; lower and upper probability; nonparametric predictive inference; ordinal data; ROC surface.

1 Introduction

Measuring the accuracy of diagnostic tests is crucial in many application areas including medicine, machine learning and credit scoring. The receiver operating characteristic (ROC) surface is a useful tool to assess the ability of a diagnostic test to discriminate among three ordered classes or groups. The construction of the ROC surface based on the probabilities of correct classification for three classes has been introduced by Mossman (1999), Nakas and Yiannoutsos (2004) and Nakas and Alonzo (2007). They also considered the volume under the ROC surface (VUS), and its relation to the probability of correctly ordered observations

from the three groups. The three-group ROC surface generalizes the popular two-group ROC curve, which in recent years has attracted much theoretical attention and has been widely applied for analysis of accuracy of diagnostic tests. For an overview of the current state of the art of ROC surface analysis and its applications, the reader is referred to Nakas (2014).

In this paper, we introduce nonparametric predictive inference (NPI) for three-group ROC analysis for ordinal outcomes. NPI is a frequentist statistical framework based only on few modelling assumptions, enabled by the use of lower and upper probabilities to quantify uncertainty (Augustin and Coolen, 2004; Coolen, 2006). In NPI, attention is restricted to one or more future observable random quantities, and Hill's assumption $A_{(n)}$ (Hill, 1968) is used to link these random quantities to data, in a way that is closely related to exchangeability (De Finetti, 1974). NPI has been introduced for assessing the accuracy of a classifier's ability to discriminate between two outcomes (or two groups) for binary data (Coolen-Maturi et al., 2012a) and for diagnostic tests with ordinal observations (Elkhafifi and Coolen, 2012) and with real-valued observations (Coolen-Maturi et al., 2012b). Recently, Coolen-Maturi et al. (2014) generalized the results in (Coolen-Maturi et al., 2012b) by introducing NPI for three-group ROC surface, with real-valued observations, to assess the ability of a diagnostic test to discriminate among three ordered classes or groups.

In this paper we generalize the results in (Elkhafifi and Coolen, 2012) by presenting NPI for three-group ROC surface with ordinal outcomes. In order to use NPI with ordinal data, we use an assumed underlying latent variable representation, with the categories represented by intervals on the real-line, reflecting the known ordering of the categories and enabling application of the assumption $A_{(n)}$ (Coolen et al., 2013; Elkhafifi and Coolen, 2012). The paper is organized as follows. Section 2 provides a brief introduction to NPI for ordinal data. Empirical three-group ROC analysis for ordinal outcomes is presented in Section 3. The main contribution of this paper, namely NPI for three-group ROC analysis for ordinal outcomes is introduced in Section 4. To illustrate our method, two examples are presented in Section 5. The paper ends with some concluding remarks in Section 6 and an appendix presenting the proofs of the main results.

2 Nonparametric predictive inference for ordinal data

2.1 Nonparametric predictive inference (NPI)

Nonparametric predictive inference (NPI) (Augustin and Coolen, 2004; Coolen, 2006) is based on the assumption $A_{(n)}$ proposed by Hill (1968). Let $X_1, \ldots, X_n, X_{n+1}$ be real-valued absolutely continuous and exchangeable random quantities. Let the ordered observed values of $X_1, X_2, \ldots, X_n$ be denoted by $x_1 < x_2 < \ldots < x_n$ and let $x_0 = -\infty$ and $x_{n+1} = \infty$ for ease of notation. For X_{n+1}, representing a future observation, $A_{(n)}$ (Hill,

1968) partially specifies a probability distribution by $P(X_{n+1} \in I_j = (x_{j-1}, x_j)) = \frac{1}{n+1}$ for $j = 1, \ldots, n+1$. $A_{(n)}$ does not assume anything else, and can be considered to be a post-data assumption related to exchangeability (De Finetti, 1974). Inferences based on $A_{(n)}$ are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides bounds for probabilities via the 'fundamental theorem of probability' (De Finetti, 1974), which are lower and upper probabilities in interval probability theory (Walley, 1991; Weichselberger, 2000). In NPI, uncertainty about the future observation X_{n+1} is quantified by lower and upper probabilities for events of interest. Lower and upper probabilities generalize classical ('precise') probabilities, and a lower (upper) probability for event A, denoted by $\underline{P}(A)$ ($\overline{P}(A)$), can be interpreted as the sharpest bounds on a probability for an event of interest when only $A_{(n)}$ is assumed. Informally, $\underline{P}(A)$ ($\overline{P}(A)$) can be considered to reflect the evidence in favour of (against) event A.

Augustin and Coolen (2004) proved that NPI has strong consistency properties in the theory of interval probability (Augustin et al., 2014; Walley, 1991; Weichselberger, 2000). Direct application of $A_{(n)}$ for inferential problems is only possible for real-valued random quantities. However, by using assumed latent variable representations and variations to $A_{(n)}$, NPI has been developed for different situations, including Bernoulli quantities (Coolen, 1998), non-ordered categorical data using so-called the 'circular-$A_{(n)}$' assumption (Coolen, 2006; Coolen and Augustin, 2009), and for ordered categorical data (Coolen et al., 2013). As this paper deals with ordered categorical data, a brief overview of NPI for ordinal data is given below (Coolen et al., 2013).

2.2 NPI for ordinal data

In situations with ordinal data, there are $K \geq 3$ categories to which observations belong, and these categories have a natural fixed ordering, hence they can be denoted by $C_1 < C_2 < \ldots < C_K$. It is attractive to base NPI for such data on the naturally related latent variable representation with the real-line partitioned into K categories, with the same ordering, and observations per category represented by corresponding values on the real-line and in the specific category. Assuming that multiple observations in a category are represented by different values in this latent variable representation, the assumption $A_{(n)}$ can be applied for the latent variables.

We assume that n observations are available, with only the number of observations in each category given. Let $n_r \geq 0$ be the number of observations in category C_r, for $r = 1, \ldots, K$, so $\sum_{r=1}^{K} n_r = n$. Let Y_{n+1} denote the random quantity representing the category a future observation will belong to. We wish to

derive the NPI lower and upper probabilities for events $Y_{n+1} \in \bigcup_{r \in R} C_r$ with $R \subset \{1, \ldots, K\}$.

Using the latent variable representation, we assume that category C_r is represented by interval IC_r, with the intervals $IC_1, \ldots, IC_K$ forming a partition of the real-line and logically ordered, that is interval IC_r has neighbouring intervals IC_{r-1} to its left and IC_{r+1} to its right on the real-line (or only one of these neighbours if $r = 1$ or $r = K$, of course). We further assume that the n observations are represented by $x_1 < \ldots < x_n$, of which n_r are in interval IC_r, these are also denoted by $x_{r,i}$ for $i = 1, \ldots, n_r$. A further latent variable X_{n+1} on the real-line corresponds to the future observation Y_{n+1}, so the event $Y_{n+1} \in C_r$ corresponds to the event $X_{n+1} \in IC_r$. This allows $A_{(n)}$ to be directly applied to X_{n+1}, and then transformed to inference on the categorical random quantity Y_{n+1}. The ordinal data structure for the latent variables is presented in Figure 1.
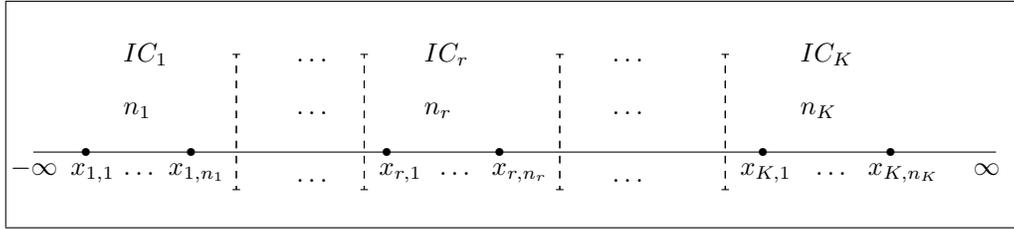

