



## Durham E-Theses

---

# *Nonparametric Predictive Inference for Ordinal Data and Accuracy of Diagnostic Tests*

ALI, FAIZA,FARAG

### How to cite:

---

ALI, FAIZA,FARAG (2012) *Nonparametric Predictive Inference for Ordinal Data and Accuracy of Diagnostic Tests*, Durham theses, Durham University. Available at Durham E-Theses Online:  
<http://etheses.dur.ac.uk/3914/>

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

---

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP  
e-mail: [e-theses.admin@dur.ac.uk](mailto:e-theses.admin@dur.ac.uk) Tel: +44 0191 334 6107  
<http://etheses.dur.ac.uk>

# Nonparametric Predictive Inference for Ordinal Data and Accuracy of Diagnostic Tests

**Faiza F. Elkhafifi**

A thesis presented for the degree of  
Doctor of Philosophy



Department of Mathematical Sciences  
University of Durham  
England

July 2012

*Dedicated to*

My parents, children, husband,  
my sister Mazoza and my brother Adal

# Nonparametric Predictive Inference for Ordinal Data and Accuracy of Diagnostic Tests

Faiza F. Elkhafifi

Submitted for the degree of Doctor of Philosophy  
July 2012

## Abstract

This thesis considers Nonparametric Predictive Inference (NPI) for ordinal data and accuracy of diagnostic tests. We introduce NPI for ordinal data, which are categorical data with an ordering of the categories. Such data occur in many application areas, for example medical and social studies. The method uses a latent variable representation of the observations and categories on the real line. Lower and upper probabilities for events involving the next observation are presented, with specific attention to comparison of multiple groups of ordinal data.

We introduce NPI for accuracy of diagnostic tests with ordinal outcomes, with the inferences based on data for a disease group and a non-disease group. We introduce empirical and NPI lower and upper Receiver Operating Characteristic (ROC) curves and the corresponding areas under the curves. We discuss the use of the Youden index related to the NPI lower and upper ROC curves in order to determine the optimal cut-off point for the test. Finally, we present NPI for assessment of accuracy of diagnostic tests involving three groups of real-valued data. This is achieved by developing NPI lower and upper ROC surfaces and the corresponding volumes under these surfaces, and we also consider the choice of cut-off points for classifications based on such diagnostic tests.

# Declaration

The work in this thesis is based on research carried out at the Statistics and Probability Research Group, Department of Mathematical Sciences, University of Durham, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

**Copyright © 2012 by Faiza F. Elkhaffi.**

“The copyright of this thesis rests with the author. No quotations from it should be published without the author’s prior written consent and information derived from it should be acknowledged”.

# Acknowledgements

First, Allah my God, I am truly grateful for your countless blessings you have bestowed on me generally and in accomplishing this thesis especially.

My main appreciation and thanks go to my supervisor Prof. Frank Coolen for his unlimited support, expert advice and guidance.

My beloved parents... words fail me... I am internally grateful your endless support not only in my education, but in my journey through life. I am truly fortunate to have your frequent loving prayers for me and the amount of unconditional love you surround me with.

My special thanks go to my husband Ali for his unwavering belief in me and unlimited support, thanks my lovely children Rawan and Mansour for lighting my life with your smile.

I would also like to thank Tahani for all her support, help and useful discussions. Thanks also to my colleagues and my friends especially, Radow, Amain and Rachel. I am indebted for your friendship.

Libya, I am thankful to you for the financial support that has awarded me the chance to complete my studies abroad.

Thanks to Durham University for offering such an enjoyable academic atmosphere and for the facilities that have enabled me to study smoothly.

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Declaration</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Introduction to literature</b>	<b>4</b>
2.1 Ordinal data . . . . .	4
2.2 Diagnostic accuracy . . . . .	6
2.3 Nonparametric Predictive Inference . . . . .	8
2.4 NPI for ordinal data . . . . .	10
2.4.1 Special cases . . . . .	13
<b>3 Multiple comparisons for ordinal data</b>	<b>17</b>
3.1 Introduction . . . . .	17
3.2 Selecting one group . . . . .	18
3.2.1 Empirical probabilities . . . . .	23
3.2.2 Examples . . . . .	24
3.2.3 Special cases . . . . .	28
3.2.4 Examples for special cases . . . . .	31
3.3 Subset containing all best groups . . . . .	36
3.4 Subset containing the best group . . . . .	44
3.5 Subset selection . . . . .	53
3.6 Maximum next observation . . . . .	58



---

3.7	Concluding remarks . . . . .	60
<b>4</b>	<b>Accuracy of ordinal diagnostic tests</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	Ordinal diagnostic tests . . . . .	62
4.2.1	ROC curve . . . . .	63
4.2.2	Area under the <i>ROC</i> curve ( <i>AUC</i> ) . . . . .	64
4.2.3	Youden's index . . . . .	66
4.3	Empirical lower and upper ROC curves . . . . .	67
4.4	NPI lower and upper ROC curves . . . . .	68
4.5	Area under the NPI lower or upper ROC curve . . . . .	70
4.6	Examples . . . . .	75
4.7	Concluding remarks . . . . .	79
<b>5</b>	<b>Three-group ROC analysis for continuous data</b>	<b>81</b>
5.1	Introduction . . . . .	81
5.2	ROC surface for three-group diagnostic tests . . . . .	82
5.2.1	Volume under the ROC Surface ( <i>VUS</i> ) . . . . .	84
5.2.2	Cut-off point selection in three-group classification . . . . .	85
5.3	NPI analysis for the ROC surface . . . . .	86
5.3.1	NPI lower and upper ROC surface . . . . .	87
5.3.2	Volumes under the NPI lower and upper ROC surfaces . . . . .	89
5.3.3	NPI cut-off points selection . . . . .	91
5.4	Examples . . . . .	93
5.5	Concluding remarks . . . . .	100
<b>6</b>	<b>Concluding remarks</b>	<b>102</b>
	<b>Bibliography</b>	<b>105</b>
	<b>Appendix</b>	<b>113</b>
<b>A</b>	<b>Multiple comparisons</b>	<b>113</b>
A.1	Proof of Theorem 3.2 . . . . .	113

---

A.2 Proof of Theorem 3.4 in case of four groups . . . . .	115
A.3 Proof of Theorem 3.5 in case of four groups . . . . .	118
A.4 Maximum next observation . . . . .	120

# Chapter 1

## Introduction

In recent years, nonparametric predictive inference (NPI) has been developed as a frequentist statistical framework which requires only few modelling assumptions and considers one or more future observations instead of a population. It is a statistical method based on Hill's assumption  $A_{(n)}$  [42]. NPI uses lower and upper probabilities for uncertainty quantification, and has strong consistency properties within theory of interval probability [20]. As NPI does not aim at inference for an entire population but instead explicitly considers one or more future observations, this provides an attractive alternative to standard methods of frequentist statistical inference.

In this thesis we introduce NPI for ordinal data and for accuracy of diagnostic tests. In many applications with categorical data, the categories are ordered, e.g. different levels of severity of a disease, in which case such data are also known as ordinal data [1]. It is important that such knowledge about ordering of categories is taken into account. The NPI method for such data uses 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  $A_{(n)}$ . We introduce NPI for multiple comparisons of groups of ordinal data.

The evaluation of the accuracy of a diagnostic test is crucial in many research fields such as medical imaging and machine learning. NPI for accuracy of diagnostic tests with ordinal data and real-valued data is presented and discussed.

In Chapter 2, we summarize some background literature relevant to the topics considered in this thesis. Some classical approaches for analysis of ordinal data are briefly reviewed. This is followed by some background information about the diagnostic accuracy of tests, which is relevant for Chapters 4 and 5. The general ideas of NPI are introduced and NPI for ordinal data is discussed.

In Chapter 3, we present NPI for multiple comparisons and subset selection for ordinal categorical data. For events which are of most practical interest, closed-form formulae for the NPI lower and upper probabilities are derived. Some special cases are discussed and illustrated with examples. The results in this chapter are also presented in the paper "Nonparametric predictive inference for ordinal data", which has been accepted for publication in *Communications in Statistics - Theory and Methods* [25]. This paper introduces NPI for ordinal data as presented in Chapter 2 of this thesis together with multiple comparisons which is the topic of Chapter 3 of this thesis.

Chapter 4 begins with a summary of the main concepts for ordinal diagnostic tests from the literature. We introduce NPI for accuracy of diagnostic tests with ordinal outcomes, with the inferences based on data for a disease group and a non-disease group. We introduce empirical and NPI lower and upper ROC curves and the corresponding areas under the curves. We discuss the use of the Youden index related to the NPI lower and upper ROC curves in order to determine the optimal cut-off point for the test. The results of this chapter have been presented at the 4th International Conference of the European Research Consortium for Informatics and Mathematics Working Group (ERCIM 2011) on Computing and Statistics held in London, UK [36] and they are presented in the paper "Nonparametric predictive inference for accuracy of ordinal diagnostic tests", which has been accepted for publication in *Journal of Statistical Theory and Practice* [35].

In Chapter 5, we first review classical methods for the ROC surface with three-groups of diagnostic tests. We present the NPI approach for assessment of accuracy of diagnostic tests involving three groups with real-valued data and we consider the choice of cut-off points for classifications based on such diagnostic tests. This is achieved by developing NPI lower and upper ROC surfaces, which are bounds to the empirical ROC surface. The volumes under these surfaces are also presented. For inferences on diagnostic tests, interest tends to be explicitly in future use of such tests, so a predictive approach seems attractive for such assessments and choice of cut-off points. A paper present the results in this chapter is in preparation.

We summarize our main results with some concluding remarks in Chapter 6. In the appendix, we enclose the proofs of some theorems in Chapter 3. Throughout this thesis, our methods are illustrated and discussed via examples with data from the literature. All the calculations have been done using the statistical software R version 2.15.0 ([www.r-project.org](http://www.r-project.org)).

# Chapter 2

## Introduction to literature

In this chapter we introduce the concepts from the literature that provide relevant background information for the topics considered in this thesis. Section 2.1 contains a brief overview of some classical methods for the analysis of ordinal data. An introduction to the concepts of the accuracy of diagnostic tests, including the Receiver Operating Characteristic (ROC) analysis, is given in Section 2.2. These concepts will be used in Chapters 4 and 5. Section 2.3 provides the main theory and concepts of Nonparametric Predictive Inference (NPI). Section 2.4 introduces NPI for ordinal data as presented by Coolen et al [26].

### 2.1 Ordinal data

In practice, ordinal data scales are prevalent in medical research and public health disciplines. For example, severity of an injury, degree of recovery from an illness and amount of exposure to a potentially harmful substance may all be measured on ordinal scales. Statistical methods for the analysis of ordinal data have received considerable attention. Excellent overviews of established statistical methods for ordinal data are presented by Agresti [1] and Liu and Agresti [45]. In what follows, we briefly introduce some classical statistical methods for ordinal data.

Everitt [37] discussed classical methods for analysis of contingency tables when columns and rows have ordered categories. For example, classification categories can be assigned numerical values, which allows the variables to be quantified and

regression methods to be used. Log-linear models for tables with ordered categories are also discussed. For more details we refer to Everitt [37]. A popular approach to analysis of classification tables is to test the null hypothesis of independence using the Chi-squared test, but in particular it is interesting to know how strong the association between groups is, which can be considered by the use of odds ratios. For more details we refer to Agresti [1].

Chen et al [15] presented multiple testing for ordinal data, considering an  $r \times c$  contingency table in which the categories are ordered. They tested the null hypothesis that log odds ratios are zero against the alternative that not all are zero, and also one-sided alternatives were studied. As test statistics, Fisher's exact statistic and Pearson's Chi-square statistic were used.

Another way to analyse contingency tables with ordered categories is order-restricted inference based on inequality constraints for parameters that recognize the ordering, as presented by Agresti [1, 2]. He used the constraint that ordinal log odds ratios are non-negative. First, the  $r \times 2$  table is considered with ordered rows, where the rows are independent binomial samples, with  $n_{i1}$  successes in  $n_i$  trials with parameters  $\pi_i$ , for  $i = 1, \dots, r$ , where  $n_{i2} = n_i - n_{i1}$  and  $n = \sum_i n_i$ . The null hypothesis  $H_0 : \pi_1 = \pi_2 = \dots = \pi_r$  is tested against the order-restricted alternative  $H_a : \pi_1 \leq \pi_2 \leq \dots \leq \pi_r$  or  $H_a : \pi_1 \geq \pi_2 \geq \dots \geq \pi_r$ . Let  $p_i = n_{i1}/n_i$  denote the sample proportions, under  $H_0$  the maximum likelihood (*ML*) estimator of  $\pi_i$  is the overall sample proportion of successes  $p = \sum_i n_{i1}/n$  for all  $i$ . If the  $r$  sample proportions satisfy  $p_1 \leq p_2 \leq \dots \leq p_r$ , then the order-restricted *ML* estimators of the  $\pi_i$  are  $\hat{\pi}_i = p_i$ . The likelihood-ratio test statistic (*LR*) for this test, suggested by Robertson et al [56], is

$$G^2(I|O) = 2 \sum_{i=1}^r n_{i1} \log \frac{\hat{\pi}_i}{p} + 2 \sum_{i=1}^r n_{i2} \log \frac{1 - \hat{\pi}_i}{1 - p} \quad (2.1)$$

where  $I$  denotes the null hypothesis of independence and  $O$  the order-restricted alternative hypothesis with  $p$ -value  $\sum_{d=1}^r w_d P(\chi_{d-1}^2 \geq G^2(I|O)_{obs})$ . Robertson et al [56] provided tables of  $w_d$ . Order-restricted inference for  $2 \times c$  tables with ordered columns is also presented in [1]. In general, for  $r \times c$  tables with ordered rows and columns, the order-restricted inference becomes complex as there are various order-restricted

alternatives, for more details we refer to Agresti [1]. In the NPI method presented in this thesis, we do not test any hypotheses and it does not matter how large  $r$  and  $c$  are. We will compare the above method with our method in Chapter 3.

## 2.2 Diagnostic accuracy

Diagnostic accuracy relates to the ability of a test to discriminate between different conditions. The evaluation of the accuracy of a diagnostic test is crucial in many research fields and applications areas such as medicine and machine learning [12,60]. Parametric and nonparametric methods have been developed for the analysis of diagnostic tests that yield results which may have only two values (binary tests), or a value in a finite number of ordered categories (ordinal tests), or real values (continuous tests) [55,73]. The ROC curve is one of the established tools for assessing the accuracy of a diagnostic test and the area under the ROC curve can be used as a summary index for diagnostic accuracy [55,73]. Fawcett [38] presented a clear introduction to ROC curves and related concepts.

As an example of application, Metz [49,50] provides a variety of studies in which ROC analyses have been employed in medical imaging. ROC analysis is used for medical image quality in terms of the ability of human observers or computer algorithms to use image data to classify patients with respect to any particular disease. Good recent overviews of statistical methods for ROC curves were presented by Zhou et al [73] and Pepe [55], considering both parametric and nonparametric approaches to inference on ROC curves and diagnostic accuracy.

To introduce the ROC curve, we use the following notation. Let  $X_j^i$ , with  $i = 0, 1$  and  $j = 1, \dots, n^i$ , be the observed continuous test results for the non-disease and disease groups, respectively, with  $n^0$  and  $n^1$  the sample sizes of the two groups and  $F_0, F_1$  the corresponding cumulative distribution functions (CDFs) for the non-disease and disease groups. We assume that the two groups are fully independent, in the sense that any information about one or more  $X_j^0$ , does not provide any information on one or more  $X_j^1$ , and the other way around. Using a cut-off point (threshold value)  $c$ , a test result is called positive if it is above the value  $c$  and



negative if below or equal  $c$ . The ROC is the plot of the true positive fraction ( $TPF$ ) and false positive fraction ( $FPF$ ) in the unit square, where  $TPF = P(X^1 > c)$  is the probability that the test result is positive, given that the condition (disease) is present and  $FPF = P(X^0 > c)$  is probability of a positive test result for an individual without the disease. The ROC curve is constructed by plotting the points  $(FPF(c), TPF(c))$  for  $c \in (-\infty, \infty)$ .

Using the empirical  $CDFs$  as estimates of the  $CDFs$  of  $X^0$  and  $X^1$ , respectively, the empirical ROC curve is the plot depicting  $(1 - \hat{F}_0(c), 1 - \hat{F}_1(c))$  for  $c \in (-\infty, \infty)$  [44, 55, 73] where

$$\hat{F}_0(c) = \sum_{j=1}^{n^0} \frac{I(X_j^0 \leq c)}{n^0}, \quad (2.2)$$

$$\hat{F}_1(c) = \sum_{j=1}^{n^1} \frac{I(X_j^1 \leq c)}{n^1} \quad (2.3)$$

The area under the ROC curve ( $AUC$ ) is a global measure of the test accuracy between the two groups. Generally, the  $AUC$  is equal to the probability that a randomly chosen individual from the non-disease group has a test result that is lower than that of a randomly selected individual from the disease group, i.e.  $AUC = P(X^0 < X^1)$ . Higher  $AUC$  values indicate more accurate tests, with  $AUC = 1$  for ideal tests and  $AUC = 0.5$  for uninformative tests [44, 55, 73].