Figure 1: Ordinal data structure

Coolen et al. (2013) explain how the NPI lower and upper probabilities for general events of the form $Y_{n+1} \in \bigcup_{r \in R} C_r$ and $R \subset \{1, \ldots, K\}$, are calculated. In this paper we only need to consider the special case with the event $Y_{n+1} \in C_R$ where C_R consists of adjoining categories, so the corresponding union of intervals IC_R forms a single interval on the real line in the latent variable representation. For this case, simple closed forms for the NPI lower and upper probabilities are available (Coolen et al., 2013). Let $R = \{s, \ldots, t\}$, with $s, t \in \{1, \ldots, K\}$, $s \leq t$, excluding the case with $s = 1$ and $t = K$ for which both the NPI lower and upper probabilities are equal to 1. Let $C_{s,t} = \bigcup_{r=s}^{t} C_r$, $IC_{s,t} = \bigcup_{r=s}^{t} IC_r$ and let $n_{s,t} = \sum_{r=s}^{t} n_r$. Thus the NPI lower and upper probabilities for the event $Y_{n+1} \in C_{s,t}$ are (Coolen et al., 2013)

$$\underline{P}(Y_{n+1} \in C_{s,t}) = \underline{P}(X_{n+1} \in IC_{s,t}) = \begin{cases} \dfrac{(n_{s,t} - 1)^+}{n+1} & \text{if } 1 < s \leq t < K \\[3mm] \dfrac{n_{s,t}}{n+1} & \text{if } s = 1 \text{ or } t = K \end{cases} \tag{1}$$

$$\overline{P}(Y_{n+1} \in C_{s,t}) = \overline{P}(X_{n+1} \in IC_{s,t}) = \frac{n_{s,t} + 1}{n+1} \quad \text{for } 1 \leq s \leq t \leq K \tag{2}$$

where $(x)^+ = \max(x, 0)$, and the case $s = t$ gives the event that the next observation belongs to one specific

| | Diagnostic test result | | | | | | | |
Condition status	C_1	$\ldots$	C_{k_1}	$\ldots$	C_{k_2}	$\ldots$	C_K	Total
Y^0	n_1^0	$\ldots$	$n_{k_1}^0$	$\ldots$	$n_{k_2}^0$	$\ldots$	n_K^0	n^0
Y^1	n_1^1	$\ldots$	$n_{k_1}^1$	$\ldots$	$n_{k_2}^1$	$\ldots$	n_K^1	n^1
Y^2	n_1^2	$\ldots$	$n_{k_1}^2$	$\ldots$	$n_{k_2}^2$	$\ldots$	n_K^2	n^2
Total	n_1	$\ldots$	n_{k_1}	$\ldots$	n_{k_2}	$\ldots$	n_K	n

Table 1: Ordinal test data

category.

3 Empirical three-group ROC analysis for ordinal outcomes

We consider a diagnostic test with ordinal test results, where the test outcome for each individual indicates one of $K \geq 3$ ordered categories, denoted by C_1 to C_K and representing an increasing level of severity with regard to their indication of the presence of the condition of interest. We assume that the data available are on individuals in three ordered groups according to known condition status, e.g. mild, moderate and severe status, indicated by Y^0, Y^1 and Y^2 respectively[1]. The notation for the numbers of individuals for each combination of condition status and test result is given in Table 1. Throughout this paper we follow the definitions and notations introduced in Elkhafifi and Coolen (2012), Coolen-Maturi et al. (2012b), Coolen et al. (2013) and Coolen-Maturi et al. (2014).

We assume throughout this paper that there are two cut-off points (or thresholds) $k_1 < k_2$ in $\{1, ..., K\}$ such that a test result in categories $\{C_1, \ldots, C_{k_1}\}$ is interpreted as indication of the least severity of the condition "mild" condition, a test result in categories $\{C_{k_1+1}, \ldots, C_{k_2}\}$ as indication of the "moderate" condition, and a test result in categories $\{C_{k_2+1}, \ldots, C_K\}$ as indication of the "severe" condition. For a pair of thresholds (k_1, k_2), the probability of correct classification of a subject from group Y^0 is $p_0(k_1) = P(Y^0 \in \{C_1, \ldots, C_{k_1}\})$, the probability of correct classification of a subject from group Y^1 is $p_1(k_1, k_2) = P(Y^1 \in \{C_{k_1+1}, \ldots, C_{k_2}\})$, and the probability of correct classification of a subject from group Y^2 is $p_2(k_2) = P(Y^2 \in \{C_{k_2+1}, \ldots, C_K\})$. The ROC surface, denoted by ROCs, can be constructed by plotting these probabilities of correct classification $(p_0(k_1), p_1(k_1, k_2), p_2(k_2))$ for all $k_1 < k_2$ in $\{1, \ldots, K\}$. The probabilities of correct classification take values in $[0, 1]$ with corner coordinates $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. The empirical estimators of these probabilities $p_0(k_1)$, $p_1(k_1, k_2)$ and $p_2(k_2)$ (and hence the empirical estimator of ROCs, denoted by $\widehat{\text{ROCs}}$) are $\hat{p}_0(k_1) = \frac{1}{n^0} \sum_{j=1}^{k_1} n_j^0$, $\hat{p}_1(k_1, k_2) = \frac{1}{n^1} \sum_{j=k_1+1}^{k_2} n_j^1$, and $\hat{p}_2(k_2) = \frac{1}{n^2} \sum_{j=k_2+1}^{K} n_j^2$, respectively.

---

[1]Throughout this paper, the superscript notation ^{2} indicates group 3, there are no squared values used in this paper.

The volumes under the ROC surface (VUS) can be used as a global measure of the three-group discriminatory ability of the test under consideration. The empirical estimator of the volume under ROC surface (VUS) for ordinal data (as presented in Table 1) is given as

$$\widehat{VUS} = \frac{1}{n^0 n^1 n^2} \left[\sum_{i=1}^{K} \sum_{j=i+1}^{K} \sum_{l=j+1}^{K} n_i^0 n_j^1 n_l^2 + \frac{1}{2} \sum_{i=1}^{K} \sum_{j=i+1}^{K} n_i^0 (n_i^1 + n_j^1) n_j^2 + \frac{1}{6} \sum_{i=1}^{K} n_i^0 n_i^1 n_i^2 \right] \quad (3)$$

The volume under ROC surface $\widehat{VUS}$ can take values from 0 to 1. The $\widehat{VUS}$ value of about 1/6 would occur if the observations from the three groups would fully overlap, in such a way that the diagnostic method would perform no better than a random allocation of subjects to the three groups. If there is a perfect separation of the test results for the three groups, then $\widehat{VUS} = 1$.

The selection of the optimal cut-off points k_1 and k_2, is an important aspect of defining the diagnostic test and analysing its quality. One approach is Youden's index (Youden, 1950), which for three-group (continuous) diagnostic tests was introduced by Nakas et al. (2010). Similarly we can define Youden's index for ordinal three-group diagnostic tests as $J(k_1, k_2) = p_0(k_1) + p_1(k_1, k_2) + p_2(k_2)$. Using this index, the optimal cut-off points are the values of k_1 and k_2 which maximise $J(k_1, k_2)$. This index $J(k_1, k_2)$ is equal to 1 if the three groups fully overlap, while $J(k_1, k_2) = 3$ if the three groups are perfectly separated. The empirical estimator for $J(k_1, k_2)$ is obtained by replacing these probabilities by their corresponding empirical estimators,

$$\widehat{J}(k_1, k_2) = \frac{1}{n^0} \sum_{j=1}^{k_1} n_j^0 + \frac{1}{n^1} \sum_{j=k_1+1}^{k_2} n_j^1 + \frac{1}{n^2} \sum_{j=k_2+1}^{K} n_j^2 \quad (4)$$

4   NPI for three-group ROC analysis with ordinal outcomes

In this section the main results of this paper are presented. First, the NPI approach for three-group ROC analysis with ordinal outcomes is introduced and corresponding results for the volumes under the ROC surfaces and the Youden's index are derived. The notation required in this section was introduced in Sections 2 and 3, and we follow Coolen-Maturi et al. (2014) in the introduction of the NPI-based structures for the next observation from each of the three groups. Recall that for the latent variable representation, we assume that category C_r is represented by interval IC_r, with the intervals $IC_1, \ldots, IC_K$ forming a partition of the real-line and logically ordered. We further assume that the n^d observations are represented by $x_1^d < \ldots < x_{n^d}^d$, of which n_r^d are in the interval IC_r, $r = 1, \ldots, K$, these are also denoted by $x_{r,i}^d$ for $d = 0, 1, 2$ and $i = 1, \ldots, n_r^d$. A further latent variable $X_{n^d+1}^d$, for $d = 0, 1, 2$, on the real-line corresponds to the future observation $Y_{n^d+1}^d$, so the event $Y_{n^d+1}^d \in C_r$ corresponds to the event $X_{n^d+1}^d \in IC_r$. This allows $A_{(n^d)}$ to be directly applied to $X_{n^d+1}^d$, and then transformed to inference on the categorical random quantity

$Y_{n^d+1}^d$.