There are many medical diagnostic situations in which the disease status has more than two groups. Therefore, the ROC methodology was extended to three-group diagnostic problems by constructing a three-dimensional surface. Mossman [51] proposed the concept of  $ROC$  surfaces for three groups of disease, so with three possible diagnostic outcomes. The volume under the ROC surface ( $VUS$ ) is a direct extension of the  $AUC$  and it is a global accuracy measure of a test to classify subjects in three groups. The interpretation of the  $VUS$  is similar to the  $AUC$ , namely it is equal to the probability that three random measurements, one from each group, appear in the correct order [52, 71]. Nakas and Yiannoutsos [53] considered nonparametric estimation of the  $VUS$  using several methods including bootstrapping for the estimation of the variance of the  $VUS$ . A nonparametric Bayesian approach for estimating the ROC surface has been presented by Inácio et al [47]. Parametric

estimators for the ROC surface and the corresponding VUS were presented by Beck [9].

## 2.3 Nonparametric Predictive Inference

The statistical method used in this thesis uses lower and upper probabilities to quantify uncertainty [20, 31]. In classical probability theory a single precise probability  $P(A)$  is used for each event  $A$ , satisfying Kolmogorov's axioms. However, when information is vague or conflicting, a unique probability distribution may be hard to identify. In this case, imprecise probability is an alternative approach which replaces the single probability  $P(A)$  with an interval specified by its bounds, called the lower and upper probabilities and denoted by  $\underline{P}(A)$  and  $\overline{P}(A)$ , respectively, where  $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ . The idea to use interval probabilities dates back at least to the middle of the nineteenth century [11, 31].

Nonparametric predictive inference (NPI) [4, 20] is based on the assumption  $A_{(n)}$  proposed by Hill [42], which gives direct probabilities for future observations based on observed values of related random quantities. Let  $X_1, \dots, X_n, X_{n+1}$  be real-valued absolutely continuous and exchangeable random quantities. Let the ordered observed values of  $X_1, X_2, \dots, X_n$  be denoted by  $x_1 < x_2 < \dots < x_n$  and let  $x_0 = -\infty$  and  $x_{n+1} = \infty$  for ease of notation. We assume that no ties occur; ties can be dealt with in NPI by assuming that tied observations differ by small amounts which tend to zero [43] but it is not relevant in Chapters 3 and 4 as in the assumed latent variable representation we assume that no ties occur (as discussed later in next section). In Chapter 5, we avoid complications, particularly with regard to notation, by assuming that the data do not contain any ties (we discuss this further in Chapter 5).

For  $X_{n+1}$ , representing a future observation, based on  $n$  observations,  $A_{(n)}$  [42] partially specifies a probability distribution by  $P(X_{n+1} \in I_i = (x_{i-1}, x_i)) = \frac{1}{n+1}$  for  $i = 1, \dots, n+1$ .  $A_{(n)}$  does not assume anything else, and can be considered to be a post-data assumption related to exchangeability [33]. 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’ [33], which are lower and upper probabilities in interval probability theory [20, 65, 67, 68].

In NPI, uncertainty about the future observation  $X_{n+1}$  is quantified by lower and upper probabilities for events of interest. As we mentioned before, lower and upper probabilities generalize classical (‘precise’) probabilities, and a lower probability  $\underline{P}(A)$  for event  $A$  in NPI can be interpreted as the maximum lower bound for the probability of  $A$  that follows from the assumptions made [20]. Similarly, an upper probability  $\overline{P}(A)$  for event  $A$  can be interpreted as the minimum upper bound for the probability of  $A$ . In NPI, we wish to explore application of  $A_{(n)}$  for inference without making further assumptions. So, NPI lower and upper probabilities are the sharpest bounds on a probability for an event of interest when only  $A_{(n)}$  is assumed. Informally,  $\underline{P}(A)$  can be considered to reflect the evidence in favour of event  $A$  and the corresponding upper probability  $\overline{P}(A)$  can be interpreted by considering that  $\overline{P}(A)$  reflects the evidence against event  $A$ , so in support of the complementary event  $A^c$ . Augustin and Coolen [4] proved that NPI has strong consistency properties in the theory of interval probability [65, 67, 68], it is also exactly calibrated [46], and it never leads to results that are in conflict with inferences based on empirical probabilities.

Direct application of  $A_{(n)}$  for inferential problems is only possible for real-valued random quantities [16, 32]. However, by assuming latent variable representations and variations to  $A_{(n)}$ , NPI has been developed for other situations, including Bernoulli data [18] and multinomial data [22]. A latent variable representation can also be used for ordinal data, as shown in the following section. Ordinal data and multinomial data differ in the important fact that the data categories are not ordered for multinomial data but they are logically ordered for ordinal data (the difference will be discussed further in Section 2.4).

While it is natural to consider inference for a single future observation in many situations, one may also be interested in multiple future observations. This is pos-

sible in NPI in a sequential way, taking the inter-dependence of the multiple future observations into account [3]. For example in NPI for Bernoulli quantities this was included throughout [18] and dependence of specific inferences on the choice of the number of future Bernoulli observations was explicitly studied in the context of comparison of proportions by Coolen and Coolen-Schrijner [24]. In this thesis, new developments for NPI for ordinal data and accuracy of diagnostic tests are presented with attention restricted to a single future observation, leaving generalization to multiple future observations as an interesting challenge for future research. Throughout this thesis, we present empirical probabilities and, where appropriate, inferences based on these, in order to compare them with the NPI results. These empirical probabilities coincide with the nonparametric maximum likelihood estimates in the scenarios considered, and are generally bounded by the corresponding NPI lower and upper probabilities.

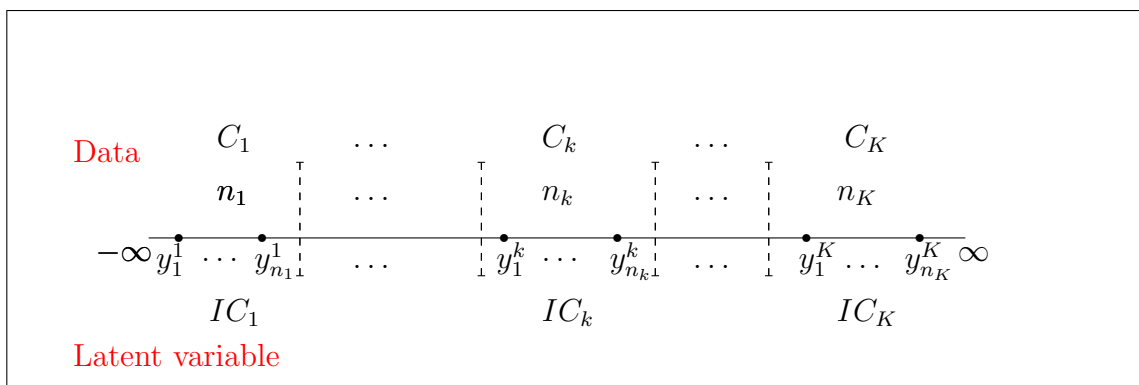
## 2.4 NPI for ordinal data

NPI for categorical data with a known number of possible categories, yet with no ordering or other known relationship between the categories, has been presented by Coolen and Augustin [22]. The inferences about the next observation are based on a latent variable representation using a probability wheel, with the assumption that each category is only allowed to be represented by a single segment of the wheel, where a segment is defined as a single part of the wheel. The NPI lower and upper probabilities with regard to the next observation are further based on a circular version of  $A_{(n)}$  [19] and optimisation over all configurations of the probability wheel that are possible corresponding to the data and this so-called circular- $A_{(n)}$  assumption. Baker [5] presents several further developments and applications of NPI for non-ordered categorical data, including consideration of sub-categories and application to classification problems. For further details of NPI for non-ordered categorical data we refer to Coolen and Augustin [22].

In many applications the categories are ordered, e.g. levels of severity of a disease, in which case such data are also known as ordinal data. It is important that such

knowledge about ordering of categories is taken into account, this section presents NPI for such data. The method uses 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  $A_{(n)}$ .

This section summarizes results from [25, 26] who introduced NPI for ordinal categorical data in terms of lower and upper probabilities. In case of ordinal data, suppose there are  $K \geq 2$  categories which have a natural fixed ordering  $C_1 < C_2 < \dots < C_K$  and there is a known number of observations in each category, say  $n_k$  in category  $C_k$  for  $k = 1, \dots, K$ . Let  $n$  be the total number of the observations, so  $n = \sum_{k=1}^K n_k$ . A latent variable representation of the observations on the real-line is used (i.e. observations per category are represented by corresponding values on the real-line). Using this method the category  $C_k$  is represented by the interval  $IC_k$  for  $k = 1, \dots, K$ , so in this case there are  $K$  ordered intervals  $IC_1, \dots, IC_K$  forming a partition of the real-line. We assume that the  $n$  observations are represented by  $y_1 < \dots < y_n$ , and those in  $C_k$  are also denoted by  $y_j^k$  for  $j = 1, \dots, n_k$ . Let  $X_{n+1}$  denote the random quantity representing a future observation and  $Y_{n+1}$  be the latent variable on the real-line corresponding to the future observation  $X_{n+1}$ , so the event  $X_{n+1} \in C_k$  corresponds to the event  $Y_{n+1} \in IC_k$ . This allows  $A_{(n)}$  to be directly applied to  $Y_{n+1}$ , and then transformed to inference on the categorical random quantity  $X_{n+1}$  [26]. The ordinal data structure is presented in Figure 2.1.



**Figure 2.1:** Structure for ordinal data and latent variables

The NPI lower and upper probabilities for general events of the form  $X_{n+1} \in \mathcal{C}_T$  have been derived with  $\mathcal{C}_T = \bigcup_{k \in T} C_k$  and  $T \subset \{1, \dots, K\}$ . It is assumed that  $T$

is a strict subset of  $\{1, \dots, K\}$ , as the event that a future observation falls into any of the  $K$  categories is necessarily true and has NPI lower and upper probabilities both equal to one. The lower probability is super-additive and upper probability is sub-additive [65], so the derivations of these do not follow straightforwardly from the NPI lower and upper probabilities for the events involving single categories.

Using the assumption  $A_{(n)}$  for  $Y_{n+1}$  in the latent variable representation, each interval  $I_i$  has been assigned probability mass  $1/(n+1)$ . Although the values  $y_i$  are not exactly known since they only exist in the latent variable representation, the number of these  $y_i$  values in each interval  $IC_k$  is known. It should be emphasized that the intervals  $I_i$  are, as before, intervals between consecutive latent points  $y_{i-1}$  and  $y_i$ , and with the number of such points in each interval  $IC_k$  known, it is therefore also known how many intervals  $I_i$  are fully within each  $IC_k$ .

To derive the NPI lower probability for the event  $X_{n+1} \in \mathcal{C}_T$ , the NPI lower probability for the corresponding latent variable event  $Y_{n+1} \in \mathcal{IC}_T$  is derived, where  $\mathcal{IC}_T = \bigcup_{k \in T} IC_k$  and  $T \subset \{1, \dots, K\}$ . To derive this lower probability, one calculates the sum of all probability masses assigned to intervals  $I_i$  that are fully within  $\mathcal{IC}_T$ , so in effect one minimises the total probability mass assigned to  $\mathcal{IC}_T$ . So, the NPI lower probability is

$$\underline{P}(X_{n+1} \in \mathcal{C}_T) = \underline{P}(Y_{n+1} \in \mathcal{IC}_T) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}\{I_i \subset \mathcal{IC}_T\} \quad (2.4)$$

where  $\mathbf{1}\{A\}$  is equal to 1 if  $A$  is true and equal to 0 else. As the exact locations of the intervals  $IC_k$  are unknown, this may appear to be vague, yet the fact that the number of  $y_i$  values within each interval  $IC_k$  is known suffices to get unique values for these NPI lower probabilities.

The corresponding NPI upper probabilities are derived by maximising the total probability mass that can be assigned to  $\mathcal{IC}_T$ , so by summing all probability masses corresponding to intervals  $I_i$  that have a non-empty intersection with  $\mathcal{IC}_T$ . Hence, the NPI upper probability is

$$\overline{P}(X_{n+1} \in \mathcal{C}_T) = \overline{P}(Y_{n+1} \in \mathcal{IC}_T) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}\{I_i \cap \mathcal{IC}_T \neq \emptyset\} \quad (2.5)$$

From the perspective of frequentist statistics, the NPI lower and upper probabilities (2.4) and (2.5) can be considered as ‘confidence statements’, in the sense

that repeated application of this procedure will lead to correct predictions of the event  $X_{n+1} \in \mathcal{C}_T$  in a proportion that, in the limit, will be in the interval  $[\underline{P}(X_{n+1} \in \mathcal{C}_T), \overline{P}(X_{n+1} \in \mathcal{C}_T)]$ . Of course, to achieve this result the method should only be applied in cases where the assumption  $A_{(n)}$  is reasonable. This would, for example, not be the case if the data show a clear patterns, e.g. for time series when the order of the data is important and informative.

### 2.4.1 Special cases

The first special case of these inferences concerns the event  $X_{n+1} \in \mathcal{C}_T$ , with  $\mathcal{C}_T$  consisting of adjoining categories, so the corresponding union of intervals  $\mathcal{IC}_T$  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. Let  $T = \{s, \dots, t\}$ , with  $s, t \in \{1, \dots, K\}$ ,  $s \leq t$ , excluding the case with  $s = 1$  and  $t = K$  for which both the NPI lower and upper probabilities are equal to one. Let  $\mathcal{C}_{s,t} = \bigcup_{k=s}^t C_k$ ,  $\mathcal{IC}_{s,t} = \bigcup_{k=s}^t IC_k$  and let  $n_{s,t} = \sum_{k=s}^t n_k$ . Using the notation  $(x)^+ = \max(x, 0)$ , the NPI lower and upper probabilities (2.4) and (2.5) for such events are

$$\underline{P}(X_{n+1} \in \mathcal{C}_{s,t}) = \underline{P}(Y_{n+1} \in \mathcal{IC}_{s,t}) = \begin{cases} \frac{(n_{s,t} - 1)^+}{n + 1} & \text{if } 1 < s \leq t < K \\ \frac{n_{s,t}}{n + 1} & \text{if } s = 1 \text{ or } t = K \end{cases} \quad (2.6)$$

$$\overline{P}(X_{n+1} \in \mathcal{C}_{s,t}) = \overline{P}(Y_{n+1} \in \mathcal{IC}_{s,t}) = \frac{n_{s,t} + 1}{n + 1} \quad \text{for } 1 \leq s \leq t \leq K \quad (2.7)$$

Of course, the case with  $s = t$  gives the event that the next observation belongs to one specific category.

The second special case for which closed-form expressions are available for the NPI lower and upper probabilities occurs if  $n_k > 0$  for all  $k \in \{1, \dots, K\}$ , so there are observations in all  $K$  categories. One needs to consider if  $C_1$  and  $C_K$  are in  $\mathcal{C}_T$  (so  $IC_1$  and  $IC_K$  in  $\mathcal{IC}_T$ ) and to take account of all pairs of neighbouring categories which are both included in  $\mathcal{C}_T$ . Let  $p_T = \sum_{r=1}^{K-1} \mathbf{1}\{r, r+1 \in T\}$  be the number of neighbouring pairs of categories included in  $\mathcal{C}_T$ , and  $e_T = \mathbf{1}\{1 \in T\} + \mathbf{1}\{K \in T\} + p_T$

the number of extra intervals. Furthermore,  $s_T$  is the number of categories in  $C_T$ . Then the NPI lower probability (2.4), with  $T$  a strict subset of  $\{1, \dots, K\}$ , is

$$\underline{P}(X_{n+1} \in C_T) = \underline{P}(Y_{n+1} \in \mathcal{IC}_T) = \frac{\sum_{k \in T} (n_k - 1) + e_T}{n + 1} = \frac{n_T - s_T + e_T}{n + 1} \quad (2.8)$$

and the corresponding NPI upper probability (2.5) is

$$\overline{P}(X_{n+1} \in C_T) = \overline{P}(Y_{n+1} \in \mathcal{IC}_T) = \frac{\sum_{k \in T} (n_k + 1) - p_T}{n + 1} = \frac{n_T + s_T - p_T}{n + 1} \quad (2.9)$$

The NPI lower and upper probabilities presented in this section satisfy the following properties:

1. The conjugacy property

$$\underline{P}(X_{n+1} \in C_T) = 1 - \overline{P}(X_{n+1} \in C_{T^c})$$

for all  $T \subset \{1, \dots, K\}$  and  $T^c = \{1, \dots, K\} \setminus T$ , which follows from  $\mathbf{1}\{I_i \in \mathcal{IC}_T\} + \mathbf{1}\{I_i \in \mathcal{IC}_{T^c}\} = 1$  for all  $i = 1, \dots, n + 1$ .

2. The NPI lower and upper probabilities (2.4) and (2.5) bound the corresponding empirical probability for the event of interest. This means that

$$\underline{P}(X_{n+1} \in C_T) \leq \frac{n_T}{n} \leq \overline{P}(X_{n+1} \in C_T) \quad (2.10)$$

Property (2.10) can be considered attractive when aiming at 'objective inference', and the possibility to satisfy this property is an important advantage of statistical methods using lower and upper probabilities [19].

**Example 2.1.** Suppose there are  $K = 7$  ordered categories,  $C_1 < \dots < C_7$ , and  $n = 17$  observations as presented in Table 2.1.