By using equations (1) and (2), we derive the NPI lower and upper bounds for the probabilities of correct classification as

$$\underline{p_0}(k_1) = \underline{P}(Y_{n^0+1}^0 \in \{C_1, \ldots, C_{k_1}\}) = \frac{1}{n^0+1} \sum_{j=1}^{k_1} n_j^0 \tag{5}$$

$$\overline{p_0}(k_1) = \overline{P}(Y_{n^0+1}^0 \in \{C_1, \ldots, C_{k_1}\}) = \frac{1}{n^0+1} \left(1 + \sum_{j=1}^{k_1} n_j^0\right) \tag{6}$$

$$\underline{p_1}(k_1, k_2) = \underline{P}(Y_{n^1+1}^1 \in \{C_{k_1+1}, \ldots, C_{k_2}\}) = \frac{1}{n^1+1} \left(-1 + \sum_{j=k_1+1}^{k_2} n_j^1\right)^+ \tag{7}$$

$$\overline{p_1}(k_1, k_2) = \overline{P}(Y_{n^1+1}^1 \in \{C_{k_1+1}, \ldots, C_{k_2}\}) = \frac{1}{n^1+1} \left(1 + \sum_{j=k_1+1}^{k_2} n_j^1\right) \tag{8}$$

$$\underline{p_2}(k_2) = \underline{P}(Y_{n^2+1}^2 \in \{C_{k_2+1}, \ldots, C_K\}) = \frac{1}{n^2+1} \sum_{j=k_2+1}^{K} n_j^2 \tag{9}$$

$$\overline{p_2}(k_2) = \overline{P}(Y_{n^2+1}^2 \in \{C_{k_2+1}, \ldots, C_K\}) = \frac{1}{n^2+1} \left(1 + \sum_{j=k_2+1}^{K} n_j^2\right) \tag{10}$$

4.1 Lower and upper envelopes of the set of NPI-based ROC surfaces

The sets of all probability distributions that correspond to the partial specifications, for $X_{n^0+1}^0$, $X_{n^1+1}^1$ and $X_{n^2+1}^2$, are the NPI-based structures and are denoted by $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$, respectively. For each combination of probability distributions for $X_{n^0+1}^0$, $X_{n^1+1}^1$ and $X_{n^2+1}^2$ in their respective NPI-based structures, $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$, the corresponding ROC surface as presented in Section 3 can be created. This will lead to a set of such NPI-based ROC surfaces, which we denote by $\mathbf{S_{roc}}$. The lower and upper envelopes of this set are of interest, they consist of the pointwise infima and suprema for this set, see Coolen-Maturi et al. (2014) for more details. These envelopes are presented below.

It is easy to show that the NPI lower ROC surface, $\underline{ROCs}^L$, goes through the points $\left\{ \left(p_0(k_1), \underline{p_1}(k_1, k_2), p_2(k_2) \right) : p_0(k_1) \in \left[\overline{p_0}(k_1) - \overline{p_0}(k_1 - 1) \right], p_2(k_2) \in \left[\underline{p_2}(k_2) - \underline{p_2}(k_2 + 1) \right], k_1 < k_2 \in \{1, \ldots, K\} \right\}$, where $\underline{p_1}(k_1, k_2)$ is obtained from (7). On the other hand, the NPI upper ROC surface, $\overline{ROCs}^U$, goes through the points $\left\{ \left(p_0(k_1), \overline{p_1}(k_1 - 1, k_2), p_2(k_2) \right) : p_0(k_1) \in \left[\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1) \right], p_2(k_2) \in \left[\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2) \right], k_1 < k_2 \in \{1, \ldots, K\} \right\}$, where $\overline{p_1}(k_1 - 1, k_2)$ is obtained from (8).

It is interesting to consider the volumes under these lower and upper envelopes, which we denote by $\underline{VUS}^L$ and $\overline{VUS}^U$, respectively. These are given in Theorem 1, the proof is presented in the appendix.

7

Theorem 1 *The volumes under the lower and upper envelopes of all NPI-based ROC surfaces in* $\mathbf{S_{roc}}$ *are*

$$\underline{VUS}^L = A \left\{ \sum_{i=1}^{K-2} \sum_{j=i+1}^{K-1} \sum_{l=j+1}^{K} n_i^0 n_j^1 n_l^2 - \sum_{i=1}^{K-2} \sum_{l=i+2}^{K} n_i^0 n_l^2 \right\} \tag{11}$$

$$\overline{VUS}^U = A \left\{ \sum_{i=1}^{K} \sum_{j=i}^{K} \sum_{l=j}^{K} n_i^0 n_j^1 n_l^2 + \sum_{i=1}^{K} \sum_{j=i}^{K} n_i^0 n_j^1 + \sum_{j=1}^{K} \sum_{l=j}^{K} n_j^1 n_l^2 + \sum_{i=1}^{K} \sum_{l=i}^{K} n_i^0 n_l^2 + n^0 + n^1 + n^2 + 1 \right\} \tag{12}$$

where $A = \frac{1}{(n^0+1)(n^1+1)(n^2+1)}$.

4.2 NPI lower and upper ROC surfaces

As the lower and upper envelopes of all ROC surfaces in $\mathbf{S_{roc}}$, result from pointwise optimisations they are too wide with regard to the set $\mathbf{S_{roc}}$ when the VUS values are considered. It should be emphasized that these envelopes are of interest as they characterize the set $\mathbf{S_{roc}}$ and can e.g. be used to graphically represent this set, as will be done in the example in Section 5. But it is also interesting to identify surfaces that provide thight bounds to all ROC surfaces in the set $\mathbf{S_{roc}}$ when the VUS values are considered, as these values play an important role for summarizing the quality of the diagnostic tests and for interpreting the ROC surfaces. So, we wish to define ROC surfaces with VUS values equal to the infimum and supremum of the VUS values for all ROC surfaces in $\mathbf{S_{roc}}$. The equality of the VUS and the probability of correctly ordered observations enables us to define lower and upper ROC surfaces in line with the optimization procedures described in the appendix to obtain $\underline{n}_j^1$ and $\overline{n}_j^1$. These lower and upper ROC surfaces are defined below.

It is easy to show that the NPI lower ROC surface, $\underline{ROCs}^E$, goes through the points $\left\{ \left(p_0(k_1), p_1^*(k_1, k_2), p_2(k_2) \right) : p_0(k_1) \in \left[\overline{p_0}(k_1) - \overline{p_0}(k_1 - 1) \right], p_2(k_2) \in \left[\underline{p_2}(k_2) - \underline{p_2}(k_2 + 1) \right], k_1 < k_2 \in \{1, \ldots, K\} \right\}$, where $p_1^*(k_1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1+1}^{k_2} \underline{n}_j^1$. On the other hand, the NPI upper ROC surface, $\overline{ROCs}^E$, goes through the points $\left\{ \left(p_0(k_1), p_1^{**}(k_1, k_2), p_2(k_2) \right) : p_0(k_1) \in \left[\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1) \right], p_2(k_2) \in \left[\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2) \right], k_1 < k_2 \in \{1, \ldots, K\} \right\}$, where $p_1^{**}(k_1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1}^{k_2} \overline{n}_j^1$.

The volumes under these NPI lower and upper surfaces, which we denote by $\underline{VUS}^E$ and $\overline{VUS}^E$, respectively, are given in Theorem 2. The proof is presented in the appendix.

Theorem 2 *The volumes under the NPI lower and upper ROC surfaces, which are equal to the NPI lower*

and upper probabilities for the event $(Y^0_{n^0+1} < Y^1_{n^1+1} < Y^2_{n^2+1})$, *respectively, are*

$$\underline{VUS}^E = A \sum_{i=1}^{K-2} \sum_{j=i+1}^{K-1} \sum_{l=j+1}^{K} n^0_i \underline{n}^1_j n^2_l \tag{13}$$

$$\overline{VUS}^E = A \left\{ \sum_{i=1}^{K} \sum_{j=i}^{K} \sum_{l=j}^{K} n^0_i \overline{n}^1_j n^2_l + \sum_{i=1}^{K} \sum_{j=i}^{K} n^0_i \overline{n}^1_j + \sum_{j=1}^{K} \sum_{l=j}^{K} \overline{n}^1_j n^2_l + \sum_{j=1}^{K} \overline{n}^1_j \right\} \tag{14}$$

where $A = \frac{1}{(n^0+1)(n^1+1)(n^2+1)}$. *Notice that* $\sum_{j=1}^{K} \underline{n}^1_j = \sum_{j=1}^{K} \overline{n}^1_j = n^1 + 1$.

4.3 Upper (lower) bound for the NPI lower (upper) ROC surface

One may want to avoid the numerical optimisations (especially for a large data set with a large number of categories) required to derive the NPI lower and upper ROC surfaces, in Section 4.2, by using the envelopes as approximations, benefiting from the fact that they are available in simple analytical expressions as given in Theorem 1. As the lower envelope provides a lower bound for the NPI lower ROC surface, it will be useful to be able to derive, also without numerical optimisations, an upper bound for this NPI lower ROC surface; together these two bounds will give some further information about the quality of the approximation. Similarly, it is of interest to derive a lower bound for the NPI upper ROC surface.

It is easy to show that the NPI lower ROC surface, $\underline{ROCs}^U$, goes through the points $\{ (p_0(k_1), \tilde{p}_1(k_1 + 1, k_2), p_2(k_2)) : p_0(k_1) \in [\overline{p_0}(k_1) - \overline{p_0}(k_1 - 1)], p_2(k_2) \in [\underline{p_2}(k_2) - \underline{p_2}(k_2 + 1)], k_1 < k_2 \in \{1, \dots, K\} \}$, where $\tilde{p}_1(k_1 + 1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1+1}^{k_2} n^1_j$. On the other hand, the NPI upper ROC surface, $\overline{ROCs}^L$, goes through the points $\{ (p_0(k_1), \tilde{p}_1(k_1, k_2), p_2(k_2)) : p_0(k_1) \in [\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1)], p_2(k_2) \in [\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2)], k_1 < k_2 \in \{1, \dots, K\} \}$, where $\tilde{p}_1(k_1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1}^{k_2} n^1_j$.

The volumes under these NPI lower and upper ROC surfaces are given by Theorem 3, the proof is presented in the appendix.

Theorem 3 *The volumes under the NPI lower and upper ROC surfaces,* $\underline{ROCs}^U$ *and* $\overline{ROCs}^L$, *are*

$$\underline{VUS}^U = A \sum_{i=1}^{K-2} \sum_{j=i+1}^{K-1} \sum_{l=j+1}^{K} n^0_i n^1_j n^2_l \tag{15}$$

$$\overline{VUS}^L = A \left\{ \sum_{i=1}^{K} \sum_{j=i}^{K} \sum_{l=j}^{K} n^0_i n^1_j n^2_l + \sum_{i=1}^{K} \sum_{j=i}^{K} n^0_i n^1_j + \sum_{j=1}^{K} \sum_{l=j}^{K} n^1_j n^2_l + \sum_{j=1}^{K} n^1_j \right\} \tag{16}$$

where $A = \frac{1}{(n^0+1)(n^1+1)(n^2+1)}$.

9

4.4 The NPI-based optimal decision thresholds

The choice of the decision thresholds k_1 and k_2 is an important aspect of designing the diagnostic method for the three groups case. One method is by maximisation of Youden's index as given in (4). The NPI lower and upper probabilities of correct classifications can be used to obtain the NPI lower and upper bounds for Youden's index. These are the sharpest possible bounds for all Youden's indices corresponding to probability distributions for $X^0_{n^0+1}$, $X^1_{n^1+1}$ and $X^2_{n^2+1}$ in their respective NPI-based structures $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$. The NPI lower and upper bounds for Youden's index are

$$\underline{J}(k_1, k_2) = \underline{p_0}(k_1) + \underline{p_1}(k_1, k_2) + \underline{p_2}(k_2)$$

$$= \frac{1}{n^0 + 1} \sum_{j=1}^{k_1} n_j^0 + \frac{1}{n^1 + 1} \left(-1 + \sum_{j=k_1+1}^{k_2} n_j^1 \right)^+ + \frac{1}{n^2 + 1} \sum_{j=k_2+1}^{K} n_j^2$$

$$\overline{J}(k_1, k_2) = \overline{p_0}(k_1) + \overline{p_1}(k_1, k_2) + \overline{p_2}(k_2)$$

$$= \frac{1}{n^0 + 1} \left(1 + \sum_{j=1}^{k_1} n_j^0 \right) + \frac{1}{n^1 + 1} \left(1 + \sum_{j=k_1+1}^{k_2} n_j^1 \right) + \frac{1}{n^2 + 1} \left(1 + \sum_{j=k_2+1}^{K} n_j^2 \right)$$

It is straightforward to show that, when $\sum_{j=k_1+1}^{k_2} n_j^1 = 0$,

$$\overline{J}(k_1, k_2) = \underline{J}(k_1, k_2) + \frac{1}{n^0 + 1} + \frac{1}{n^1 + 1} + \frac{1}{n^2 + 1}$$

and when $\sum_{j=k_1+1}^{k_2} n_j^1 > 0$,

$$\overline{J}(k_1, k_2) = \underline{J}(k_1, k_2) + \frac{1}{n^0 + 1} + \frac{2}{n^1 + 1} + \frac{1}{n^2 + 1}$$

this constant difference between the NPI upper and lower Youden's indices implies that both will be maximised at the same values of k_1 and k_2. It is further easy to show that, for all k_1 and k_2, $\underline{J}(k_1, k_2) \leq \widehat{J}(k_1, k_2) \leq \overline{J}(k_1, k_2)$, where $\widehat{J}(k_1, k_2)$ is the empirical estimate of Youden's index in (4). These inequalities do not imply that the empirical estimate of Youden's index is maximal for the same values of k_1 and k_2 as the NPI lower and upper Youden's indices, but we expect that in many situations the maxima will be attained as the same values, in particular for small K.

Scenario 1 ($K = 5, n = 100$)									
Y^0	54	28	15	2	1				
Y^1	3	26	44	25	2				
Y^2	1	2	11	26	60				
Scenario 1 ($K = 10, n = 1000$)									
Y^0	451	207	110	80	60	36	24	26	5	1
Y^1	1	20	70	150	244	242	169	83	20	1
Y^2	2	6	18	26	41	70	100	139	186	412
Scenario 2 ($K = 5, n = 100$)									
Y^0	22	36	34	7	1				
Y^1	4	30	39	25	2				
Y^2	1	7	30	47	15				
Scenario 2 ($K = 10, n = 1000$)									
Y^0	53	175	228	217	155	101	48	21	1	1
Y^1	1	18	63	180	229	226	182	81	20	0
Y^2	0	2	22	61	89	170	203	230	166	57

Table 2: Simulated data (Example 1)

5　Example

In this section, we illustrate our method using two examples, the first example based on simulated data from Beta distributions while the second example uses a real medical data set[2].

Example 1 This example is based on simulated data from Beta distributions. We consider two scenarios here, in the first scenario we assume weak overlap between the three ordered groups, while in the second scenario we assume that the three ordered groups considerably overlap. So, we expect that the volume under the ROC surface for the first scenario is close to 1, while for the second scenario it is close to 1/6. For each scenario we will consider two cases, ($K = 5, n = 100$) and ($K = 10, n = 1000$), where $n = n^0 = n^1 = n^2$. We use the cut-points 0.2(0.2)0.8 to categorize the simulated values from Beta distributions into $K = 5$ categories, and the cut-points 0.1(0.1)0.9 to categorize the simulated values into $K = 10$ categories.

	K	n	$\underline{VUS}^L$	$\underline{VUS}^E$	$\underline{VUS}^U$	$\widehat{VUS}_e$	$\overline{VUS}^L$	$\overline{VUS}^E$	$\overline{VUS}^U$
Scenario 1	5	100	0.5663	0.5677	0.5746	0.7256	0.8588	0.8682	0.8686
	10	1000	0.7054	0.7055	0.7063	0.7625	0.8175	0.8184	0.8185
Scenario 2	5	100	0.2239	0.2250	0.2285	0.4330	0.6642	0.6726	0.6735
	10	1000	0.4334	0.4337	0.4342	0.5416	0.6554	0.6561	0.6563

Table 3: Simulated data results (Example 1)

	K	n	k_1	k_2	$\underline{J}(k_1, k_2)$	$\widehat{J}(k_1, k_2)$	$\overline{J}(k_1, k_2)$
Scenario 1	5	100	2	3	2.089	2.120	2.129
	10	1000	3	7	2.307	2.310	2.311
Scenario 2	5	100	2	3	1.564	1.590	1.604
	10	1000	4	6	1.781	1.784	1.785

Table 4: Youden index results (Example 1)

---

[2]The R codes for implementing the proposed method will become available soon from http://npi-statistics.com

	Reverse Surgical Apgar Score (RevSAS)					
	C_1	C_2	C_3	C_4	C_5	Total
No morbidity	9 (90%)	149 (62%)	135 (56%)	20 (39%)	1 (17%)	314
Minor morbidity	1 (10%)	61 (25%)	71 (29%)	19 (37%)	1 (17%)	153
Major morbidity	0 (0%)	35 (14%)	35 (15%)	12 (24%)	4 (67%)	86
Total	10	245	241	51	6	553

Table 5: Thirty-day morbidity and mortality by revSAS (Example 2)

For the first scenario (weak overlap), we simulate from the Beta distributions $B(0.6, 2.5)$, $B(5, 5)$ and $B(2.5, 0.6)$ for groups Y^0, Y^1, and Y^2, respectively. For the second scenario (considerably overlap), we simulate from Beta distributions $B(5, 5)$, $B(2.5, 5)$ and $B(5, 2.5)$, for groups Y^0, Y^1, and Y^2, respectively. The simulated data sets are presented in Table 2. The empirical VUS and the NPI lower and upper bounds of VUS are provided in Table 3. We can see that $\underline{VUS}^L < \underline{VUS}^E < \underline{VUS}^U < \widehat{VUS_e} < \overline{VUS}^L < \overline{VUS}^E < \overline{VUS}^U$. And obviously these values are large especially for $K = 10$ and $n = 1000$. As expected the imprecision is lower when more data are available. Finally, the values of the cut-off points that maximize the Youden indices are given in Table 4, where $\underline{J}(k_1, k_2) < \widehat{J}(k_1, k_2) < \overline{J}(k_1, k_2)$. These cut-off points are the same whether we use the empirical Youden index or the NPI lower and upper bounds for the Youden index, which is often the case given that we have few categories, as discussed earlier in Section 4.4.