Category	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
Observations	2	3	0	5	2	1	4

**Table 2.1:** Data



$T$	$\underline{P}$	$\overline{P}$	$n_T/n$
{1}	2/18	3/18	2/17
{3}	0/18	1/18	0/17
{4}	4/18	6/18	5/17
{5}	1/18	3/18	2/17
{6}	0/18	2/18	1/17
{7}	4/18	5/18	4/17
{6, 7}	5/18	6/18	5/17
{1, 2, 3}	5/18	6/18	5/17
{3, 4, 5}	6/18	8/18	7/17
{2, 4, 6}	6/18	11/18	9/17
{1, 4, 7}	10/18	14/18	11/17
{3, 4, 6, 7}	9/18	12/18	10/17
{4, 5, 6, 7}	12/18	13/18	12/17
{2, 3, 5, 6}	4/18	8/18	6/17
{1, 2, 5, 6, 7}	12/18	14/18	12/17

**Table 2.2:** NPI lower and upper probabilities

Table 2.2 gives the NPI lower and upper probabilities for several events  $Y_{18} \in \mathcal{C}_T$ , together with the corresponding empirical probability  $n_T/n$ . It is clear that these lower and upper probabilities are bounds for the empirical probability and they also show that the difference between corresponding upper and lower probabilities is not constant. This difference is called imprecision, denoted by  $\Delta$ , and it provides insight into the link between these lower and upper probabilities and the amount of information available [24, 65], namely large numbers of observations lead to small imprecision. The lower and upper probabilities for the events with  $T$  consisting of a single category or a group of adjoining categories are also illustrated. The first special case is used when  $T$  is  $\{6, 7\}$ ,  $\{1, 2, 3\}$ ,  $\{3, 4, 5\}$  and  $\{4, 5, 6, 7\}$ , the rest of Table 2.2 is calculated directly from the main formula (2.4) and (2.5). In addition, it is clear that these NPI lower and upper probabilities satisfy the conjugacy property, for example  $\underline{P}(\{1, 4, 7\}) = 1 - \overline{P}(\{2, 3, 5, 6\})$ .

To illustrate the second special case, we change the data in Table 2.1 by adding

one observation to category 3 taken from category 4. Table 2.3 shows the new data

Category	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$
Observations	2	3	1	4	2	1	4

**Table 2.3:** Data

and Table 2.4 gives the NPI lower and upper probabilities for these data. It is

$T$	$\underline{P}$	$\bar{P}$	$n_T/n$
{1}	2/18	3/18	2/17
{3}	0/18	2/18	1/17
{4}	3/18	5/18	4/17
{5}	1/18	3/18	2/17
{6}	0/18	2/18	1/17
{7}	4/18	5/18	4/17
{6, 7}	5/18	6/18	5/17
{1, 2, 3}	6/18	7/18	6/17
{3, 4, 5}	6/18	8/18	7/17
{2, 4, 6}	5/18	10/18	8/17
{1, 4, 7}	9/18	13/18	10/17
{3, 4, 6, 7}	9/18	12/18	10/17
{4, 5, 6, 7}	11/18	12/18	11/17
{2, 3, 5, 6}	5/18	9/18	6/17
{1, 2, 5, 6, 7}	12/18	14/18	12/17

**Table 2.4:** NPI lower and upper probabilities

clear that the second special case can be applied for  $\{2, 4, 6\}$ ,  $\{1, 4, 7\}$ ,  $\{3, 4, 6, 7\}$ ,  $\{2, 3, 5, 6\}$  and  $\{1, 2, 5, 6, 7\}$  in Table 2.4. For example,  $\{3, 4, 6, 7\}$  has  $n_T = 10$ ,  $s_T = 4$ ,  $p_T = 2$ ,  $e_T = 3$ .

△

# Chapter 3

## Multiple comparisons for ordinal data

### 3.1 Introduction

In this chapter we introduce NPI multiple comparisons for ordinal data, including two different problems of subset selection for groups of ordinal data. Methods for multiple comparisons, including subset selection, have been studied extensively in the statistics literature [8, 41, 64]. Similar NPI methods for multiple comparisons, with some important variations, have been presented for real-valued data [16, 32], for proportions data [23, 24] and for lifetime data including right-censored observations [29, 30].

As before, we assume that there are  $K$  ordered categories. Suppose that there are  $J \geq 2$  independent groups and  $n^j$  observations for group  $j$  ( $j = 1, \dots, J$ ) of which  $n_k^j$  are in category  $C_k$ ,  $k = 1, \dots, K$ . So  $n = \sum_{j=1}^J n^j = \sum_{j=1}^J \sum_{k=1}^K n_k^j$ , and let  $n_{s,t}^j = \sum_{k=s}^t n_k^j$  where  $s \leq t$ . The assumption of ‘independence of the groups’ means that any information about a random quantity in one group does not provide any information about a random quantity in any other group. We apply the  $A_{(n)}$ -based inferences per group to consider one future observation from each group. Let  $X_{n^j+1}^j$  denote the next observation from group  $j$  and let the corresponding latent variable be denoted by  $Y_{n^j+1}^j$ .

In Section 3.2 we derive the NPI lower and upper probabilities for the event

that the next observation from group  $j$ ,  $X_{n^j+1}^j$  is less than (or equal to) the next observation from each of the other groups,  $X_{n^l+1}^l$ ,  $l = 1, \dots, J$ ,  $l \neq j$ .

Thereafter, we divide the multiple groups into two non-empty subsets of groups,  $S$  and  $S^c$ . Again we apply the  $A_{(n)}$ -based inferences per group to consider one future observation for each group. In Section 3.3 we derive the NPI lower and upper probabilities for the event that all these future observations for groups in  $S$  are less (so in terms of categories ‘further to the left’) than all the future observations for groups in  $S^c$ . In Section 3.4 we consider the variation that at least one of the future observations for groups in  $S$  is less than all the future observations for groups in  $S^c$ . The group which provides the smallest next observation is also called the best group in this chapter. In Section 3.5 we discuss how to use these NPI lower and upper probabilities for selecting the best subset. Our method is presented in terms of minimum next observation(s), in practice we may be interested in the maximum next observation(s), this is briefly discussed in Section 3.6. This chapter ends with concluding remarks in Section 3.7.

## 3.2 Selecting one group

In this section we present NPI for selecting a single group from  $J$  groups by deriving the NPI lower and upper probabilities for the event that the next observation of group  $j$ ,  $X_{n^j+1}^j$ , is less than the next observation from each of the other groups,  $X_{n^l+1}^l$ ,  $l = 1, \dots, J$ ,  $l \neq j$ . We use the notation  $X_{n^j+1}^j < X_{n^l+1}^l$  for the event  $X_{n^j+1}^j \in C_{k_j}$  and  $X_{n^l+1}^l \in C_{k_l}$  for  $k_j, k_l \in \{1, \dots, K\}$  with  $k_j < k_l$ , so  $X_l$  will be in a category to the right of  $X_j$ , and similar for  $X_{n^j+1}^j \leq X_{n^l+1}^l$ . Two specific configurations of the probability masses in the intervals that are important for what follows, are defined by putting all the probability masses at the left endpoints ( $L$ ) or at the right endpoints ( $R$ ), so

$$P^L(X_{n^j+1}^j \in \mathcal{C}_{s,t}) = P^L(Y_{n^j+1}^j \in \mathcal{IC}_{s,t}) = \begin{cases} \frac{n_{s,t}^j + 1}{n^j + 1} & \text{if } s = 1 \\ \frac{n_{s,t}^j}{n^j + 1} & \text{if } 1 < s \leq t \leq K \end{cases} \quad (3.1)$$

and

$$P^R(X_{n^{j+1}}^j \in \mathcal{C}_{s,t}) = P^R(Y_{n^{j+1}}^j \in \mathcal{IC}_{s,t}) = \begin{cases} \frac{n_{s,t}^j}{n^j + 1} & \text{if } 1 \leq s \leq t < K \\ \frac{n_{s,t}^j + 1}{n^j + 1} & \text{if } t = K \end{cases} \quad (3.2)$$

where if  $t = s$  we have  $\mathcal{C}_{s,s} = C_s$  and  $n_{s,s}^j = n_s^j$ . Probability (3.1) corresponds to the situation where, in the latent variable representation of the categories and the data, all probability masses for the next observation following from the  $A_{(n)}$  assumption are put at the left endpoint per interval (left configurations). Similarly, probability (3.2) corresponds to the situation with all these probability masses put at the right endpoint per interval (right configurations).

Theorem 3.1 gives the NPI lower and upper probabilities for the event that the next observation of group  $j$ ,  $X_{n^{j+1}}^j$ , is less than the next observation from each of the other groups,  $X_{n^{l+1}}^l$  for  $l = 1, \dots, J, l \neq j$ , denoted by

$$\underline{P}_j^< = \underline{P} \left( X_{n^{j+1}}^j < \min_{l=1, \dots, J, l \neq j} X_{n^{l+1}}^l \right)$$

and

$$\overline{P}_j^< = \overline{P} \left( X_{n^{j+1}}^j < \min_{l=1, \dots, J, l \neq j} X_{n^{l+1}}^l \right)$$

We use the notation  $\mathcal{C}_{a,b} = \{C_a, \dots, C_b\}$ .

**Theorem 3.1.** The NPI lower and upper probabilities for the event that the next observation of group  $j$  is less than the next observation from each of the other groups are

$$\underline{P}_j^< = A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J n_{k+1, K}^l \right) \right] \quad (3.3)$$

$$\overline{P}_j^< = A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k+1, K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n_{2, K}^l + 1) \right] \quad (3.4)$$

where  $A = \left[ \prod_{j=1}^J (n^j + 1) \right]^{-1}$ .

*Proof.* The NPI lower probability is derived as follows. Firstly observe that

$$P\left(X_{n^{j+1}}^j < \min_{l=1, \dots, J, l \neq j} X_{n^{l+1}}^l\right) = P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{j+1}}^j < X_{n^{l+1}}^l\}\right) \quad (3.5)$$

By the law of total probability, Equation (3.5) is equivalent to

$$\sum_{k=1}^K P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P(X_{n^{j+1}}^j \in C_k) \quad (3.6)$$

In this derivation, we use the assumptions  $A_{(n^j)}$  for all groups with assumed independence of the  $J$  groups. To get the lower bound we put the probability mass per interval at the endpoints; for the group  $j$  of interest at the right endpoint and for all other groups at the left endpoints, by using (3.1) and (3.2), as we are interested in the lower bound for the probability of  $X_{n^{j+1}}^j < X_{n^{l+1}}^l$ . This probability must be as small as possible, which happens if  $X^j$  is as large as possible and  $X^l$  is as small as possible, so we used the right configuration for  $X_{n^{j+1}}^j$ . So, for  $k = 1, \dots, K-1$ ,

$$P(X_{n^{j+1}}^j \in C_k) \geq P^R(X_{n^{j+1}}^j \in C_k). \quad (3.7)$$

For  $k = K$ , since we are interested in all the cases where  $X_{n^{j+1}}^j < X_{n^{l+1}}^l$  (i.e. if  $X_{n^{j+1}}^j \in C_K$  then it is impossible for  $X_{n^{l+1}}^l$  to be greater than  $X_{n^{j+1}}^j$ )

$$P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_K\right) = 0. \quad (3.8)$$

Using (3.6), (3.7) and (3.8), we get

$$\begin{aligned} P\left(X_{n^{j+1}}^j < \min_{l=1, \dots, J, l \neq j} X_{n^{l+1}}^l\right) &\geq \sum_{k=1}^{K-1} P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \\ &\quad \times P^R(X_{n^{j+1}}^j \in C_k) \\ &= \sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} P\left(\bigcap_{l=1, l \neq j}^J (X_{n^{l+1}}^l \in C_{k+1, K})\right) \end{aligned} \quad (3.9)$$

Using the independence of the groups Equation (3.9) is equivalent to

$$\sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in C_{k+1, K}) \quad (3.10)$$

Using the the left configuration for the groups  $l, l \neq j$  for  $k = 1, \dots, K - 1$  leads to

$$P(X_{n^{l+1}}^l \in \mathcal{C}_{k+1,K}) \geq P^L(X_{n^{l+1}}^l \in \mathcal{C}_{k+1,K}) \quad (3.11)$$

Using (3.10) and (3.11), we get

$$\begin{aligned} P\left(X_{n^{j+1}}^j < \min_{l=1,\dots,J,l \neq j} X_{n^{l+1}}^l\right) &\geq \sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P^L(X_{n^{l+1}}^l \in \mathcal{C}_{k+1,K}) \\ &= \sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left(\sum_{f=k+1}^K \frac{n_f^l}{n^l + 1}\right) \\ &= A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J n_{k+1,K}^l \right) \right] \end{aligned} \quad (3.12)$$

Formula (3.12) is the maximum lower bound for  $P\left(X_{n^{j+1}}^j < \min_{l=1,\dots,J,l \neq j} X_{n^{l+1}}^l\right)$  which can be achieved and therefore it is a lower probability [65], so the NPI lower probability is

$$\underline{P}_j^< = A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J n_{k+1,K}^l \right) \right] \quad (3.13)$$

The corresponding NPI upper probability is obtained by putting the probability masses per interval at the other endpoints; for group  $j$  at the left endpoints, and for all other groups at the right endpoints. The upper bound for the probability for the event  $X_{n^{j+1}}^j < \min_{l=1,\dots,J,l \neq j} X_{n^{l+1}}^l$  can be derived as follows:

$$\begin{aligned} P\left(X_{n^{j+1}}^j < \min_{l=1,\dots,J,l \neq j} X_{n^{l+1}}^l\right) &= P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{j+1}}^j < X_{n^{l+1}}^l\}\right) \\ &= \sum_{k=1}^K P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P(X_{n^{j+1}}^j \in C_k) \\ &\leq \sum_{k=1}^{K-1} P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P^L(X_{n^{j+1}}^j \in C_k) \\ &\quad + P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_K\right) \times P^L(X_{n^{j+1}}^j \in C_K) \end{aligned} \quad (3.14)$$

$$= \sum_{k=1}^{K-1} P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P^L(X_{n^{j+1}}^j \in C_k) \quad (3.15)$$

$$\begin{aligned}
&= \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} P \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k \right) \\
&\quad + \frac{n_1^j + 1}{n^j + 1} P \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l > X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_1 \right) \tag{3.16}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in \mathcal{C}_{k+1, K}) + \frac{n_1^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in \mathcal{C}_{2, K}) \\
&\leq \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P^R(X_{n^{l+1}}^l \in \mathcal{C}_{k+1, K}) + \frac{n_1^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J P^R(X_{n^{l+1}}^l \in \mathcal{C}_{2, K}) \\
&= \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \sum_{f=k+1}^{K-1} \frac{n_f^l}{n^l + 1} + \frac{n_K^l + 1}{n^l + 1} \right) + \frac{n_1^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \sum_{f=2}^{K-1} \frac{n_f^l}{n^l + 1} + \frac{n_K^l + 1}{n^l + 1} \right) \tag{3.17}
\end{aligned}$$

$$\begin{aligned}
&= A \left[ \sum_{k=2}^{K-1} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k+1}^K n_f^l + 1 \right) + (n_1^j + 1) \prod_{l=1, l \neq j}^J \left( \sum_{f=2}^K n_f^l + 1 \right) \right] \\
&= A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k+1, K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n_{2, K}^l + 1) \right] \tag{3.18}
\end{aligned}$$

This is the minimal upper bound which can be justified and it can actually be attained, so this upper bound (3.18) is the upper probability [65]

$$\bar{P}_j^< = A \left[ \sum_{k=1}^{K-1} \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k+1, K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n_{2, K}^l + 1) \right] \tag{3.19}$$

Similarly to the derivation of the lower probability, to go from (3.14) to (3.15) we use that, when  $X_{n^{j+1}}^j \in C_K$ , the probability of the event  $X_{n^{l+1}}^l < X_{n^{j+1}}^j$  is zero. We get Equation (3.16) by using the left configuration for  $X_{n^{j+1}}^j$  and Equation (3.17) by using the right configuration for the other groups.  $\square$

It is also of interest to consider the event that the next observation from group  $j$ ,  $X_{n^{j+1}}^j$ , is less than or equal to the next observation from each of the other groups,  $X_{n^{l+1}}^l$  for  $l = 1, \dots, J$ ,  $l \neq j$ . The NPI lower and upper probabilities for these events, are denoted by

$$\underline{P}_j^{\leq} = P \left( X_{n^{j+1}}^j = \min_{l=1, \dots, J} X_{n^{l+1}}^l \right)$$

and

$$\bar{P}_j^{\leq} = \bar{P} \left( X_{n^{j+1}}^j = \min_{l=1, \dots, J} X_{n^{l+1}}^l \right)$$



These are given in Theorem 3.2.

**Theorem 3.2.** The NPI lower and upper probabilities for the event that the next observation from group  $j$  is less than or equal to next observation from each of the other groups are

$$\underline{P}_j^{\leq} = A \left[ n_1^j \prod_{l=1, l \neq j}^J (n^l + 1) + \sum_{k=2}^K \left( n_k^j \prod_{l=1, l \neq j}^J n_{k,K}^l \right) + \prod_{l=1, l \neq j}^J n_K^l \right] \quad (3.20)$$

$$\overline{P}_j^{\leq} = A \left[ \sum_{k=1}^K \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k,K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \quad (3.21)$$

where  $A = \left[ \prod_{j=1}^J (n^j + 1) \right]^{-1}$ .