Example 2 In this example, a data set from Assifi et al. (2012) is used to illustrate the method presented in this paper. The data set consist of 553 patients undergoing a medical procedure over 10 years. Assifi et al. (2012) investigated whether the 10-point Surgical Apgar Score (SAS) accurately predicts the postoperative complications, such as major complications or death within 30 days. To match the presentation in this paper, we reverse the original SAS scores, denoted now by 'revSAS' scores, and we combine the last two classes in the original data set into one class as 'Major morbidity'. The revSAS scores are now grouped into five ordered categories as $C_1 : 9 - 10$, $C_2 : 7 - 8$, $C_3 : 5 - 6$, $C_4 : 3 - 4$ and $C_5 : 0 - 2$. This data set is presented in Table 5, from which we can see that as revSAS increases, the percentage of patients without morbidity decreases substantially. Likewise, as revSAS increases, the number of patients who had major morbidity increases.

The empirical estimators of the probabilities of correct classification, for $k_1 < k_2$ in $\{1, 2, 3, 4, 5\}$, and the NPI lower and upper bounds are given in Table 6. This table illustrates that $\underline{p_0}(k_1) \leq \hat{p}_0(k_1) \leq \overline{p_0}(k_1)$, $\underline{p_1}(k_1, k_2) \leq \hat{p}_1(k_1, k_2) \leq \overline{p_1}(k_1, k_2)$ and $\underline{p_2}(k_2) \leq \hat{p}_2(k_2) \leq \overline{p_2}(k_2)$. The empirical estimate of the volume under the ROC surface is $\widehat{VUS_e} = 0.2152$, and the NPI lower and upper bounds are $\underline{VUS}^L = 0.0536$, $\underline{VUS}^E = 0.0538$, $\underline{VUS}^U = 0.0544$, $\overline{VUS}^L = 0.4817$, $\overline{VUS}^E = 0.4854$, $\overline{VUS}^U = 0.4868$, so $\underline{VUS}^L < \underline{VUS}^E < \underline{VUS}^U < \widehat{VUS_e} < \overline{VUS}^L < \overline{VUS}^E < \overline{VUS}^U$. The NPI lower (upper) bound for the lower (upper) ROC surface is plotted in Figure 2 (Figure 3). The values of the cut-off points that maximize the Youden

(k_1,k_2)	$\underline{p_0}(k_1)$	$\hat{p}_0(k_1)$	$\overline{p_0}(k_1)$	$\underline{p_1}(k_1,k_2)$	$\hat{p}_1(k_1,k_2)$	$\overline{p_1}(k_1,k_2)$	$\underline{p_2}(k_2)$	$\hat{p}_2(k_2)$	$\overline{p_2}(k_2)$
$(1,2)$	0.02857	0.02866	0.03175	0.38961	0.39869	0.40260	0.58621	0.59302	0.59770
$(1,3)$	0.02857	0.02866	0.03175	0.85065	0.86275	0.86364	0.18391	0.18605	0.19540
$(1,4)$	0.02857	0.02866	0.03175	0.97403	0.98693	0.98701	0.04598	0.04651	0.05747
$(1,5)$	0.02857	0.02866	0.03175	0.98052	0.99346	0.99351	0.00000	0.00000	0.01149
$(2,3)$	0.50159	0.50318	0.50476	0.45455	0.46405	0.46753	0.18391	0.18605	0.19540
$(2,4)$	0.50159	0.50318	0.50476	0.57792	0.58824	0.59091	0.04598	0.04651	0.05747
$(2,5)$	0.50159	0.50318	0.50476	0.58442	0.59477	0.59740	0.00000	0.00000	0.01149
$(3,4)$	0.93016	0.93312	0.93333	0.11688	0.12418	0.12987	0.04598	0.04651	0.05747
$(3,5)$	0.93016	0.93312	0.93333	0.12338	0.13072	0.13636	0.00000	0.00000	0.01149
$(4,5)$	0.99365	0.99682	0.99683	0.00000	0.00654	0.01299	0.00000	0.00000	0.01149

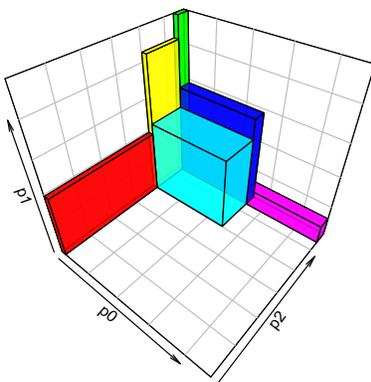Table 6: Probabilities of correct classification (Example 2)

Figure 2: The lower bound for the lower ROC surface (Example 2)

indices are $k_1 = 2$ and $k_2 = 3$, where $\underline{J}(k_1,k_2) = 1.1400$, $\widehat{J}(k_1,k_2) = 1.1533$, $\overline{J}(k_1,k_2) = 1.1677$. In this example the cut-off points are the same whether we use the empirical Youden index or the NPI lower and upper bounds for the Youden index, but this is not always the case as mentioned earlier in Section 4.4.

6 Concluding Remarks

In this paper we introduced the NPI approach for three-group ROC surfaces with ordinal outcomes. This can be used to asses the accuracy of a diagnostic test, with the NPI setting ensuring, due to its predictive nature, specific focus on the next patient. The NPI lower probability reflects the evidence in favour of the event of interest, while the NPI upper probability reflects the evidence against the event of interest. The difference between corresponding upper and lower probabilities reflects the amount of information available.

NPI typically leads to lower and upper probabilities for events of interest, which are based on Hill's
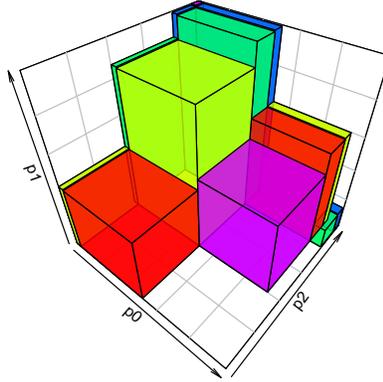
Figure 3: The upper bound for the upper ROC surface (Example 2)

assumption $A_{(n)}$ and have strong properties from frequentist statistics perspective. As events of interest are explicitly about a future observation, or a function of such an observation, NPI is indeed explicitly about prediction. The NPI lower and upper probabilities have a frequentist interpretation that could be regarded as 'confidence statements' related to repeated application of the same procedure. From this perspective, corresponding lower and upper probabilities can be interpreted as bounds for the confidence level for the predictive event of interest. However, the method does not provide prediction intervals in the classical sense, as e.g. appear in frequentist regression methods. Those tend to relate to confidence intervals for model parameter estimates combined with variability included in the model, in NPI no variability is explicitly included in a model and there are clearly no parameters to be estimated.

The concepts and ideas presented in this paper can be generalized to classification into more than three groups, but the computation of NPI lower and upper ROC hypersurfaces, in line with Section 4, will require numerical optimisations that will quickly become complicated for larger data sets with substantial overlap between the observations from different groups. Development of NPI-based methods for diagnostic accuracy with explicit focus on $m \geq 2$ future observations is an interesting topic for future research, where particularly the strength of the inferences as function of m should be studied carefully, see Coolen and Coolen-Schrijner (2007) for a similar study with focus on the role of m for comparison of groups of Bernoulli data.

Acknowledgement

Appendix

Proof [Proof of Theorem 1]

The proof consists of two parts, the first part is to prove that the NPI lower and upper probabilities for the event $Y_{n^0+1}^0 < Y_{n^1+1}^1 < Y_{n^2+1}^2$, are given by equations (11) and (12), respectively. The second part is to prove that the volumes under the NPI lower and upper ROC surfaces (envelopes) are actually given by (11) and (12), respectively.

First, in order to prove the NPI lower and upper probabilities for the event $Y_{n^0+1}^0 < Y_{n^1+1}^1 < Y_{n^2+1}^2$, we need to work with the latent variable representation, i.e. to find the lower and upper probabilities for the event $X_{n^0+1}^0 < X_{n^1+1}^1 < X_{n^2+1}^2$. Recall that, for the latent variable representation, we assume that category C_r is represented by interval IC_r, with the intervals $IC_1, \ldots, IC_K$ forming a partition of the real-line and logically ordered (Section 2.2). We further assume that the n^d observations are represented by $x_1^d < \ldots < x_{n^d}^d$, of which n_r^d are in the interval IC_r, $r = 1, \ldots, K$, these are also denoted by $x_{r,i}^d$ for $d = 0, 1, 2$ and $i = 1, \ldots, n_r^d$. A further latent variable $X_{n^d+1}^d$, for $d = 0, 1, 2$, on the real-line corresponds to the future observation $Y_{n^d+1}^d$, so the event $Y_{n^d+1}^d \in C_r$ corresponds to the event $X_{n^d+1}^d \in IC_r$. This allows $A_{(n^d)}$ to be directly applied to $X_{n^d+1}^d$, and then transformed to inference on the categorical random quantity $Y_{n^d+1}^d$.