The derivation of the NPI lower and upper probabilities in this theorem is similar to the proof of Theorem 3.1. The detailed proof is given in Appendix A.1.

### 3.2.1 Empirical probabilities

Let us consider the same cases as in Theorems 3.1 and 3.2, but now in terms of empirical probability, with  $EP$  the notation for empirical probability. So, let  $EP_j^{\leq}$  be the empirical probability for the event that  $X^j < \min_{l=1, \dots, J; l \neq j} X^l$ , this is given by

$$\begin{aligned} EP_j^{\leq} &= P \left( X^j < \min_{l=1, \dots, J; l \neq j} X^l \right) = P \left( \bigcap_{l=1, l \neq j}^J \{X^j < X^l\} \right) \\ &= \sum_{k=1}^{K-1} P \left( \bigcap_{l=1, l \neq j}^J \{X^l > X^j\} \mid X^j \in C_k \right) \times P(X^j \in C_k) \\ &= \sum_{k=1}^{K-1} \frac{n_k^j}{n^j} P \left( \bigcap_{l=1, l \neq j}^J \{X^l > X^j\} \mid X^j \in C_k \right) = \sum_{k=1}^{K-1} \frac{n_k^j}{n^j} P \left( \bigcap_{l=1, l \neq j}^J (X^l \in C_{k+1, K}) \right) \\ &= \sum_{k=1}^{K-1} \frac{n_k^j}{n^j} \prod_{l=1, l \neq j}^J P(X^l \in C_{k+1, K}) = \sum_{k=1}^{K-1} \frac{n_k^j}{n^j} \prod_{l=1, l \neq j}^J \left( \sum_{f=k+1}^K \frac{n_f^l}{n^l} \right) \\ &= \frac{1}{\prod_{j=1}^J n^j} \left[ \sum_{k=1}^{K-1} \left\{ n_k^j \prod_{l=1, l \neq j}^J n_{k+1, K}^l \right\} \right] \end{aligned} \quad (3.22)$$

Similarly, the empirical probability for the event  $X^j \leq X^l$  for all  $l \neq j$ ,  $EP_j^{\leq}$ , is

$$\begin{aligned}
EP_j^{\leq} &= P\left(X^j = \min_{l=1, \dots, J} X^l\right) = P\left(\bigcap_{l=1, l \neq j}^J \{X^j \leq X^l\}\right) \\
&= \sum_{k=1}^K P\left(\bigcap_{l=1, l \neq j}^J \{X^l \geq X^j\} \mid X^j \in C_k\right) \times P(X^j \in C_k) \\
&= \sum_{k=1}^K \frac{n_k^j}{n^j} P\left(\bigcap_{l=1, l \neq j}^J \{X^l \geq X^j\} \mid X^j \in C_k\right) = \sum_{k=1}^K \frac{n_k^j}{n^j} P\left(\bigcap_{l=1, l \neq j}^J (X^l \in \mathcal{C}_{k,K})\right) \\
&= \sum_{k=1}^K \frac{n_k^j}{n^j} \prod_{l=1, l \neq j}^J P(X^l \in \mathcal{C}_{k,K}) = \sum_{k=1}^K \frac{n_k^j}{n^j} \prod_{l=1, l \neq j}^J \left(\sum_{f=k}^K \frac{n_f^l}{n^l}\right) \\
&= \frac{1}{\prod_{j=1}^J n^j} \left[ \sum_{k=1}^K \left\{ n_k^j \prod_{l=1, l \neq j}^J n_{k,K}^l \right\} \right] \tag{3.23}
\end{aligned}$$

### 3.2.2 Examples

We illustrate the lower and upper probabilities presented in this section via examples, considering data sets presented in the literature.

**Example 3.1.** The data of this example are taken from Simonoff [59]. The independence was rejected for this data set and the outlire identification and Robust estimation show that small cars are more likely to be predictive to have good reliability, whereas the medium ones are more likely to be average [59]. We use this data set to illustrate the NPI lower and upper probabilities and the empirical probabilities presented in this section. There are 7 groups of cars with 3 categories representing levels of predicted reliability: Below average, Average and Above average, which we label as categories  $C_1$ ,  $C_2$  and  $C_3$ , respectively. The data are given in Table 3.1. We are interested in the event that the next observation of group  $j$ , for  $j \in \{1, \dots, 7\}$ , is less than (or equal to) the next observation from each other group, so this event indicates lower predicted reliability for the next car from group  $j$ . Using the results in this section, and with the appropriate assumption  $A_{(n^j)}$  per group, the lower, empirical and upper probabilities are given in Table 3.2.

Table 3.2 shows that group 5 (Coupe) has the largest lower probability and upper probability to give the smallest next observation, ( so the next car from group coupe

Group $j$	Cars type	Predicted reliability			Total
		Below	Average	Above	
1	Small	4	3	10	17
2	Medium	4	19	9	32
3	Large	1	8	4	13
4	Sporty	5	3	6	14
5	Coupe	4	3	2	9
6	Minivan	6	5	3	14
7	SUV	3	10	5	18

**Table 3.1:** Data: Predicted reliability for 7 type of cars

Group $j$	$\underline{P}_j^<$	$EP_j^<$	$\overline{P}_j^<$	$\Delta_j^<$	$\underline{P}_j^{\leq}$	$EP_j^{\leq}$	$\overline{P}_j^{\leq}$	$\Delta_j^{\leq}$
1	0.0205	0.0324	0.0460	0.0255	0.2377	0.2598	0.3060	0.0683
2	0.0097	0.0156	0.0236	0.0139	0.1664	0.1966	0.2363	0.0699
3	0.0059	0.0093	0.0209	0.0150	0.1159	0.1472	0.2224	0.1065
4	0.0369	0.0585	0.0773	0.0404	0.3557	0.3924	0.4392	0.0835
5	0.0534	0.0845	0.1077	0.0543	0.4401	0.5078	0.5647	0.1246
6	0.0501	0.0793	0.1007	0.0506	0.4417	0.4945	0.5385	0.0968
7	0.0136	0.0216	0.0334	0.0198	0.2023	0.2370	0.2913	0.0890

**Table 3.2:** Selecting one group: NPI lower and upper probabilities (Example 3.1)

is more likely to has lower predicted reliability), and also has the largest imprecision  $\Delta_5^< = \overline{P}_5^< - \underline{P}_5^<$ , namely  $\Delta_5^< = 0.0543$  and  $\Delta_5^{\leq} = 0.1246$ . This large imprecision reflects the fact that this group has the smallest number of observations, whereas the smallest imprecision is for the second group in the case ( $<$ ). In the case ( $\leq$ ) the first group has the smallest imprecision  $\Delta_1^{\leq} = \overline{P}_1^{\leq} - \underline{P}_1^{\leq} = 0.0683$ .

△

**Example 3.2.** The data of this example are given in Table 3.3 and were used by Simonoff [59] to study statistical methods for tables with ordered categories. The data describe the relationship between the condition of books and the strength of the paper in the books. The rows are ordered in increasing strength of the paper, whereas the columns are ordered in increasing level of deterioration. Simonoff fitted several

models to these data. For example, the column effects model which indicated that books in worse condition have weaker paper. This example illustrates our methods when there are large numbers of observations. Most of the data are in one category, which leads to big differences in the NPI lower and upper probabilities for ( $<$ ) and ( $\leq$ ) cases.

Group $j$	Strength	Degree of book deterioration				Total
		Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	1 fold	181	14	18	43	256
2	2-4 folds	140	6	1	15	162
3	5-15 folds	44	2	0	0	46
4	> 15 folds	369	7	0	0	376

**Table 3.3:** Data: degree of book deterioration and strength of the paper

Group $j$	$\underline{P}_j^<$	$EP_j^<$	$\overline{P}_j^<$	$\Delta_j^<$	$\underline{P}_j^{\leq}$	$EP_j^{\leq}$	$\overline{P}_j^{\leq}$	$\Delta_j^{\leq}$
1	0.000075	0.000078	0.000136	0.000061	0.7043	0.7070	0.7082	0.0039
2	0.00019	0.00020	0.00035	0.00016	0.8589	0.8642	0.8651	0.0062
3	0.00068	0.00071	0.00085	0.00017	0.9362	0.9566	0.9575	0.0213
4	0.00164	0.00169	0.00262	0.00098	0.9788	0.9814	0.9815	0.0027

**Table 3.4:** Selection one group: NPI lower and upper probabilities (Example 3.2)

Table 3.4 gives the NPI lower and upper probabilities for the events that the next observation of the selected group  $j$  is less than (or equal to) the next observation from each of the other groups. It shows that the  $\underline{P}_j^<$  are all close to zero, because most observations in all groups are in the first category, whereas most  $\underline{P}_j^{\leq}$  and  $\overline{P}_j^{\leq}$  are close to one, also because most observations in all groups are in the first category. For example, if we consider the first group,  $\underline{P}_1^{\leq} < EP_1^{\leq} < \overline{P}_1^{\leq}$  are all close to 0.7, as there are 181 observations from 256 in the first category. These lower, empirical and upper probabilities increase as the number of observations in the first category increases, for example  $\underline{P}_4^{\leq} < EP_4^{\leq} < \overline{P}_4^{\leq}$  are all close to 0.98. Of course, if all the observations for one group  $j$  are in the first category, then  $\overline{P}_j^{\leq} = EP_j^{\leq} = 1$  while  $\underline{P}_j^{\leq}$

would be closed to one. Group 1 is most unlikely to give the next observation smaller than that for all other group (the next book with good condition), while group 4 (> 15 folds) has the largest lower and upper probabilities to give such a smallest next observation (so the next book with more than 15 folds is more likely to be in good condition). The smallest imprecision appears for the event  $X_{n^1+1}^1 < \min_{l=1,\dots,4, l \neq 1} X_{n^l+1}^l$  as both lower and upper probabilities for the first group are close to zero, and the largest imprecision is for the event  $X_{n^3+1}^3 = \min_{l=1,\dots,4} X_{n^l+1}^l$  which reflects the small number of observations in the third group.

△

**Example 3.3.** Table 3.5 comes from the same study of book deterioration in Example 3.2 but now the rows represent preservation strength. This example illustrates our methods when there are some groups with small numbers of observations and others with large numbers of observations.

Group $j$	Strength	Degree of book deterioration				Total
		Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	Repair	27	1	2	0	30
2	Microfilm	50	6	3	34	93
3	Restore	7	0	1	30	38
4	No preservation	676	22	13	0	711

**Table 3.5:** Data: degree of book deterioration and preservation strength

Table 3.6 gives the lower, empirical and upper probabilities for the event that the next observation of the selected group  $j$  is less than (or equal to) the next observation from each of the other groups, in order to compare the four groups. From Table 3.6 it can be seen that the smallest lower probability is  $\underline{P}_3^< = 0.0004$ , due to the fact that most of the observations for group 3 are in the last category, so this group is unlikely to be the best group (so the next book from restore group is unlikely to be in good condition). The largest lower probability is  $\underline{P}_4^< = 0.9509$ , as group 4 has most observations in the first category and no observations in the

last category,  $\overline{P}_4^<$  is near to one for the same reason, so group 4 is the best group when aiming at the smallest next observation (so the next book from this group is more likely to be in good condition). Moreover, the largest imprecision is  $\Delta_1^< = 0.0324$ , which reflects that the first group has the smallest number of observations, while the smallest imprecision occurs for group 4 which has the largest number of observations.

Group $j$	$\underline{P}_j^<$	$EP_j^<$	$\overline{P}_j^<$	$\Delta_j^<$	$\underline{P}_j^{\leq}$	$EP_j^{\leq}$	$\overline{P}_j^{\leq}$	$\Delta_j^{\leq}$
1	0.0158	0.0169	0.0178	0.0020	0.8719	0.9010	0.9043	0.0324
2	0.0021	0.0022	0.0030	0.0009	0.5322	0.5379	0.5429	0.0107
3	0.0004	0.00042	0.0006	0.0002	0.1795	0.1842	0.2052	0.0257
4	0.0340	0.0365	0.0483	0.0143	0.9509	0.9523	0.9530	0.0021

**Table 3.6:** Selection one group: NPI lower and upper probabilities (Example 3.3)

△

### 3.2.3 Special cases

1. If there are only two groups to be compared, so  $J = 2$ , then Equations (3.3), (3.4), (3.20), and (3.21) lead to, with  $A = [(n^1 + 1)(n^2 + 1)]^{-1}$ ,

$$\begin{aligned} \underline{P}(X_{n^1+1}^1 < X_{n^2+1}^2) &= A \left[ \sum_{k=1}^{K-1} n_k^1 \sum_{f=k+1}^K n_f^2 \right] \\ \overline{P}(X_{n^1+1}^1 < X_{n^2+1}^2) &= A \left[ \sum_{k=1}^{K-1} n_k^1 \sum_{f=k+1}^K n_f^2 + n^1 - n_K^1 + n^2 - n_1^2 + 1 \right] \\ \underline{P}(X_{n^1+1}^1 \leq X_{n^2+1}^2) &= A \left[ \sum_{k=1}^K n_k^1 \sum_{f=k}^K n_f^2 + n_1^1 + n_K^2 \right] \\ \overline{P}(X_{n^1+1}^1 \leq X_{n^2+1}^2) &= A \left[ \sum_{k=1}^K n_k^1 \sum_{f=k}^K n_f^2 + n^1 + n^2 + 1 \right] \end{aligned}$$

These results for  $J = 2$  were presented by Coolen et al [26].

2. If there are only two categories,  $K = 2$ , the NPI lower and upper probabilities presented in Theorems 3.1 and 3.2 become, with  $A = \left[ \prod_{j=1}^J (n^j + 1) \right]^{-1}$ ,

$$\underline{P}_j^< = A \left[ n_1^j \prod_{l=1, l \neq j}^J n_2^l \right]$$

$$\overline{P}_j^< = A \left[ n_1^j \prod_{l=1, l \neq j}^J (n_2^l + 1) + \prod_{l=1, l \neq j}^J (n_2^l + 1) \right] = A \left[ (n_1^j + 1) \prod_{l=1, l \neq j}^J (n_2^l + 1) \right]$$

$$\underline{P}_j^{\leq} = A \left[ n_1^j \prod_{l=1, l \neq j}^J (n^l + 1) + (n_2^j + 1) \prod_{l=1, l \neq j}^J n_2^l \right]$$

$$\overline{P}_j^{\leq} = A \left[ (n_1^j + 1) \prod_{l=1, l \neq j}^J (n^l + 1) + n_2^j \prod_{l=1, l \neq j}^J (n_2^l + 1) \right]$$

These are identical to the NPI lower and upper probabilities for comparison of proportions presented by Coolen and Coolen-Schrijner [24]. They presented the NPI lower and upper probabilities for the event that the number of successes in  $m$  future trials from one group exceeds the number of successes in  $m$  future trials from the other group, using the same latent variable representation as we do for ordinal data but with only two categories.

3. The lower probability  $\underline{P}_j^<$  as given in (3.3) is zero if either  $n_k^j = 0$  for all  $k = 1, \dots, K - 1$  (so all the observations for group  $j$  are in the last category  $C_K$ ) or if  $\prod_{l=1, l \neq j}^J n_{k+1, K}^l = 0$ , which occurs if for one or more  $l, l \neq j, n_{k+1, K}^l = 0$ , for every  $k = 1, \dots, K - 1$ . The corresponding NPI upper probability  $\overline{P}_j^<$  does not simplify greatly from Equation (3.4) for the first case, only terms with  $n_k^j = 0$  disappear from the formula. If  $n_{k+1, K}^l = 0, k = 1, \dots, K - 1$ , for all groups  $l \neq j$ , then

$$\overline{P}_j^< = A \left[ \sum_{k=1}^{K-1} n_k^j + 1 \right] \quad (3.24)$$

4. The lower probability (3.20)  $\underline{P}_j^{\leq}$  is zero in two cases. First, if  $n_k^j = 0$  for group  $j$  for  $k = 1, \dots, K - 1$  and  $n_K^l = 0$  for at least one group other than group  $j$ ; this is easily seen from the proof of Theorem 3.2 in Appendix A.1. The corresponding NPI upper probability (3.21) is

$$\overline{P}_j^{\leq} = A \left[ n_K^j \prod_{l=1, l \neq j}^J (n_K^l + 1) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \quad (3.25)$$

Secondly, if  $n_1^j = 0$  and  $\prod_{l=1, l \neq j}^J n_{k,K}^l = 0$ , for all  $k = 2, \dots, K$  (i.e.  $n_{k,K}^l = 0$  for at least one group  $l \neq j$ ), then the corresponding NPI upper probability (3.21) is

$$A \left[ \sum_{k=2}^K \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k,K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \quad (3.26)$$

5. The upper probability  $\overline{P}_j^<$  as given in (3.4) is one in the following case. Suppose the set of categories  $\{1, \dots, K\}$  is split up into  $\{1, \dots, k_i\}$  and  $\{k_i + 1, \dots, K\}$ , with  $n_k^j = 0$  if  $k \in \{k_i + 1, \dots, K\}$  and  $n_k^l = 0$ , if  $k \in \{1, \dots, k_i\}$  for all  $l = 1, \dots, j, l \neq j$ , so in this case  $n^j = \sum_{k=1}^{k_i} n_k^j$  and  $n^l = \sum_{f=k_i+1}^K n_f^l$ . Then

$$\overline{P}_j^< = A \left[ \sum_{k=1}^{k_i} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k+1}^K n_f^l + 1 \right) + \prod_{l=1, l \neq j}^J \left( \sum_{f=2}^K n_f^l + 1 \right) \right]$$

As  $n_k^l = 0$ , if  $k \in \{1, \dots, k_i\}$ , for all  $l = 1, \dots, j, l \neq j$ ,

$$\begin{aligned} \overline{P}_j^< &= A \left[ \sum_{k=1}^{k_i} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l + 1 \right) + \prod_{l=1, l \neq j}^J \left( \sum_{f=2}^K n_f^l + 1 \right) \right] \\ &= A \left[ n^j \prod_{l=1, l \neq j}^J (n^l + 1) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \\ &= A \left[ \prod_{l=1, l \neq j}^J (n^l + 1) (n^j + 1) \right] = A \left[ \prod_{j=1}^J (n^j + 1) \right] = 1 \end{aligned} \quad (3.27)$$

In this case, the corresponding NPI lower probability (3.3) is

$$\underline{P}_j^< = A \left[ \sum_{k=1}^{k_i} \left\{ n_k^j \prod_{l=1, l \neq j}^J \sum_{f=k_i+1}^K n_f^l \right\} \right] = A \left[ n^j \prod_{l=1, l \neq j}^J n^l \right] = A \left[ \prod_{j=1}^J n^j \right] \quad (3.28)$$

6. The upper probability  $\overline{P}_j^{\leq}$  as given in (3.21) is one in the following case. Using the same notation as above, suppose we have  $k_i \in \{1, \dots, K\}$  and the set is split up into  $\{1, \dots, k_i\}$  and  $\{k_i + 1, \dots, K\}$ , with  $n_k^j = 0$  if  $k \in \{k_i + 1, \dots, K\}$  and  $n_k^l = 0$  if  $k \in \{1, \dots, k_i - 1\}$ , for all  $l = 1, \dots, j, l \neq j$ . In this case  $n^j = \sum_{k=1}^{k_i} n_k^j$  and  $n^l = \sum_{f=k_i}^K n_f^l$ , and

$$\overline{P}_j^{\leq} = A \left[ \sum_{k=1}^{k_i} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l + 1 \right) + \prod_{l=1, l \neq j}^J (n^l + 1) \right]$$



As  $n_k^l = 0$ , if  $k \in \{1, \dots, k_i - 1\}$ , for all  $l = 1, \dots, j$ ,  $l \neq j$

$$\begin{aligned} \overline{P}_j^{\leq} &= A \left[ \sum_{k=1}^{k_i} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k_i}^K n_f^l + 1 \right) + \prod_{l=1, l \neq j}^J n^l + 1 \right] \\ &= A \left[ n^j \prod_{l=1, l \neq j}^J (n^l + 1) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \\ &= A \left[ \prod_{l=1, l \neq j}^J (n^l + 1) (n^j + 1) \right] = A \left[ \prod_{j=1}^J (n^j + 1) \right] = 1 \end{aligned}$$

In this case, the corresponding NPI lower probability (3.20) is

$$\begin{aligned} \underline{P}_j^{\leq} &= A \left[ n_1^j \prod_{l=1, l \neq j}^J (n^l + 1) + \sum_{k=2}^{k_i} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l \right) + \prod_{l=1, l \neq j}^J (n_K^l) \right] \\ &= A \left[ n_1^j \prod_{l=1, l \neq j}^J (n^l + 1) + \sum_{k=2}^{k_i} n_k^j \prod_{l=1, l \neq j}^J n^l + \prod_{l=1, l \neq j}^J (n_K^l) \right] \quad (3.29) \end{aligned}$$

### 3.2.4 Examples for special cases

We illustrate the lower and upper probabilities for the special cases presented in Subsection 3.2.3 by some examples. We start by considering the first special case, comparison of two groups, this is followed by several examples for the other special cases.

**Example 3.4.** We use the data set of Example 3.1 to illustrate the first special case. We consider pairwise comparison of all pairs of groups of these data, the NPI lower and upper probabilities are presented in Table 3.7.

For example, the NPI lower and upper probabilities for the event that the next observation of the first group (Small) is less than (or equal to) the next observation of the fifth group (Coupe) are  $\underline{P}(X_{18}^1 < X_{10}^5) = 0.144$  and  $\overline{P}(X_{18}^1 < X_{10}^5) = 0.217$ . In other words, the NPI lower and upper probabilities for the event that the next car from small cars group has lower predicted reliability than the next car from coupe group are 0.144 and 0.217, respectively. The NPI lower and upper probabilities for the event  $X_{18}^1 \leq X_{10}^5$  are not in this table, but these can be calculated using the conjugacy relation  $\underline{P}(A) = 1 - \overline{P}(A^c)$ , so  $\underline{P}(X_{18}^1 \leq X_{10}^5) = 1 - \overline{P}(X_{10}^5 < X_{18}^1) =$

	j	Small ( $l = 1$ )	Medium ( $l = 2$ )	Large ( $l = 3$ )	Sporty ( $l = 4$ )
Small	1		0.234, 0.256, 0.295	0.238, 0.272, 0.318	0.200, 0.227, 0.263
Medium	2	0.407, 0.445, 0.469		0.268, 0.298, 0.346	0.303, 0.360, 0.369
Large	3	0.369, 0.421, 0.460	0.217, 0.240, 0.299		0.271, 0.313, 0.362
Sporty	4	0.351, 0.399, 0.433	0.337, 0.373, 0.412	0.343, 0.396, 0.443	
Coupe	5	0.456, 0.536, 0.572	0.421, 0.483, 0.530	0.429, 0.339, 0.571	0.360, 0.429, 0.473
Minivan	6	0.474, 0.538, 0.567	0.430, 0.475, 0.511	0.438, 0.505, 0.552	0.373, 0.429, 0.467
SUV	7	0.406, 0.454, 0.485	0.278, 0.302, 0.345	0.286, 0.325, 0.384	0.305, 0.345, 0.386
	j	Coupe ( $l = 5$ )	Minivan ( $l = 6$ )	SUV ( $l = 7$ )	
Small	1	0.144, 0.169, 0.217	0.152, 0.172, 0.211	0.219, 0.225, 0.287	
Medium	2	0.176, 0.201, 0.264	0.178, 0.199, 0.244	0.247, 0.269, 0.309	
Large	3	0.150, 0.179, 0.257	0.152, 0.176, 0.238	0.207, 0.235, 0.301	
Sporty	4	0.207, 0.246, 0.300	0.218, 0.250, 0.293	0.316, 0.357, 0.400	
Coupe	5		0.273, 0.325, 0.380	0.395, 0.463, 0.516	
Minivan	6	0.2667, 0.318, 0.380		0.404, 0.456, 0.498	
SUV	7	0.184, 0.216, 0.284	0.189, 0.214, 0.267		

**Table 3.7:** Lower, empirical and upper probabilities for the event  $X_{n^j+1}^j < X_{n^l+1}^l$

$1 - 0.572 = 0.428$  and  $\bar{P}(X_{18}^1 \leq X_{10}^5) = 1 - \underline{P}(X_{10}^5 < X_{18}^1) = 1 - 0.456 = 0.544$ . It is clear from this example that the small number of observations in each group leads to large imprecision.

△

**Example 3.5.** This example illustrates the second special case when  $K = 2$ . Table 3.8 shows the number of heart operations on children under one year old at 12 medical centres, which are a part of data presented by Spiegelhalter et al [61]. This data set was also used to illustrate the NPI comparison method for proportions data presented by Coolen and Coolen-Schrijner [24] with,  $C_1$  and  $C_2$  representing the number of mortalities and number of surviving patients.

Table 3.9 shows the results of comparing multiple groups when there are two categories, so  $K = 2$ . It is clear that the largest lower probability occurs for the first group, for both events  $X_{n^1+1}^1 < \min_{l=1, \dots, J, l \neq 1} X_{n^l+1}^l$  and  $X_{n^1+1}^1 = \min_{l=1, \dots, J} X_{n^l+1}^l$ , so this group is more likely to give the smallest next observation than any other group.

Group $j$	$C_1$	$C_2$	Total	Group $j$	$C_1$	$C_2$	Total
1	43	138	181	7	27	226	253
2	27	173	200	8	57	312	369
3	26	131	157	9	28	186	214
4	15	127	142	10	31	153	184
5	36	181	217	11	67	673	740
6	49	368	417	12	32	236	268

**Table 3.8:** Heart operations mortality data for 12 medical centres

Group $j$	$\underline{P}_j^<$	$\overline{P}_j^<$	$\Delta_j^<$	$\underline{P}_j^{\leq}$	$\overline{P}_j^{\leq}$	$\Delta_j^{\leq}$
1	0.0468	0.0506	0.0038	0.3876	0.4004	0.0128
2	0.0234	0.0257	0.0023	0.2854	0.2982	0.0128
3	0.0298	0.0327	0.0029	0.3160	0.3295	0.0135
4	0.0177	0.0200	0.0023	0.2563	0.2705	0.0142
5	0.0299	0.0324	0.0026	0.3162	0.3287	0.0125
6	0.0200	0.0217	0.0017	0.2678	0.2790	0.0112
7	0.0180	0.0197	0.0017	0.2572	0.2694	0.0122
8	0.0274	0.0296	0.0022	0.3048	0.3161	0.0113
9	0.0226	0.0248	0.0474	0.2813	0.2939	0.0126
10	0.0304	0.0332	0.0028	0.3188	0.3318	0.0130
11	0.0150	0.0161	0.0011	0.2409	0.2514	0.0105
12	0.0204	0.0223	0.0019	0.2698	0.2818	0.0120

**Table 3.9:** NPI lower and upper probabilities (special case 2)

These results are equal to the results from nonparametric predictive comparison of proportions which were presented by Coolen and Coolen-Schrijner [24], as we use the same latent variable representation if  $K = 2$ .

△

**Example 3.6.** In this example we illustrate the special cases 3, 4, 5 and 6 from Subsection 3.2.3 by using the data set in Table 3.3, but with some changes applied to the data to show the special cases. Throughout this example the event of interest is that the next observation of the first group is less than (or equal to) the next

observation from each of the other groups. Table 3.10 illustrates the special case 3 when  $n_1^k = 0$  for  $k = 1, \dots, K - 1$ , so these data lead to the lower probability (3.3)  $\underline{P}_1^< = 0$ . The corresponding NPI upper probability (3.4)  $\overline{P}_1^< = 7.44 * 10^{-7}$ . Furthermore, this table is also related to special case 4 for the event  $X_{n^1+1}^1 = \min_{l=1, \dots, J} X_{n^l+1}^l$ , as  $n_k^1 = 0$  for  $k = 1, \dots, K - 1$  and  $n_K^l = 0$  for at least one group  $l$ ,  $l \neq 1$ . Therefore, the lower probability (3.20)  $\underline{P}_1^< = 0$ , with corresponding upper probability (3.25)  $\overline{P}_1^< = 0.003896$ .

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	0	0	0	256	256
2	140	6	1	15	162
3	44	2	0	0	46
4	369	7	0	0	376

**Table 3.10:** Data: degree of book deterioration (special cases 3 and 4).

Table 3.11 shows also special case 3 as  $n_{k+1,K}^l = 0, k = 1, \dots, K - 1$  for the third group, so the lower probability (3.3)  $\underline{P}_1^< = 0$  with corresponding upper probability (3.4)  $\overline{P}_1^< = 4.58 * 10^{-5}$ . Also Table 3.12 shows this special case as  $n_{k+1,K}^l = 0, k = 1, \dots, K - 1$  for all  $l$  groups, so the data in this table lead to lower probability (3.3)  $\underline{P}_1^< = 0$  with corresponding upper probability (3.24)  $\overline{P}_1^< = 2.88 * 10^{-7}$ .

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	181	14	18	43	256
2	140	6	1	15	162
3	46	0	0	0	46
4	369	7	0	0	376

**Table 3.11:** Data: degree of book deterioration (special case 3).

Table 3.13 shows the special case 4 as  $n_1^1 = 0$  and  $n_{k,K}^l = 0, k = 2, \dots, K$  for  $l = 3$ . So these data lead to lower probability (3.20)  $\underline{P}_1^< = 0$  with corresponding

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	181	14	18	43	256
2	162	0	0	0	162
3	46	0	0	0	46
4	376	0	0	0	376

**Table 3.12:** Data: degree of book deterioration (special case 3).

upper probability (3.26)  $\overline{P}_1^{\leq} = 0.00394$ .

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	0	195	18	43	256
2	140	6	1	15	162
3	46	0	0	0	46
4	369	7	0	0	376

**Table 3.13:** Data: degree of book deterioration (special case 4).

Finally, the data in Tables 3.14 and 3.15 relate to the special cases 5 and 6, respectively. For Table 3.14, the upper probability (3.4)  $\overline{P}_1^{\leq} = 1$  with corresponding lower probability (3.28)  $\underline{P}_1^{\leq} = 0.9664$ , which illustrates special case 5. For Table 3.15, the upper probability (3.21)  $\overline{P}_1^{\leq} = 1$  with corresponding lower probability (3.29)  $\underline{P}_1^{\leq} = 0.9874$ , which illustrates special case 6.

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	181	75	0	0	256
2	0	0	141	21	162
3	0	0	44	2	46
4	0	0	7	369	376

**Table 3.14:** Data: degree of book deterioration (special case 5).

Group $j$	Degree of book deterioration				Total
	Intact $C_1$	Slight $C_2$	Moderate $C_3$	Extreme $C_4$	
1	181	75	0	0	256
2	0	146	1	15	162
3	0	46	0	0	46
4	0	376	0	0	376

**Table 3.15:** Data: degree of book deterioration (special case 6).

△

### 3.3 Subset containing all best groups

In this section we consider  $J$  independent groups denoted by  $G_1, \dots, G_J$ , so no information about any aspect related to one or more group(s) contains information about any aspect of the other groups. We divide these multiple groups into two non-empty subsets of groups,  $S$  and  $S^c$ , and we apply the  $A_{(n)}$ -based inferences per group to consider one future observation for each group. We derive the NPI lower and upper probabilities for the event that all these future observations for groups in  $S$  are less (so in categories ‘further to the left’) than all the future observations for groups in  $S^c$ . Let  $S = \{j_1, \dots, j_w\} \subset \{1, \dots, J\}$  be the selected subset containing  $w$  groups, for  $1 \leq w \leq J - 1$ , and let  $S^c = \{1, \dots, J\} \setminus S$  be the subset of the not selected groups (i.e. the complementary subset to  $S$ ) which contains  $J - w$  groups.

The NPI lower and upper probabilities for the event that the next observation of each group in  $S$  is less than the next observation from each group in  $S^c$ , denoted by

$$\underline{P}_S^< = \underline{P} \left( \max_{j \in S} X_{n^j+1}^j < \min_{l \in S^c} X_{n^l+1}^l \right)$$

and

$$\overline{P}_S^< = \overline{P} \left( \max_{j \in S} X_{n^j+1}^j < \min_{l \in S^c} X_{n^l+1}^l \right)$$

These are given in Theorem 3.3, we use the following notation for the left and right configurations

$$L_k^j = \begin{cases} n_k^j + 1 & \text{if } k = 1 \\ n_k^j & \text{if } k = 2, \dots, K \end{cases} \quad (3.30)$$

$$R_k^j = \begin{cases} n_k^j & \text{if } k = 1, \dots, K - 1 \\ n_k^j + 1 & \text{if } k = K \end{cases} \quad (3.31)$$

**Theorem 3.3.** The NPI lower and upper probabilities for the event that the next observation of each group in  $S$  is less than the next observation from each group in the complementary set  $S^c$  are

$$P_S^< = \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^L (X_{n^l+1}^l \in \mathcal{C}_{M_W+1,K}) \quad (3.32)$$

$$\bar{P}_S^< = \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R (X_{n^l+1}^l \in \mathcal{C}_{M_W+1,K}) \quad (3.33)$$

where  $M_W = \max \{k_{j_1}, \dots, k_{j_w}\}$  and with the sums over  $k_j = 1, \dots, K - 1$  for all  $j \in S = \{j_1, \dots, j_w\}$ .