To derive the NPI lower (upper) probability for this event, the probability mass $1/(n^0 + 1)$ for $X_{n^0+1}^0$, as assigned to each interval in the partition of the real-line created by the observations $x_1^0 < \ldots < x_{n^0}^0$, is put at the right-end (left-end) point of each interval. Simultaneously, the probability mass $1/(n^2 + 1)$ for $X_{n^2+1}^2$, as assigned to each interval in the partition of the real-line created by the observations from $x_1^2 < \ldots < x_{n^2}^2$, is put at the left-end (right-end) point of each interval. This leads to, for the lower

$$P \geq \frac{1}{(n^0+1)(n^2+1)} \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} P(x_i^0 < X_{n^1+1}^1 < x_{l-1}^2 | X_{n^1+1}^1 \in (x_{j-1}^1, x_j^1)) P(X_{n^1+1}^1 \in (x_{j-1}^1, x_j^1)) \quad (17)$$

and for the upper

$$P \leq \frac{1}{(n^0+1)(n^2+1)} \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} P(x_{i-1}^0 < X_{n^1+1}^1 < x_l^2 | X_{n^1+1}^1 \in (x_{j-1}^1, x_j^1)) P(X_{n^1+1}^1 \in (x_{j-1}^1, x_j^1)) \quad (18)$$

The data structure for the lower and upper probabilities, after this process, are visualized in Figures 4 and 5, respectively. The notation a_r and b_r in these figures will be introduced later in the proof of theorem 2.
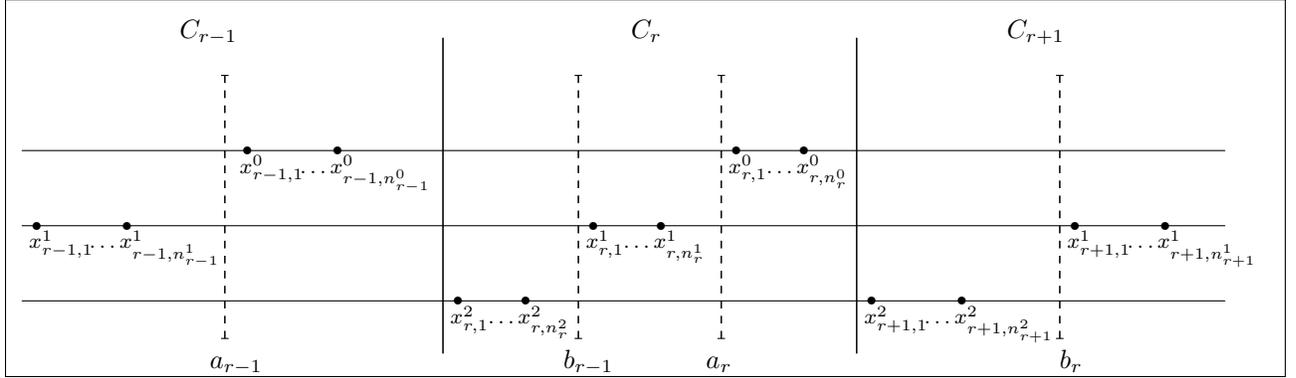

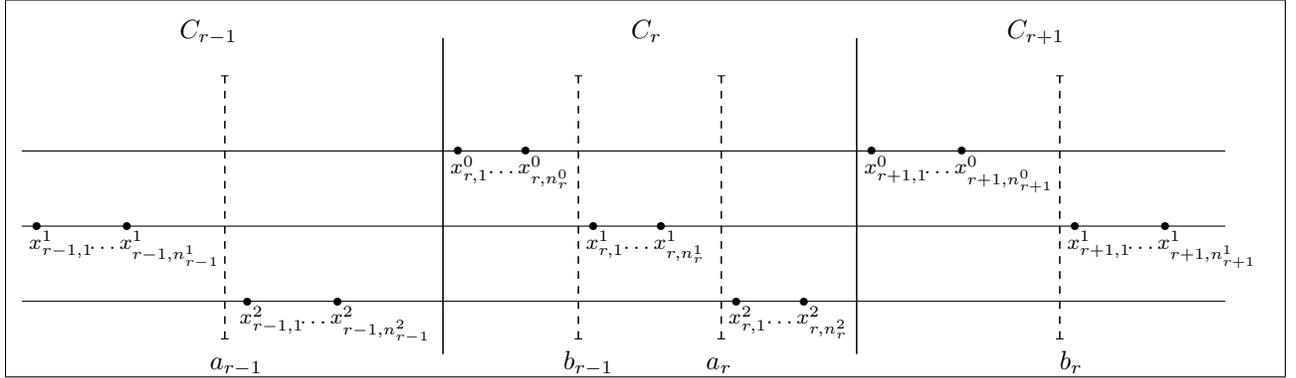Figure 4: Ordinal data structure (Lower probability)

Figure 5: Ordinal data structure (Upper probability)

The question now is how to assign the probability mass $1/(n^1 + 1)$ for $X^1_{n^1+1}$ within each interval (x^1_{j-1}, x^1_j), $j = 1, \ldots, n^1 + 1$, for the NPI lower and upper probabilities. We use the notation $x^1_0 = -\infty$ and $x^1_{n^1+1} = \infty$ for convenience.

For $k_1 < k_2$, the NPI lower probability in (11) is obtained by taking the NPI lower probability for the event that $X^1_{n^1+1}$ will be in the interval $(x^0_{k_1,n^0_{k_1}}, x^2_{k_2+1,1})$. This is can be achieved by counting the number of intervals (x^1_{j-1}, x^1_j) that are totally included in $(x^0_{k_1,n^0_{k_1}}, x^2_{k_2+1,1})$. The NPI upper probability in (12) is obtained by taking the NPI upper probability for the event that $X^1_{n^1+1}$ will be in the interval $(x^0_{k_1,n^0_{k_1}}, x^2_{k_2+1,1})$. This is can be achieved by counting the number of intervals (x^1_{j-1}, x^1_j) that have non-empty intersection with $(x^0_{k_1,n^0_{k_1}}, x^2_{k_2+1,1})$.

In the second part, we need to prove that the volumes under the NPI lower and upper ROC surfaces are actually given by (11) and (12), respectively. The volume under the NPI lower ROC surface consists of the volumes of $(K-2)(K-1)/2$ rectangular prisms, where the volume of the rectangular prism $V_{(k_1,k_2)}$ (for

$k_1 = 1, \ldots, K - 2$, and $k_2 = k_1 + 1, \ldots, K - 1$) is

$$[\overline{p_0}(k_1) - \overline{p_0}(k_1 - 1)]\, \underline{p_1}(k_1, k_2)\, [\underline{p_2}(k_2) - \underline{p_2}(k_2 + 1)] = \frac{n^0_{k_1}}{n^0 + 1} \left[\frac{1}{n^1 + 1} \left(-1 + \sum_{j=k_1+1}^{k_2} n^1_j \right)^+ \right] \frac{n^2_{k_2+1}}{n^2 + 1}$$

Summing the volumes of these rectangular prisms gives the volume under the NPI lower ROC surface as in equation (11), as follows.

$$\begin{aligned}
\underline{VUS}^L &= \sum_{k_1=1}^{K-2} \sum_{k_2=k_1+1}^{K-1} [\overline{p_0}(k_1) - \overline{p_0}(k_1 - 1)]\, \underline{p_1}(k_1, k_2)\, [\underline{p_2}(k_2) - \underline{p_2}(k_2 + 1)] \\
&= A \sum_{k_1=1}^{K-2} \sum_{k_2=k_1+1}^{K-1} n^0_{k_1} \left(-1 + \sum_{j=k_1+1}^{k_2} n^1_j \right)^+ n^2_{k_2+1} \\
&= A \left[\sum_{k_1=1}^{K-2} n^0_{k_1} \sum_{j=k_1+1}^{K-1} \sum_{k_2=j}^{K-1} n^1_j n^2_{k_2+1} - \sum_{k_1=1}^{K-2} n^0_{k_1} \sum_{k_2=k_1+1}^{K-1} n^2_{k_2+1} \right] \\
&= A \left[\sum_{i=1}^{K-2} n^0_i \sum_{j=i+1}^{K-1} \sum_{l=j+1}^{K} n^1_j n^2_l - \sum_{i=1}^{K-2} n^0_{k_1} \sum_{l=i+2}^{K} n^2_l \right]
\end{aligned}$$

The volume under the NPI upper ROC surface consists of the volumes of $K(K + 1)/2$ rectangular prisms, where the volume of the rectangular prism $V_{(k_1,k_2)}$ (for $k_1 = 1, \ldots, K$, and $k_2 = k_1, \ldots, K$) is

$$[\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1)]\, \overline{p_1}(k_1 - 1, k_2)\, [\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2)]$$

Summing the volumes of these rectangular prisms gives the volume under the NPI upper ROC surface as in equation (12), as follows.

$$\overline{VUS}^U = \sum_{k_1=1}^{K} \sum_{k_2=k_1}^{K} [\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1)]\, \overline{p_1}(k_1 - 1, k_2)\, [\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2)]$$

This double summation can be obtained in 4 steps, as follows.