Before we prove these results in general, we present the proof for a case of four groups, as this is easier with regard to notation and explaining all steps well. Let there be four groups and suppose we are interested in the subset which contains groups 1 and 3, so  $S = \{1, 3\}$  and  $S^c = \{2, 4\}$ . We derive the NPI lower and upper probabilities (3.32) and (3.33) for the event that the next observations of groups 1 and 3 are both less than the next observations from both groups 2 and 4. The lower probability can be derived as follows, where we use the notation  $M_W = \max \{k_1, k_3\}$ ,

$$\begin{aligned} P \left( \max_{j=1,3} X_{n^j+1}^j < \min_{l=2,4} X_{n^l+1}^l \right) &= P \left( \bigcap_{l \in S^c} \left\{ \max_{j=1,3} X_{n^j+1}^j < X_{n^l+1}^l \right\} \right) \\ &= P \left( \left\{ X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j \right\} \cap \left\{ X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j \right\} \right) \\ &= \sum_{k_1=1}^K \sum_{k_3=1}^K P \left( \left\{ X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j \right\} \cap \left\{ X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j \right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\} \right) \\ &\quad \times P (X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}) \end{aligned}$$

$$\begin{aligned}
&\geq \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{K-1} P \left( \left\{ X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j \right\} \cap \left\{ X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j \right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\} \right) \\
&\quad \times P^R (X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}) \\
&= \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{K-1} \binom{n_{k_1}^1}{n^1+1} \binom{n_{k_3}^3}{n^3+1} P (X_{n^2+1}^2 \in \mathcal{C}_{M_{W+1},K}) P (X_{n^4+1}^4 \in \mathcal{C}_{M_{W+1},K}) \\
&\geq \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{k_1} \binom{n_{k_1}^1}{n^1+1} \binom{n_{k_3}^3}{n^3+1} P^L (X_{n^2+1}^2 \in \mathcal{C}_{k_1+1,K}) P^L (X_{n^4+1}^4 \in \mathcal{C}_{k_1+1,K}) \\
&\quad + \sum_{k_1=1}^{K-1} \sum_{k_3=k_1+1}^{K-1} \binom{n_{k_1}^1}{n^1+1} \binom{n_{k_3}^3}{n^3+1} P^L (X_{n^2+1}^2 \in \mathcal{C}_{k_3+1,K}) P^L (X_{n^4+1}^4 \in \mathcal{C}_{k_3+1,K}) \\
&= \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{k_1} \binom{n_{k_1}^1}{n^1+1} \binom{n_{k_3}^3}{n^3+1} \sum_{f=k_1+1}^K \frac{n_f^2}{n^2+1} \sum_{f=k_1+1}^K \frac{n_f^4}{n^4+1} \\
&\quad + \sum_{k_1=1}^{K-1} \sum_{k_3=k_1+1}^{K-1} \binom{n_{k_1}^1}{n^1+1} \binom{n_{k_3}^3}{n^3+1} \sum_{f=k_3+1}^K \frac{n_f^2}{n^2+1} \sum_{f=k_3+1}^K \frac{n_f^4}{n^4+1} \tag{3.34}
\end{aligned}$$

Equation (3.34) is the maximum lower bound for  $P \left( \max_{j=1,3} X_{n^j+1}^j < \min_{l=2,4} X_{n^l+1}^l \right)$  which can be achieved, so the NPI lower probability is

$$\begin{aligned}
&\underline{P}_{\{1,3\}}^< = \\
&A \left[ \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{k_1} (n_{k_1}^1)(n_{k_3}^3) \prod_{l=2,4} \left( \sum_{f=k_1+1}^K n_f^l \right) + \sum_{k_1=1}^{K-1} \sum_{k_3=k_1+1}^{K-1} (n_{k_1}^1)(n_{k_3}^3) \prod_{l=2,4} \sum_{f=k_3+1}^K n_f^l \right] \tag{3.35}
\end{aligned}$$

with  $A = \left[ \prod_{j=1}^4 (n^j + 1) \right]^{-1}$ . In the derivation of this lower bound, we use the assumption  $A_{(n^j)}$  for group  $j$ , with assumed independence of the groups. The first inequality follows by putting the probability mass per interval for groups 1 and 3 at the right endpoints. The term for category  $K$  disappears from the summation because if  $X_{n^j+1}^j \in C_K, j = 1, 3$ , then the probability of the next observations of group 2 and 4 being greater than the next observations from both groups 1 and 3 is zero. The second inequality contains two terms; one when  $k_1 \geq k_3$  and the other when  $k_1 < k_3$ . Moreover, we get Equation (3.34) by using the left configuration of the latent variable representations for groups 2 and 4.

The corresponding upper probability for this event is again derived similarly but



by putting the probability masses at the opposite endpoints,

$$\begin{aligned}
P\left(\max_{j=1,3} X_{n^j+1}^j < \min_{l=2,4} X_{n^l+1}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{\max_{j \in S} X_{n^j+1}^j < X_{n^l+1}^l\right\}\right) \\
&= P\left(\left\{X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j\right\}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\}\right) \\
&\quad \times P\left(X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}\right) \\
&\leq \sum_{k_1=1}^{K-1} \sum_{k_3=1}^{K-1} P\left(\left\{X_{n^2+1}^2 > \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 > \max_{j=1,3} X_{n^j+1}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\}\right) \\
&\quad \times P^L\left(X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}\right) \\
&= \left[ \frac{n_1^1+1}{n^1+1} \left( \frac{n_1^3+1}{n^3+1} + \sum_{k_3=2}^{K-1} \left( \frac{n_{k_3}^3}{n^3+1} \right) \right) + \sum_{k_1=2}^{K-1} \left( \frac{n_{k_1}^1}{n^1+1} \right) \left( \frac{n_1^3+1}{n^3+1} + \sum_{k_3=2}^{K-1} \left( \frac{n_{k_3}^3}{n^3+1} \right) \right) \right] \\
&\quad \times P\left(X_{n^2+1}^2 \in \mathcal{C}_{M_W+1,K}\right) P\left(X_{n^4+1}^4 \in \mathcal{C}_{M_W+1,K}\right) \\
&= \left[ \left( \frac{n_1^1+1}{n^1+1} \right) \left( \frac{n_1^3+1}{n^3+1} \right) + \left( \frac{n_1^1+1}{n^1+1} \right) \sum_{k_3=2}^{K-1} \left( \frac{n_{k_3}^3}{n^3+1} \right) + \left( \frac{n_1^3+1}{n^3+1} \right) \sum_{k_1=2}^{K-1} \left( \frac{n_{k_1}^1}{n^1+1} \right) \right. \\
&\quad \left. + \sum_{k_1=2}^{K-1} \sum_{k_3=2}^{K-1} \left( \frac{n_{k_1}^1}{n^1+1} \right) \left( \frac{n_{k_3}^3}{n^3+1} \right) \right] P\left(X_{n^2+1}^2 \in \mathcal{C}_{M_W+1,K}\right) P\left(X_{n^4+1}^4 \in \mathcal{C}_{M_W+1,K}\right) \quad (*)
\end{aligned} \tag{3.36}$$

In the next step Equation (3.36) becomes more complicated as we will take the right configuration for groups 2 and 4, so it is useful to use the notations  $L_{k_j}^j$  and  $R_{k_j}^j$  in the rest of the proof as follows:

$$\begin{aligned}
(*) &\leq \left[ \prod_{j=1,3} \left( \frac{L_1^j}{n^j+1} \right) + \left( \frac{L_1^3}{n^3+1} \right) \sum_{k_1=2}^{K-1} \left( \frac{L_{k_1}^1}{n^1+1} \right) + \sum_{k_1=2}^{K-1} \sum_{k_3=2}^{k_1} \prod_{j=1,3} \left( \frac{L_{k_j}^j}{n^j+1} \right) \right] \\
&\quad \times P^R\left(X_{n^2+1}^2 \in \mathcal{C}_{k_1+1,K}\right) P^R\left(X_{n^4+1}^4 \in \mathcal{C}_{k_1+1,K}\right) \\
&\quad + \left[ \left( \frac{L_1^j}{n^1+1} \right) \sum_{k_3=2}^{K-1} \left( \frac{L_{k_3}^3}{n^3+1} \right) + \sum_{k_1=2}^{K-1} \sum_{k_3=k_1+1}^{K-1} \prod_{j=1,3} \left( \frac{L_{k_j}^j}{n^j+1} \right) \right] \\
&\quad \times P^R\left(X_{n^2+1}^2 \in \mathcal{C}_{k_3+1,K}\right) P^R\left(X_{n^4+1}^4 \in \mathcal{C}_{k_3+1,K}\right)
\end{aligned}$$

$$\begin{aligned}
&= \prod_{j=1,3} \left( \frac{L_1^j}{n^j + 1} \right) \prod_{l=2,4} \left( \sum_{f=k_1+1}^K \frac{R_f^l}{n^l + 1} \right) \\
&\quad + \left( \frac{L_1^3}{n^3 + 1} \right) \sum_{k_1=2}^{K-1} \left( \frac{L_{k_1}^1}{n^1 + 1} \right) \left( \prod_{l=2,4} \left( \sum_{f=k_1+1}^K \frac{R_f^l}{n^l + 1} \right) \right) \\
&\quad + \sum_{k_1=2}^{K-1} \sum_{k_3=2}^{k_1} \prod_{j=1,3} \left( \frac{L_{k_j}^j}{n^j + 1} \right) \left( \prod_{l=2,4} \left( \sum_{f=k_1+1}^K \frac{R_f^l}{n^l + 1} \right) \right) \\
&\quad + \left( \frac{L_1^j}{n^1 + 1} \right) \sum_{k_3=2}^{K-1} \left( \frac{L_{k_3}^3}{n^3 + 1} \right) \left( \prod_{l=2,4} \left( \sum_{f=k_3+1}^K \frac{R_f^l}{n^l + 1} \right) \right) \\
&\quad + \sum_{k_1=2}^{K-1} \sum_{k_3=k_1+1}^{K-1} \prod_{j=1,3} \left( \frac{L_{k_j}^j}{n^j + 1} \right) \left( \prod_{l=2,4} \left( \sum_{f=k_3+1}^K \frac{R_f^l}{n^l + 1} \right) \right) \tag{3.37}
\end{aligned}$$

Formula (3.37) is the minimal upper bound which can be achieved for

$P \left( \max_{j=1,3} X_{n^j+1}^j < \min_{l=2,4} X_{n^l+1}^l \right)$ , so the upper probability is

$$\begin{aligned}
\bar{P}_{\{1,3\}}^< &= A \left[ \left( \prod_{j=1,3} L_1^j \right) \prod_{l=2,4} \left( \sum_{f=k_1+1}^K R_f^l \right) + L_1^3 \sum_{k_1=2}^{K-1} L_{k_1}^1 \prod_{l=2,4} \left( \sum_{f=k_1+1}^K R_f^l \right) \right. \\
&\quad + \sum_{k_1=2}^{K-1} \sum_{k_3=2}^{k_1} \prod_{j=1,3} \left( L_{k_j}^j \right) \left( \prod_{l=2,4} \left( \sum_{f=k_1+1}^K R_f^l \right) \right) + L_1^1 \sum_{k_3=2}^{K-1} L_{k_3}^3 \prod_{l=2,4} \left( \sum_{f=k_3+1}^K R_f^l \right) \\
&\quad \left. + \sum_{k_1=2}^{K-1} \sum_{k_3=k_1+1}^{K-1} \prod_{j=1,3} \left( L_{k_j}^j \right) \prod_{l=2,4} \left( \sum_{f=k_1+1}^K R_f^l \right) \right] \tag{3.38}
\end{aligned}$$

again with  $A = \left[ \prod_{j=1}^4 (n^j + 1) \right]^{-1}$ .

We now prove Theorem 3.3 for the general case, along the same lines as above.

*Proof.* We derive the NPI lower probability (3.32), with  $M_W = \max \{k_{j_1}, \dots, k_{j_w}\}$ .

$$\begin{aligned}
P \left( \max_{j \in S} X_{n^j+1}^j < \min_{l \in S^c} X_{n^l+1}^l \right) &= P \left( \bigcap_{l \in S^c} \left\{ X_{n^l+1}^l > \max_{j \in S} X_{n^j+1}^j \right\} \right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P \left( \bigcap_{l \in S^c} \left\{ X_{n^l+1}^l > \max_{j \in S} X_{n^j+1}^j \right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S \right) \right. \\
&\quad \left. \times \prod_{j \in S} P \left( X_{n^j+1}^j \in C_{k_j} \right) \right]
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[ P \left( \bigcap_{l \in S^c} \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\} \mid X_{n^{j+1}}^j \in C_{k_j}, j \in S \right) \right. \\
&\quad \left. \times \prod_{j \in S} P^R (X_{n^{j+1}}^j \in C_{k_j}) \right] \\
&= \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P \left( \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\} \mid X_{n^{j+1}}^j \in C_{k_j}, j \in S \right)
\end{aligned} \tag{3.39}$$

$$\begin{aligned}
&= \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P (X_{n^{l+1}}^l \in \mathcal{C}_{M_{W+1}, K}) \\
&\geq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^L (X_{n^{l+1}}^l \in \mathcal{C}_{M_{W+1}, K})
\end{aligned} \tag{3.40}$$

We stop at this step, as many terms have to be written to get the final formula. However, this can be easily computed with the statistical software R. In this derivation, we use the assumptions  $A_{(n^j)}$  for all groups with assumed independence of the  $J$  groups. The term for categories  $C_K$  disappears from the summations because for any combination, if one or more  $X_{n^{j+1}}^j$ ,  $j \in S$ , are in the last category  $C_K$ , then the probability for the event  $\bigcap_{l \in S^c} \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\}$  is zero. Inequality (3.39) follows by using the right configuration for groups  $j \in S$ , so by putting the probability masses per interval for all groups in  $S$  at the right endpoints and for all groups in  $S^c$  at the left endpoints in the latent variable representation, using Equations (3.30) and (3.31). The corresponding NPI upper probability (3.33) is

$$\begin{aligned}
&P \left( \max_{j \in S} X_{n^{j+1}}^j < \min_{l \in S^c} X_{n^{l+1}}^l \right) = P \left( \bigcap_{l \in S^c} \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\} \right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P \left( \bigcap_{l \in S^c} \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\} \mid X_{n^{j+1}}^j \in C_{k_j}, j \in S \right) \right. \\
&\quad \left. \times \prod_{j \in S} P (X_{n^{j+1}}^j \in C_{k_j}) \right] \\
&\leq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[ P \left( \bigcap_{l \in S^c} \left\{ X_{n^{l+1}}^l > \max_{j \in S} X_{n^{j+1}}^j \right\} \mid X_{n^{j+1}}^j \in C_{k_j}, j \in S \right) \right. \\
&\quad \left. \times \prod_{j \in S} P^L (X_{n^{j+1}}^j \in C_{k_j}) \right]
\end{aligned}$$

$$= \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P(X_{n^l+1}^l \in \mathcal{C}_{M_W+1,K}) \quad (3.41)$$

$$\leq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R(X_{n^l+1}^l \in \mathcal{C}_{M_W+1,K}) \quad (3.42)$$

This upper probability is also derived by using Equations (3.30) and (3.31) with the assumed independence between different groups. Inequality (3.41) follows by putting the probability mass per interval at left endpoints for groups  $j \in S$  in the latent variable representation. Equation (3.42) is quite complicated as we have to consider all combinations where one or more  $X_{n^j+1}^j$ ,  $j \in S$ , are in the first category, but it can easily be computed using the statistical software R. Because the expressions in (3.40) and (3.42) are sharp, in the sense that they are attained for the specified configurations, they are the optimal lower and upper bounds for the probability of interest under the assumptions made, and hence they are the NPI lower and upper probabilities for the event considered.  $\square$

It can also be of interest to consider the event that the next observation of each group in  $S$  is less than or equal to the next observation from each group in the complementary set  $S^c$ . We denote the NPI lower and upper probabilities for this event by

$$\underline{P}_S^{\leq} = \underline{P} \left( \max_{j \in S} X_{n^j+1}^j \leq \min_{l \in S^c} X_{n^l+1}^l \right)$$

and

$$\overline{P}_S^{\leq} = \overline{P} \left( \max_{j \in S} X_{n^j+1}^j \leq \min_{l \in S^c} X_{n^l+1}^l \right)$$

These are presented in Theorem 3.4.

**Theorem 3.4.** The NPI lower and upper probabilities for the event that the next observation of each group in  $S$  is less than or equal to the next observation from each group in the complementary set  $S^c$  are

$$\underline{P}_S^{\leq} = \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^L(X_{n^l+1}^l \in \mathcal{C}_{M_W,K}) \quad (3.43)$$

$$\overline{P}_S^{\leq} = \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R(X_{n^l+1}^l \in \mathcal{C}_{M_W,K}) \quad (3.44)$$

where  $M_W = \max\{k_{j_1}, \dots, k_{j_w}\}$  and with the sums over  $k_j = 1, \dots, K$  for all  $j \in \{j_1, \dots, j_w\}$ .

*Proof.* The lower probability (3.43) is derived as follows:

$$\begin{aligned}
P\left(\max_{j \in S} X_{n^j+1}^j \leq \min_{l \in S^c} X_{n^l+1}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j\right\}\right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \right. \\
&\quad \left. \times \prod_{j \in S} P(X_{n^j+1}^j \in C_{k_j}) \right] \\
&\geq \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \right. \\
&\quad \left. \times \prod_{j \in S} P^R(X_{n^j+1}^j \in C_{k_j}) \right] \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j+1} P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j+1} \prod_{l \in S^c} P(X_{n^l+1}^l \in \mathcal{C}_{M_W, K}) \\
&\geq \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j+1} \prod_{l \in S^c} P^L(X_{n^l+1}^l \in \mathcal{C}_{M_W, K}) \tag{3.45}
\end{aligned}$$

The upper probability (3.44) is derived as follows:

$$\begin{aligned}
P\left(\max_{j \in S} X_{n^j+1}^j \leq \min_{l \in S^c} X_{n^l+1}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j\right\}\right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right\}\right) \right. \\
&\quad \left. \times \prod_{j \in S} P(X_{n^j+1}^j \in C_{k_j}) \right] \\
&\leq \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right\}\right) \right. \\
&\quad \left. \times \prod_{j \in S} P^L(X_{n^j+1}^j \in C_{k_j}) \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P \left( \bigcap_{l \in S^c} \left\{ X_{n^l+1}^l \geq \max_{j \in S} X_{n^j+1}^j \mid X_{n^j+1}^j \in C_{k_j}, j \in S \right\} \right) \right] \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P(X_{n^l+1}^l \in \mathcal{C}_{M_W, K}) \\
&\leq \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R(X_{n^l+1}^l \in \mathcal{C}_{M_W, K}) \tag{3.46}
\end{aligned}$$

□

These lower and upper probabilities are derived similarly to the lower and upper probabilities (3.32) and (3.33) but with  $<$  everywhere replaced by  $\leq$ , and the  $K$  terms in the summation that were equal to zero in (3.32) and (3.33) are now not equal to zero, giving some additional terms in the summation. Equations (3.45) and (3.46) can be easily computed using R.