(1) The volume of the rectangular prism $V_{(k_1,k_2)}$ (for $k_1 = 2, \ldots, K$, and $k_2 = k_1, \ldots, K - 1$) is

$$[\underline{p_0}(k_1) - \underline{p_0}(k_1 - 1)]\, \overline{p_1}(k_1 - 1, k_2)\, [\overline{p_2}(k_2 - 1) - \overline{p_2}(k_2)] = \frac{n^0_{k_1}}{n^0 + 1} \left[\frac{1}{n^1 + 1} \left(1 + \sum_{j=k_1}^{k_2} n^1_j \right) \right] \frac{n^2_{k_2}}{n^2 + 1}$$

Summing the volumes of these rectangular prisms gives,

$$A\left[\sum_{k_1=2}^{K}\sum_{k_2=k_1}^{K-1}\sum_{j=k_1}^{k_2}n_{k_1}^0 n_j^1 n_{k_2}^2 + \sum_{k_1=2}^{K}\sum_{k_2=k_1}^{K-1}n_{k_1}^0 n_{k_2}^2\right] = A\left[\sum_{i=2}^{K}\sum_{j=i}^{K-1}\sum_{l=j}^{K-1}n_i^0 n_j^1 n_l^2 + \sum_{i=2}^{K}\sum_{l=i}^{K-1}n_i^0 n_l^2\right] \tag{19}$$

(2) The volume of the $(k-1)$ rectangular prism $V_{(k_1,k_2)}$ (for $k_1 = 1$, and $k_2 = k_1,\ldots,K-1$) is

$$\frac{n_1^0+1}{n^0+1} \times \overline{p_1}(k_1-1,k_2)\left[\overline{p_2}(k_2-1) - \overline{p_2}(k_2)\right] = \frac{n_1^0+1}{n^0+1}\left[\frac{1}{n^1+1}\left(1+\sum_{j=k_1}^{k_2}n_j^1\right)\right]\frac{n_{k_2}^2}{n^2+1}$$

Summing the volumes of these rectangular prisms gives,

$$A(n_1^0+1)\left[\sum_{k_2=1}^{K-1}\sum_{j=1}^{k_2}n_j^1 n_{k_2}^2 + \sum_{k_2=1}^{K-1}n_{k_2}^2\right] = A(n_1^0+1)\left[\sum_{j=1}^{K-1}\sum_{l=j}^{K-1}n_j^1 n_l^2 + \sum_{l=1}^{K-1}n_l^2\right] \tag{20}$$

(3) The volume of the $(k-1)$ rectangular prism $V_{(k_1,k_2)}$ (for $k_1 = 2,\ldots,K$, and $k_2 = K$) is

$$\left[\underline{p_0}(k_1) - \underline{p_0}(k_1-1)\right]\overline{p_1}(k_1-1,k_2) \times \frac{n_K^2+1}{n^2+1} = \frac{n_{k_1}^0}{n^0+1}\left[\frac{1}{n^1+1}\left(1+\sum_{j=k_1}^{k_2}n_j^1\right)\right]\frac{n_K^2+1}{n^2+1}$$

Summing the volumes of these rectangular prisms gives,

$$A(n_K^2+1)\left[\sum_{k_1=2}^{K}\sum_{j=k_1}^{K}n_{k_1}^0 n_j^1 + \sum_{k_1=2}^{K}n_{k_1}^0\right] = A(n_K^2+1)\left[\sum_{i=2}^{K}\sum_{j=i}^{K}n_i^0 n_j^1 + \sum_{i=2}^{K}n_i^0\right] \tag{21}$$

(4) The volume of the rectangular prism $V_{(k_1,k_2)}$ (for $k_1 = 1$, and $k_2 = K$) is

$$\frac{n_1^0+1}{n^0+1} \times \overline{p_1}(k_1-1,k_2) \times \frac{n_K^2+1}{n^2+1} = \frac{n_1^0+1}{n^0+1}\left[\frac{1}{n^1+1}\left(1+\sum_{j=1}^{k_2}n_j^1\right)\right]\frac{n_K^2+1}{n^2+1}$$

$$= A\left[(n_1^0+1)(n_K^2+1)\sum_{j=1}^{K}n_j^1 + (n_1^0+1)(n_K^2+1)\right] \tag{22}$$

Summing the volumes of these rectangular prisms in (19), (20), (21) and (22) gives the volume under the NPI upper ROC surface in equation (12). $\qquad\square$

Proof [Proof of Theorem 2]

For the first part of the proof, we use the same setting as in the proof of Theorem 1, with regard to how the corresponding probability masses are assigned to the intervals for both $X_{n^0+1}^0$ and $X_{n^2+1}^2$. The question again is how to assign the probability mass $1/(n^1+1)$ for $X_{n^1+1}^1$ within each interval (x_{j-1}^1, x_j^1), $j = 1,\ldots,n^1+1$,

for the NPI lower and upper probabilities. We use the notation $x_0^1 = -\infty$ and $x_{n^1+1}^1 = \infty$ for convenience.

For the NPI lower probability, as we are dealing with ordinal data, it is easy to see that this optimization problem can be solved in three main steps: (1) by putting the probability mass $1/(n^1+1)$ in a single point within $(x_{r,j}^1, x_{r,j+1}^1)$, for $r = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, (n_r^1 - 1)$; (2) for the first and the last interval, $(-\infty, x_{1,1}^1)$ and $(x_{K,n_K^1}^1, \infty)$, the corresponding probability mass $1/(n^1+1)$ is assigned to a single point in the interval $(x_{1,n_1^2}^2, x_{1,1}^1)$ and $(x_{K,n_K^1}^1, x_{K,1}^0)$, respectively; (3) for the intervals between every two adjoining categories, that is for the intervals $(x_{r,n_r^1}^1, x_{r+1,1}^1)$, $r = 1, 2, \ldots, K-1$, the probability mass $1/(n^1+1)$ should be assigned to a single point in $(x_{r,n_r^1}^1, x_{r,1}^0)$ (represented as a_r in Figure 4) if $\sum_{i=1}^{r-1} \sum_{l=r+1}^K n_i^0 n_l^2 < \sum_{i=1}^r \sum_{l=r+2}^K n_i^0 n_l^2$, otherwise the probability mass $1/(n^1+1)$ should be assigned to a single point in $(x_{r+1,n_{r+1}^2}^2, x_{r+1,1}^1)$ (represented as b_r in Figure 4).

Once these optimization steps have been performed, we denote the points to which the probability masses for $X_{n^1+1}^1$ in the intervals (x_{j-1}^1, x_j^1) are assigned by $t_{\min}^j$, $j = 1, \ldots, n^1 + 1$, then equation (17) becomes

$$\underline{P}(X_{n^0+1}^0 < X_{n^1+1}^1 < X_{n^2+1}^2) = A \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} \mathbf{1}\{x_i^0 < t_{\min}^j < x_{l-1}^2\} \tag{23}$$

To transform back to the inference on the categorical random quantities, $Y_{n^0+1}^0, Y_{n^1+1}^1$ and $Y_{n^2+1}^2$, let $\underline{n}_r^1$ be the number of $t_{\min}^j$ ($j = 1, \ldots, n^1 + 1$) in category C_r, that is $\sum_{r=1}^K \underline{n}_r^1 = n^1 + 1$. Then it is easy to show that the lower probability in equation (23) is equivalent to

$$\underline{VUS}^E = \underline{P}(Y_{n^0+1}^0 < Y_{n^1+1}^1 < Y_{n^2+1}^2) = A \sum_{i=1}^{K-2} \sum_{j=i+1}^{K-1} \sum_{l=j+1}^K n_i^0 \underline{n}_j^1 n_l^2 \tag{24}$$

Now for the NPI upper probability, the optimization problem can be solved by the following three steps: (1) by putting the probability mass $1/(n^1+1)$ within $(x_{r,j}^1, x_{r,j+1}^1)$ in a single point, $r = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, (n_r^1 - 1)$; (2) for the first and last intervals, $(-\infty, x_{1,1}^1)$ and $(x_{K,n_K^1}^1, \infty)$, the corresponding probability mass $1/(n^1+1)$ is assigned to a single point in the interval $(x_{1,n_1^0}^0, x_{1,1}^1)$ and $(x_{K,n_K^1}^1, x_{K,1}^2)$, respectively; (3) for the intervals between every two adjoining categories, that is for the intervals $(x_{r,n_r^1}^1, x_{r+1,1}^1)$, $r = 1, 2, \ldots, K-1$, the probability mass $1/(n^1+1)$ should be assigned to a single point in $(x_{r,n_r^1}^1, x_{r,1}^2)$ (represented as a_r in Figure 5) if $\sum_{i=1}^{r-1} \sum_{l=r+1}^K n_i^0 n_l^2 > \sum_{i=1}^r \sum_{l=r+2}^K n_i^0 n_l^2$, otherwise the probability mass $1/(n^1+1)$ should be assigned to a single point in $(x_{r+1,n_{r+1}^0}^0, x_{r+1,1}^1)$ (represented as b_r in Figure 5).