To explain the notation and steps in this proof in more detail, we present the case with four groups in Appendix A.2.

### 3.4 Subset containing the best group

An alternative event that may be of interest in multiple comparisons is that the next observation of at least one group in  $S = \{j_1, \dots, j_w\} \subset \{1, \dots, J\}$  is less than the next observation of each group in  $S^c$ , so that  $S$  contains the group with the minimal next observation. The NPI lower and upper probabilities for this event are denoted by

$$\underline{P}_{S^1}^< = \underline{P} \left( \min_{j \in S} X_{n^j+1}^j < \min_{l \in S^c} X_{n^l+1}^l \right)$$

and

$$\overline{P}_{S^1}^< = \overline{P} \left( \min_{j \in S} X_{n^j+1}^j < \min_{l \in S^c} X_{n^l+1}^l \right)$$

These lower and upper probabilities are given in Theorem 3.5, using notation  $L_{k_j}^j$  and  $R_{k_j}^j$  as defined in (3.30) and (3.31).

**Theorem 3.5.** The NPI lower and upper probabilities for the event that the next observation for at least one of the selected groups is less than the next observation

from each group in the complementary set  $S^c$  are

$$\underline{P}_{S^1}^< = \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^L (X_{n^l+1}^l \in \mathcal{C}_{M_w+1,K}) \quad (3.47)$$

$$\overline{P}_{S^1}^< = \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R (X_{n^l+1}^l \in \mathcal{C}_{M_w+1,K}) \quad (3.48)$$

where  $M_w = \min \{k_{j_1}, \dots, k_{j_w}\}$  and the sums are over  $k_j = 1, \dots, K$  for all  $j \in \{j_1, \dots, j_w\}$ .

*Proof.* The proof of this theorem is similar to the proof of Theorem 3.3, with ‘max’ replaced by ‘min’ everywhere and ‘ $M_W$ ’ replaced by ‘ $M_w$ ’. Moreover, here all  $K$  terms remain in the summation, as there is only one term which is equal to zero, namely when the next observation from all groups in  $S$  are in the last category  $C_K$ .  $\square$

The corresponding NPI lower and upper probabilities for the event that the next observation of at least one group in  $S$  is less than or equal to the next observation of each group in  $S^c$ , denoted by  $\underline{P}_{S^1}^<$  and  $\overline{P}_{S^1}^<$ , respectively, are similar to (3.47) and (3.48) but with  $M_w + 1$  replaced by  $M_w$  in the events involving the groups in  $S^c$  in the probabilities on the right-hand side.

For more detailed explanation, we give the proof of Theorem 3.5 for the case with four groups,  $S = \{1, 3\}$  and  $S^c = \{2, 4\}$ , in Appendix A.3.

**Example 3.7.** The data in Table 3.16, taken from [1, 2], refer to a clinical trial involving 802 patients who experienced trauma due to sub-arachnoid haemorrhage (SAH). There are four treatment groups ( $J = 4$ ), representing a control group and three groups corresponding to different dose levels. The Glasgow outcome scale is presented by five ordered categories ( $K = 5$ ).

The inequality-constrained test discussed in Section 2.1 is applied to test the null hypothesis  $H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4$  against  $H_a : \pi_1 \geq \pi_2 \geq \pi_3 \geq \pi_4$ , with categories  $C_2 - C_5$  are combined to illustrate the inequality-constrained test for binomial parameters. The test statistic, as given in Equation (2.1) is  $G^2(I|O) = 3.27$  and has  $p$ -value 0.095, so the null hypothesis is not rejected if significance level 0.05

Treatment Group ( $j$ )	Glasgow outcome scale					Total
	Death	Vegetative State	Major Disability	Minor Disability	Good Recovery	
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	
Placebo (1)	59	25	46	48	32	210
Low dose (2)	48	21	44	47	30	190
Medium dose (3)	44	14	54	64	31	207
High dose (4)	43	4	49	58	41	195

**Table 3.16:** Sub-Arachnoid Haemorrhage (SAH) data

is used. For the full  $4 \times 5$  table the inequality-constrained test shows evidence of association with  $p$ -value = 0.002 [1].

We apply our NPI method by considering explicitly the next patient for each group. We use this data set to illustrate the NPI lower and upper probabilities for the events discussed in this section and Sections 3.3.

Table 3.17 presents the NPI lower and upper probabilities (3.32), (3.33), (3.43) and (3.44) for the event that the next observation of each group in  $S$  is less than (or equal to) the next observation of all groups in  $S^c$ , so  $S$  contains the groups which provide the smallest next observations, which in this example actually implies the worst outcomes in real-world terms. We present the results for all subsets  $S$  to illustrate the new approach presented in this chapter. Of course, some such subsets may not be of much practical interest, for example it may be more logical to consider subsets consisting of neighbouring groups than other subsets, restricting attention to such subsets, if deemed appropriate, is straightforward. For  $S$  containing only a single group, group 4 (High dose) has the smallest lower and upper probabilities of providing the minimal next observation (so worst outcome) while group 1 (Placebo) has the largest lower and upper probabilities for this event. The NPI lower and upper probabilities for these events are not monotone if  $S$  increases, which is logical as the events corresponding to increasing subset  $S$  are not such that one implies the other. As the event  $\max_{j \in S} X_{nj+1}^j < \min_{l \in S^c} X_{n'+1}^l$  implies the event  $\max_{j \in S} X_{nj+1}^j \leq \min_{l \in S^c} X_{n'+1}^l$  and because there are multiple observations of each group in each category, the NPI lower and upper probabilities  $\underline{P}_S^<$  and  $\overline{P}_S^<$  are greater than  $\underline{P}_S^{\leq}$  and  $\overline{P}_S^{\leq}$ . The imprecision



is less than 0.01 for all events considered, reflecting the substantial numbers of data for all four groups.

Subset ( $S$ )	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$
{1}	0.1883	0.1947	0.0064	0.4298	0.4378	0.0082
{2}	0.1661	0.1721	0.0060	0.3958	0.4044	0.0086
{3}	0.1281	0.1332	0.0051	0.3380	0.3460	0.0080
{4}	0.1155	0.1204	0.0049	0.3171	0.3247	0.0076
{1, 2}	0.1441	0.1497	0.0056	0.3224	0.3317	0.0093
{1, 3}	0.1164	0.1214	0.0050	0.2818	0.2905	0.0087
{1, 4}	0.0977	0.1021	0.0044	0.2531	0.2572	0.0041
{2, 3}	0.1051	0.1097	0.0046	0.2618	0.2700	0.0082
{2, 4}	0.0879	0.0920	0.0041	0.2312	0.2389	0.0077
{3, 4}	0.0721	0.0757	0.0036	0.2011	0.2081	0.0070
{1, 2, 3}	0.1935	0.2008	0.0073	0.4394	0.4490	0.0096
{1, 2, 4}	0.1477	0.1539	0.0062	0.3719	0.3814	0.0095
{1, 3, 4}	0.1356	0.1417	0.0061	0.3388	0.3478	0.0090
{2, 3, 4}	0.1266	0.1322	0.0056	0.3185	0.3269	0.0084

**Table 3.17:** SAH data:  $\max_{j \in S} X_{n^j+1}^j < (\leq) \min_{l \in S^c} X_{n^l+1}^l$

Before we consider other events, we introduce a variation to this data set to illustrate our approach further. The changed SAH data set in Table 3.18 has the same numbers per group as the original data in Table 3.16, but the numbers per category are changed such that for groups 1 and 2 the numbers in  $C_1$  and  $C_2$  are substantially increased while those in  $C_4$  and  $C_5$  are decreased, and for groups 3 and 4 this change is the other way around. This leads to the outcomes being far worse

Group ( $j$ )	Glasgow outcome scale					Total
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	
Placebo (1)	89	55	46	8	12	210
Low dose (2)	78	41	44	17	10	190
Medium dose (3)	5	4	54	74	70	207
High dose (4)	3	4	49	78	61	195

**Table 3.18:** Changed SAH data

for groups 1 and 2 than for groups 3 and 4, which is clearly reflected by the NPI lower and upper probabilities presented in Table 3.19.

Subset ( $S$ )	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$
{1}	0.3391	0.3493	0.0102	0.6442	0.6537	0.0095
{2}	0.2953	0.3047	0.0094	0.5958	0.6051	0.0093
{3}	0.0129	0.0151	0.0022	0.0630	0.0694	0.0064
{4}	0.0100	0.0122	0.0022	0.0537	0.0603	0.0066
{1, 2}	0.5755	0.5897	0.0142	0.7953	0.8100	0.0147
{1, 3}	0.0427	0.0467	0.0040	0.1488	0.1568	0.0080
{1, 4}	0.0396	0.0436	0.0040	0.1430	0.1511	0.0081
{2, 3}	0.0326	0.0361	0.0035	0.1200	0.1271	0.0071
{2, 4}	0.0305	0.0340	0.0035	0.1142	0.1215	0.0073
{3, 4}	0.0025	0.0029	0.0004	0.0172	0.0188	0.0016
{1, 2, 3}	0.2786	0.2880	0.0094	0.6017	0.6126	0.0109
{1, 2, 4}	0.2856	0.2952	0.0096	0.6090	0.6195	0.0105
{1, 3, 4}	0.0293	0.0324	0.0031	0.1082	0.1150	0.0068
{2, 3, 4}	0.0272	0.0299	0.0027	0.0894	0.0952	0.0058

**Table 3.19:** Changed SAH data:  $\max_{j \in S} X_{n^j+1}^j < (\leq) \min_{l \in S^c} X_{n^l+1}^l$

The imprecision varies now more for the different events than was the case for the original data, even though the numbers of observations are the same. This is due mostly to the fact that some of the lower and upper probabilities are now very small while others are closer to 0.5. Typically, if corresponding lower and upper probabilities are both close to either 0 or 1, the imprecision tends to be smaller than for lower and upper probabilities which are close to 0.5, when based on similar numbers of data. One might compare this typical behaviour of imprecision with the variance  $p(1-p)$  for a single Bernoulli-distributed random quantity in the classical setting, if the success probability  $p$  is known. This variance is maximal for  $p = 0.5$  and decreases to zero when  $p$  goes to 0 or 1. However, imprecision reflects lack of information which is different to the stochastic variability reflected by the variance.

One detail of Table 3.19 that is of interest is seen by comparing the lower and upper probabilities for subsets including either group 3 or group 4. When considering  $S$  of size 1 or 2, the lower and upper probabilities for  $S$  with group 3 included are

slightly larger than those with group 4 included. However, for  $S = \{1, 2, 3\}$  and  $S = \{1, 2, 4\}$  the effect is the other way around, so adding group 3 to groups 1 and 2 increases these lower and upper probabilities less than adding group 4. This is just a consequence of the detailed data, such an effect did not occur for the original data above. It shows again that there are no straightforward monotonicities in this approach, of course this is a reflection of the data.

Subset ( $S$ )	$\underline{P}_{S^1}^<$	$\overline{P}_{S^1}^<$	$\Delta_{S^1}^<$	$\underline{P}_{S^1}^{\leq}$	$\overline{P}_{S^1}^{\leq}$	$\Delta_{S^1}^{\leq}$
{1}	0.1883	0.1947	0.0064	0.4298	0.4378	0.0082
{2}	0.1661	0.1721	0.0060	0.3958	0.4044	0.0086
{3}	0.1281	0.1332	0.0051	0.3380	0.3460	0.0080
{4}	0.1155	0.1204	0.0049	0.3171	0.3247	0.0076
{1, 2}	0.4204	0.4296	0.0092	0.7090	0.7173	0.0083
{1, 3}	0.3705	0.3796	0.0091	0.6641	0.6727	0.0086
{1, 4}	0.3545	0.3635	0.0090	0.6481	0.6568	0.0087
{2, 3}	0.3432	0.3519	0.0087	0.6365	0.6455	0.0090
{2, 4}	0.3273	0.3359	0.0086	0.6204	0.6295	0.0091
{3, 4}	0.2827	0.2910	0.0083	0.5704	0.5796	0.0092
{1, 2, 3}	0.6753	0.6829	0.0076	0.8796	0.8845	0.0049
{1, 2, 4}	0.6540	0.6620	0.0080	0.8668	0.8719	0.0051
{1, 3, 4}	0.5956	0.6042	0.0086	0.8279	0.8339	0.0060
{2, 3, 4}	0.5620	0.5702	0.0082	0.8053	0.8117	0.0064

**Table 3.20:** SAH data:  $\min_{j \in S} X_{n^j+1}^j < (\leq) \min_{l \in S^c} X_{n^l+1}^l$

Table 3.20 presents the NPI lower and upper probabilities for the event that the next observation for at least one of the groups in  $S$  is less than (or equal to) the next observation of each group in  $S^c$ , so this means that  $S$  contains the group with the worst next outcome and this illustrates the results presented in this section. Table 3.20 corresponds to the original SAH data in Table 3.16. Table 3.21 presents the NPI lower and upper probabilities for the same events but with the changed SAH data from Table 3.18. The lower and upper probabilities in these tables are of course monotonously increasing if  $S$  is expanded, as for example the event that the subset  $\{1, 2\}$  contains the smallest next observation implies that this also holds for the subset  $\{1, 2, 3\}$ . For these events the conjugacy property is nicely illustrated,

for example in Table 3.20 the NPI lower probability for the event that  $S = \{1, 2\}$  contains a group leading to the smallest next observation (with equal observations for the groups in  $S^c$  allowed) is equal to 0.7090, and the NPI upper probability for the event that  $S = \{3, 4\}$  contains the strictly smallest next observation is equal to  $0.2910 = 1 - 0.7090$ . For  $S$  containing only a single group, the values in Tables 3.20 and 3.21 are equal to the corresponding values in Tables 3.17 and 3.19 as the events are identical. The change in the data is clearly reflected in the different values in these two tables.

Subset ( $S$ )	$\underline{P}_{S^1}^<$	$\overline{P}_{S^1}^<$	$\Delta_{S^1}^<$	$\underline{P}_{S^1}^<$	$\overline{P}_{S^1}^<$	$\Delta_{S^1}^<$
{1}	0.3391	0.3493	0.0102	0.6442	0.6537	0.0095
{2}	0.2953	0.3047	0.0094	0.5958	0.6051	0.0093
{3}	0.0129	0.0150	0.0021	0.0630	0.0694	0.0064
{4}	0.0100	0.0122	0.0022	0.0537	0.0603	0.0066
{1, 2}	0.8817	0.8932	0.0115	0.9712	0.9756	0.0044
{1, 3}	0.3674	0.3775	0.0101	0.6741	0.6837	0.0096
{1, 4}	0.3623	0.3722	0.0099	0.6691	0.6787	0.0096
{2, 3}	0.3213	0.3309	0.0096	0.6278	0.6377	0.0099
{2, 4}	0.3163	0.3259	0.0096	0.6225	0.6326	0.0101
{3, 4}	0.0244	0.0288	0.0044	0.1068	0.1183	0.0115
{1, 2, 3}	0.9397	0.9463	0.0066	0.9878	0.9900	0.0022
{1, 2, 4}	0.9306	0.9370	0.0064	0.9850	0.9871	0.0021
{1, 3, 4}	0.3949	0.4042	0.0093	0.6953	0.7047	0.0094
{2, 3, 4}	0.3463	0.3558	0.0095	0.6507	0.6609	0.0102

**Table 3.21:** Changed SAH data:  $\min_{j \in S} X_{n^{j+1}}^j < (\leq) \min_{l \in S^c} X_{n^{l+1}}^l$

We briefly consider what happens if, with the same data, some of the groups or some of the categories are combined. In Table 3.22 the original SAH data are presented following combination of groups 3 and 4 (now just called group 3) and also combination of categories 2 to 5 into a single category ‘not death’ (now just called category 2). The corresponding NPI lower and upper probabilities for the event that the next observation for each group in  $S$  is less than (or equal to) the next observation for all groups in  $S^c$  are presented in Table 3.23.

The main difference with the original results in Table 3.17 is that the differences

Group $j$	$C_1$	$C_2$	Total
1	59	151	210
2	48	142	190
3	87	315	402

**Table 3.22:** Combined SAH data

Subset ( $S$ )	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^{\leq}$	$\overline{P}_S^{\leq}$	$\Delta_S^{\leq}$
{1}	0.1625	0.1669	0.0044	0.6982	0.7045	0.0063
{2}	0.1406	0.1449	0.0043	0.6701	0.6765	0.0064
{3}	0.1149	0.1178	0.0029	0.6331	0.6399	0.0068
{1, 2}	0.0549	0.0572	0.0023	0.7970	0.7998	0.0028
{1, 3}	0.0449	0.0465	0.0016	0.7589	0.7643	0.0054
{2, 3}	0.0388	0.0404	0.0016	0.7311	0.7360	0.0049

**Table 3.23:** Combined SAH data:  $\max_{j \in S} X_{n^{j+1}}^j < (\leq) \min_{l \in S^c} X_{n^{l+1}}^l$ 

between the  $\underline{P}_S^<$  and  $\overline{P}_S^<$  and the corresponding  $\underline{P}_S^{\leq}$  and  $\overline{P}_S^{\leq}$  have become substantially larger. This is due to the grouping of categories 2 to 5, which of course leads to far fewer opportunities for future observations of different groups to be different, as now all observations in the new category 2 cannot be distinguished. Of course, group 1 is still worst in the sense of giving the largest lower and upper probabilities of the next observation being ‘death’, and the combined group 3 is best in this respect, reflecting that the original groups 3 and 4 were both better than groups 1 and 2. Most important, however, is to be aware that the NPI lower and upper probabilities for specific events depend on the representation of all the groups and categories.