Once these optimization steps have been performed, we denote the points to which the probability masses

for $X^1_{n^1+1}$ in the intervals (x^1_{j-1}, x^1_j) are assigned by $t^j_{\max}$, $j = 1, \ldots, n^1 + 1$. Then equation (18) becomes

$$\overline{P}(X^0_{n^0+1} < X^1_{n^1+1} < X^2_{n^2+1}) = A \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} \mathbf{1}\{x^0_{i-1} < t^j_{\max} < x^2_l\} \tag{25}$$

To transform back to the inference on the categorical random quantities, $Y^0_{n^0+1}, Y^1_{n^1+1}$ and $Y^2_{n^2+1}$, let $\overline{n}^1_r$ be the number of $t^j_{\max}$ $(j = 1, \ldots, n^1 + 1)$ in category C_r, that is $\sum_{r=1}^{K} \overline{n}^1_r = n^1 + 1$. Then it is easy to show that the upper probability in equation (25) is equal to

$$\overline{VUS}^E = \overline{P}(Y^0_{n^0+1} < Y^1_{n^1+1} < Y^2_{n^2+1}) = A \left\{ \sum_{i=1}^{K}\sum_{j=i}^{K}\sum_{l=j}^{K} n^0_i \overline{n}^1_j n^2_l + \sum_{i=1}^{K}\sum_{j=i}^{K} n^0_i \overline{n}^1_j + \sum_{j=1}^{K}\sum_{l=j}^{K} \overline{n}^1_j n^2_l + \sum_{j=1}^{K} \overline{n}^1_j \right\} \tag{26}$$

For the second part, we need to prove that the volumes under the NPI lower and upper ROC surfaces are actually given by (13) and (14), respectively. The proof is very similar to the proof of Theorem 1, by replacing $\underline{p}_1(k_1, k_2)$ and $\overline{p_1}(k_1 - 1, k_2)$ everywhere, in the proof of Theorem 1, by $p^*_1(k_1, k_2) = \frac{1}{n^1+1}\sum_{j=k_1+1}^{k_2} \underline{n}^1_j$ and $p^{**}_1(k_1, k_2) = \frac{1}{n^1+1}\sum_{j=k_1}^{k_2} \overline{n}^1_j$, respectively.

$\square$

Proof [Proof of Theorem 3]

For the first part of the proof, we use the same setting as in the proof of Theorem 1, with regard to how the corresponding probability masses are assigned to the intervals for both $X^0_{n^0+1}$ and $X^2_{n^2+1}$. Regarding $X^1_{n^1+1}$, for the NPI lower and upper probabilities, the probability mass $1/(n^1 + 1)$ for $X^1_{n^1+1}$ within each interval (x^1_{j-1}, x^1_j), $j = 1, \ldots, n^1 + 1$, is put at the left-end (or at the right-end) point of each interval. In this proof we will put the probability mass $1/(n^1 + 1)$ for $X^1_{n^1+1}$ at the left-end point (the same formulas will be obtained if the probability masses are put at the right-end points instead). Thus equation (17) can be written as

$$\underline{P}(X^0_{n^0+1} < X^1_{n^1+1} < X^2_{n^2+1}) = A \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} \mathbf{1}\{x^0_i < x^1_{j-1} < x^2_{l-1}\} \tag{27}$$

Then it is easy to show that the lower probability in equation (27) is equal to

$$\underline{VUS}^U = \underline{P}(Y^0_{n^0+1} < Y^1_{n^1+1} < Y^2_{n^2+1}) = A \sum_{i=1}^{K-2} \sum_{j=i+1}^{K-1} \sum_{l=j+1}^{K} n^0_i n^1_j n^2_l \tag{28}$$

Similarity, equation (18) can be written as

$$\overline{P}(X^0_{n^0+1} < X^1_{n^1+1} < X^2_{n^2+1}) = A \sum_{i=1}^{n^0+1} \sum_{j=1}^{n^1+1} \sum_{l=1}^{n^2+1} \mathbf{1}\{x^0_{i-1} < x^1_{j-1} < x^2_l\} \tag{29}$$

and it is easily shown that the upper probability in equation (29) is equal to

$$\overline{VUS}^L = \overline{P}(Y^0_{n^0+1} < Y^1_{n^1+1} < Y^2_{n^2+1}) = A \left\{ \sum_{i=1}^{K} \sum_{j=i}^{K} \sum_{l=j}^{K} n_i^0 n_j^1 n_l^2 + \sum_{i=1}^{K} \sum_{j=i}^{K} n_i^0 n_j^1 + \sum_{j=1}^{K} \sum_{l=j}^{K} n_j^1 n_l^2 + \sum_{j=1}^{K} n_j^1 \right\}$$
(30)

For the second part, we need to prove that the volumes under the NPI lower and upper ROC surfaces are actually given by (15) and (16), respectively. The proof is very similar to the proof of Theorem 1, by replacing $\underline{p_1}(k_1, k_2)$ and $\overline{p_1}(k_1 - 1, k_2)$ everywhere, in the proof of Theorem 1, by $\tilde{p}_1(k_1 + 1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1+1}^{k_2} n_j^1$ and $\tilde{p}_1(k_1, k_2) = (n^1 + 1)^{-1} \sum_{j=k_1}^{k_2} n_j^1$, respectively.

$\square$

References

Assifi, M., Lindenmeyer, J., Leiby, B., Grunwald, Z., Rosato, E., Kennedy, E., Yeo, C., and Berger, A. (2012). Surgical apgar score predicts perioperative morbidity in patients undergoing pancreaticoduodenectomy at a high-volume center. *Journal of Gastrointestinal Surgery*, 16(2):275–281.

Augustin, T. and Coolen, F. P. A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124(2):251–272.

Augustin, T., Coolen, F. P. A., de Cooman, G., and Troffaes, M. C. M. (2014). *Introduction to Imprecise Probabilities*. Wiley, Chichester.

Coolen, F., Coolen-Schrijner, P., Coolen-Maturi, T., and Elkhafifi, F. (2013). Nonparametric predictive inference for ordinal data. *Communications in Statistics-Theory and Methods*, 42(19):3478–3496.

Coolen, F. P. A. (1998). Low structure imprecise predictive inference for bayes' problem. *Statistics & Probability Letters*, 36(4):349–357.

Coolen, F. P. A. (2006). On nonparametric predictive inference and objective bayesianism. *Journal of Logic, Language and Information*, 15(1-2):21–47.

Coolen, F. P. A. and Augustin, T. (2009). A nonparametric predictive alternative to the imprecise dirichlet model: the case of a known number of categories. *International Journal of Approximate Reasoning*, 50(2):217–230.

Coolen, F. P. A. and Coolen-Schrijner, P. (2007). Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference*, 137(1):23–33.

Coolen-Maturi, T., Coolen-Schrijner, P., and Coolen, F. P. A. (2012a). Nonparametric predictive inference for binary diagnostic tests. *Journal of Statistical Theory and Practice*, 6(5):665–680.

Coolen-Maturi, T., Coolen-Schrijner, P., and Coolen, F. P. A. (2012b). Nonparametric predictive inference for diagnostic accuracy. *Journal of Statistical Planning and Inference*, 142(5):1141–1150.

Coolen-Maturi, T., Elkhafifi, F. F., and Coolen, F. P. (2014). Three-group roc analysis: A nonparametric predictive approach. *Computational Statistics & Data Analysis*, 78:69–81.

De Finetti, B. (1974). *Theory of Probability: A Critical Introductory Treatment*. Wiley, London.

Elkhafifi, F. F. and Coolen, F. P. A. (2012). Nonparametric predictive inference for accuracy of ordinal diagnostic tests. *Journal of Statistical Theory and Practice*, 6(4):681–697.

Hill, B. M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63(322):677–691.

Mossman, D. (1999). Three-way rocs. *Medical Decision Making*, 19:78–89.

Nakas, C. T. (2014). Developments in roc surface analysis and assessment of diagnostic markers in three-class classification problems. *REVSTAT Statistical Journal*, 12(1):43–65.

Nakas, C. T. and Alonzo, T. A. (2007). Roc graphs for assessing the ability of a diagnostic marker to detect three disease classes with an umbrella ordering. *Biometrics*, 63:603–609.

Nakas, C. T., Alonzo, T. A., and Yiannoutsos, C. T. (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, C. T. and Yiannoutsos, C. T. (2004). Ordered multiple-class roc analysis with continuous measurements. *Statistics in Medicine*, 23:3437–3449.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.

Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170.

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