Finally, Table 3.24 presents the NPI lower and upper probabilities for the comparison of groups 1 to 3 after group 4 is removed, using the data from Table 3.16. So now no information of group 4 is taken into account and it does not appear in either  $S$  or  $S^c$ . It is clear by comparing Table 3.24 to Table 3.17 that deleting group 4 leads to increased NPI lower and upper probabilities for the events represented, which is logical as group 4 no longer ‘competes’. The imprecision also seems to have increased a bit, but this is most likely due to the lower and upper probabilities

moving a bit closer to 0.5. This may actually mask an effect in the other direction, as removing a group will probably have the effect of reducing imprecision a bit [32].

Subset ( $S$ )	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_{S^c}^<$	$\overline{P}_{S^c}^<$	$\Delta_{S^c}^<$
{1}	0.2621	0.2692	0.0071	0.4931	0.5014	0.0083
{2}	0.2334	0.2402	0.0068	0.4578	0.4664	0.0086
{3}	0.1850	0.1911	0.0061	0.3963	0.4045	0.0082
{1, 2}	0.2595	0.2673	0.0078	0.4860	0.4954	0.0094
{1, 3}	0.2258	0.2333	0.0075	0.4365	0.4457	0.0092
{2, 3}	0.2082	0.2152	0.0068	0.4094	0.4181	0.0087

**Table 3.24:** SAH data without group 4:  $\max_{j \in S} X_{n^j+1}^j < (\leq) \min_{l \in S^c} X_{n^l+1}^l$

△

**Example 3.8.** In this example we illustrate our methods when we have a small number of observations in each group. The data are presented in Table 3.25.

	Categories					Total
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	
$G_1$	9	6	4	1	1	21
$G_2$	8	4	4	2	1	19
$G_3$	1	1	5	6	7	20
$G_4$	1	1	5	8	6	21

**Table 3.25:** Data: Small number of observation

The NPI lower and upper probabilities for the event that the next observation of each group in  $S$  is less than (or equal to) the next observation of all groups in the complementary set, as presented Section 3.3, are given in Table 3.26. The conclusions from Table 3.26 are in line with those discussed for Example 3.7 and as given in Table 3.17, but Table 3.26 shows more imprecision than Table 3.17, due to the substantially smaller numbers of observations in Example 3.8 than in Example 3.7.

Subset ( $S$ )	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$
{1}	0.2683	0.3561	0.0878	0.5837	0.6685	0.0848
{2}	0.2287	0.3086	0.0799	0.5349	0.6171	0.0822
{3}	0.0183	0.0412	0.0229	0.0796	0.1394	0.0598
{4}	0.0184	0.0415	0.0231	0.0800	0.1403	0.0603
{1, 2}	0.4434	0.5630	0.1196	0.6440	0.7700	0.1260
{1, 3}	0.0469	0.0864	0.0395	0.1394	0.2134	0.0740
{1, 4}	0.0482	0.0895	0.0413	0.1436	0.2193	0.0757
{2, 3}	0.0331	0.0672	0.0341	0.1070	0.1733	0.0663
{2, 4}	0.0343	0.0700	0.0357	0.1097	0.1773	0.0676
{3, 4}	0.0032	0.0107	0.0075	0.0186	0.0409	0.0223
{1, 2, 3}	0.2307	0.3144	0.0837	0.5144	0.6106	0.0962
{1, 2, 4}	0.2562	0.3447	0.0885	0.5480	0.6411	0.0931
{1, 3, 4}	0.0298	0.0628	0.0330	0.1084	0.1756	0.0672
{2, 3, 4}	0.0226	0.0511	0.0285	0.0797	0.1376	0.0579

**Table 3.26:** Small data:  $\max_{j \in S} X_{n^j+1}^j < (\leq) \min_{l \in S^c} X_{n^l+1}^l$  (Example 3.8)

## 3.5 Subset selection

In this section we discuss selecting a subset according to any criterion in terms of the NPI lower or upper probabilities presented in previous sections. For example, we may be interested in selecting a subset of minimal size for which  $\underline{P}_{(S)} > 0.5$ . In addition, if multiple subsets with the same size satisfy this criterion, we may choose the subset with the largest lower probability. Selection problems are common in statistics and several methods have been presented [8, 41, 64]. For example, they are used in screening experiments where, to end up with a small number of preferred treatments, one starts with all those available, then after a number of observations one wishes to continue with only a subset of all treatments, which should be likely to contain (all) the best treatment(s). Such subset selection will be discussed in detail in two examples.

**Example 3.9.** In this example we use Table 3.19 to select a subset which has NPI lower probability for the event that the next observation of each group in  $S$  is less than (or equal to) the next observation of all groups in the complementary set greater

than 0.5. From Table 3.19 we can see that groups 1 and 2 have the largest lower probabilities and also only the subset  $\{1, 2\}$  has lower probability greater than 0.5 for the event  $\max_{j \in S} X_{n^{j+1}}^j < (\leq) \min_{l \in S^c} X_{n^{l+1}}^l$ . In addition, any subset containing groups 1 and 2 also has the largest lower probability for the event  $\max_{j \in S} X_{n^{j+1}}^j \leq \min_{l \in S^c} X_{n^{l+1}}^l$ ,  $\underline{P}_{\{1,2,3\}}^{\leq} = 0.6017$  and  $\underline{P}_{\{1,2,4\}}^{\leq} = 0.6090$ . Using Table 3.19, if we want to choose the subset of size 2 which contains all the worst groups, then we will choose  $S = \{1, 2\}$ .

Furthermore, table 3.21 shows that again the subset  $\{1, 2\}$  has the largest lower probability for the event  $\min_{j \in S} X_{n^{j+1}}^j < (\leq) \min_{l \in S^c} X_{n^{l+1}}^l$ ,  $\underline{P}_{\{1,2\}}^{\leq} = 0.8817$  and  $\underline{P}_{\{1,2\}}^{\leq} = 0.9712$ , so it is the best choice for the subset of size two, which contains the worst group, considering the event that the next observation for at least one of the selected groups is less than the next observation for each group in the complementary set  $S^c$ . For subsets of size 3, again the subsets  $\{1, 2, 3\}$  and  $\{1, 2, 4\}$  have the largest lower probabilities and we choose the subset  $\{1, 2, 3\}$  as the best subset containing the worst group, where  $\underline{P}_{\{1,2,3\}}^{\leq} = 0.9397$  and  $\underline{P}_{\{1,2,3\}}^{\leq} = 0.9878$ .

△

**Example 3.10.** The data in Table 3.27 were presented by Agresti [1], and describe the relationship between children's mental health status and parents' socio-economic status (SES) for 1660 residents of Manhattan. Various models were fitted for these data, for example a row and column effects model which indicated positive association [1]. The children's mental health status is presented in 4 categories; well, mild symptoms, moderate symptoms and impaired.

We consider all possible subsets for this example to illustrate our method, in practice one may wish to restrict attention to some of these subsets. Tables 3.28 and 3.29 give the NPI lower and upper probabilities for the event that the next observation for at least one of the selected groups is less than (or equal to) the next observation from each of the non-selected groups, as presented in Section 3.4, for all possible subsets of sizes 2, 3, 4 and 5. The results from these tables can be used to select the best subset. For example, if we want to select the subset with size 2 that has NPI lower probability greater than 0.6 for the event that the next observation for at least one of the selected groups is less than or equal to the next



SES	Mental Health				Total
	Well	Mild	Moderate	Impaired	
	$C_1$	$C_2$	$C_3$	$C_4$	
A	64	94	58	46	262
B	57	94	54	40	245
C	57	105	65	60	287
D	72	141	77	94	384
E	36	97	54	78	265
F	21	71	54	71	217

**Table 3.27:** Mental health and parents' socio-economic status

observation from each of the non-selected groups, then subset  $\{1, 2\}$  will be the best choice. Clearly, the lower and upper probabilities increase with extending subsets. Finally, the particularly large number of observations in this example is reflected by small imprecision.

△

Subset ( $S$ )	$P_{S^1}^<$	$\bar{P}_{S^1}^<$	$\Delta_{S^1}^<$	$P_{S^1}^<$	$\bar{P}_{S^1}^<$	$\Delta_{S^1}^<$
{1}	0.10081	0.1047	0.0039	0.3849	0.3920	0.0071
{2}	0.0959	0.0997	0.0038	0.3807	0.3882	0.0075
{3}	0.0785	0.0818	0.0033	0.3339	0.3407	0.0068
{4}	0.0735	0.0763	0.0028	0.3211	0.3271	0.0060
{5}	0.0515	0.0541	0.0026	0.2610	0.2677	0.0067
{6}	0.0357	0.0381	0.0024	0.2045	0.2116	0.0071
{1, 2}	0.2338	0.2410	0.0072	0.6396	0.6479	0.0083
{1, 3}	0.2098	0.2166	0.0068	0.6070	0.6153	0.0083
{1, 4}	0.2029	0.2094	0.0065	0.5977	0.6057	0.0080
{1, 5}	0.1733	0.1794	0.0061	0.5542	0.5629	0.0087
{1, 6}	0.1516	0.1574	0.0058	0.5155	0.5247	0.0092
{2, 3}	0.2041	0.2108	0.0067	0.6018	0.6103	0.0085
{2, 4}	0.1972	0.2037	0.0065	0.5923	0.6006	0.0083
{2, 5}	0.1681	0.1741	0.0060	0.5485	0.5574	0.0089
{2, 6}	0.1465	0.1522	0.0057	0.5098	0.5192	0.0094
{3, 4}	0.1750	0.1809	0.0059	0.5576	0.5657	0.0081
{3, 5}	0.1471	0.1526	0.0055	0.5116	0.5204	0.0088
{3, 6}	0.1265	0.1318	0.0053	0.4709	0.4801	0.0092
{4, 5}	0.1410	0.1463	0.0053	0.5009	0.5094	0.0085
{4, 6}	0.1208	0.1257	0.0049	0.4597	0.4685	0.0088
{5, 6}	0.0960	0.1007	0.0047	0.4077	0.4174	0.0097
{1, 2, 3}	0.3866	0.3956	0.0090	0.7918	0.7986	0.0068
{1, 2, 4}	0.3771	0.3862	0.0091	0.7847	0.7916	0.0069
{1, 2, 5}	0.3376	0.3460	0.0084	0.7527	0.7600	0.0073
{1, 2, 6}	0.3074	0.3155	0.0081	0.7254	0.7332	0.0078
{1, 3, 4}	0.3459	0.3548	0.0089	0.7599	0.7672	0.0073
{1, 3, 5}	0.3082	0.3164	0.0082	0.7264	.7341	0.0077
{1, 3, 6}	0.2795	0.2873	0.0078	0.6977	0.7058	0.0081
{1, 4, 5}	0.2998	0.3080	0.0082	0.7184	0.7262	0.0078
{1, 4, 6}	0.2715	0.2793	0.0024	0.6895	0.6976	0.0081
{1, 5, 6}	0.2380	0.2454	0.0074	0.6515	0.6602	0.0087
{2, 3, 4}	0.3398	0.3485	0.0087	0.7546	0.7620	0.0074
{2, 3, 5}	0.3024	0.3105	0.0081	0.7207	0.7285	0.0078

Table 3.28: Subset containing the best group: NPI lower and upper probabilities

Subset ( $S$ )	$\underline{P}_{S^1}^<$	$\overline{P}_{S^1}^<$	$\Delta_{S^1}^<$	$\underline{P}_{S^1}^{\leq}$	$\overline{P}_{S^1}^{\leq}$	$\Delta_{S^1}^{\leq}$
{2, 3, 6}	0.2738	0.2816	0.0078	0.6920	0.7002	0.0082
{2, 4, 5}	0.2942	0.3023	0.0082	0.7127	0.7205	0.0078
{2, 4, 6}	0.2659	0.2736	0.0077	0.6836	0.6918	0.0082
{2, 5, 6}	0.2328	0.2401	0.0073	0.6452	0.654	0.0089
{3, 4, 5}	0.2668	0.2746	0.0078	0.6845	0.6926	0.0081
{3, 4, 6}	0.2424	0.2473	0.0049	0.6540	0.6624	0.0084
{3, 5, 6}	0.2084	0.2153	0.0069	0.6138	0.6229	0.0091
{4, 5, 6}	0.2014	0.2082	0.0068	0.6043	0.6134	0.0091
{1, 2, 3, 4}	0.5826	0.5923	0.0097	0.8993	0.9040	0.0047
{1, 2, 3, 5}	0.5315	0.5403	0.0088	0.8743	0.8792	0.0049
{1, 2, 3, 6}	0.4906	0.4991	0.0085	0.8537	0.8590	0.0053
{1, 2, 4, 5}	0.5199	0.5291	0.0092	0.8682	0.8735	0.0053
{1, 2, 4, 6}	0.4796	0.4884	0.0088	0.8474	0.8529	0.0055
{1, 2, 5, 6}	0.4343	0.4424	0.0081	0.8191	0.8250	0.0059
{1, 3, 4, 5}	0.4808	0.4902	0.0094	0.8478	0.8535	0.0057
{1, 3, 4, 6}	0.4426	0.4515	0.0089	0.8259	0.8319	0.0060
{1, 3, 5, 6}	0.3994	0.4077	0.0083	0.7963	0.8028	0.0065
{1, 4, 5, 6}	0.3897	0.3982	0.0085	0.7892	0.7959	0.0067
{2, 3, 4, 5}	0.4753	0.4844	0.0091	0.8426	0.8484	0.0058
{2, 3, 4, 6}	0.4371	0.4458	0.0087	0.8206	0.8267	0.0061
{2, 3, 5, 6}	0.3943	0.4023	0.0080	0.7905	0.7971	0.0066
{2, 4, 5, 6}	0.3847	0.3930	0.0083	0.7834	0.7902	0.0068
{3, 4, 5, 6}	0.3521	0.3604	0.0083	0.7590	0.7662	0.0072
{1, 2, 3, 4, 5}	0.7884	0.7955	0.0071	0.9619	0.9643	0.0024
{1, 2, 3, 4, 6}	0.7323	0.7390	0.0067	0.9459	0.9485	0.0026
{1, 2, 3, 5, 6}	0.6729	0.6789	0.0060	0.9237	0.9265	0.0028
{1, 2, 4, 5, 6}	0.6593	0.6661	0.0068	0.9182	0.9215	0.0033
{1, 3, 4, 5, 6}	0.6118	0.6193	0.0075	0.9003	0.9041	0.0038
{2, 3, 4, 5, 6}	0.6080	0.6151	0.0071	0.8953	0.8992	0.0039

**Table 3.29:** Subset containing the best group: NPI lower and upper probabilities

### 3.6 Maximum next observation

The multiple comparisons have so far been presented in terms of minimum value(s) to be in the set  $S$ , but the same approach can be used if interest is in maximum value(s). One can turn around the order of the  $K$  categories and return to a formulation in terms of minimum value(s), or for some events one can of course just exchange the roles of  $S$  and  $S^c$  to return to an event of interest in terms of minimum value(s). One can also derive the NPI lower and upper probabilities directly, along the same lines as done for the minimum values above. The NPI lower and upper probabilities for the event that the next observation from each group in  $S$  is greater than the next observation from each group in  $S^c$  are denoted by

$$\underline{P}_{S^*}^> = \underline{P} \left( \min_{j \in S} X_{n^{j+1}}^j > \max_{l \in S^c} X_{n^{l+1}}^l \right)$$

and

$$\overline{P}_{S^*}^> = \overline{P} \left( \min_{j \in S} X_{n^{j+1}}^j > \max_{l \in S^c} X_{n^{l+1}}^l \right)$$

These are given in Equations (3.49) and (3.50), respectively and derived in detail in Appendix A.4.

$$\underline{P}_{S^*}^> = \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^R (X_{n^{l+1}}^l \in \mathcal{C}_{1, M_w - 1}) \quad (3.49)$$

$$\overline{P}_{S^*}^> = \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \prod_{j \in S} \frac{R_{k_j}^j}{n^j + 1} \prod_{l \in S^c} P^L (X_{n^{l+1}}^l \in \mathcal{C}_{1, M_w - 1}) \quad (3.50)$$

where  $M_w = \min \{k_{j_1}, \dots, k_{j_w}\}$ .

The NPI lower and upper probabilities for the corresponding event with ‘greater than or equal to’ are derived by replacing  $M_w - 1$  by  $M_w$  in Equations (3.49) and (3.50). The NPI lower and upper probabilities for the event that the next observation of at least one group in  $S$  is greater than (or equal to) the next observation of all groups in  $S^c$ ,  $\max_{j \in S} X_{n^{j+1}}^j > (\geq) \max_{l \in S^c} X_{n^{l+1}}^l$ , follow from Equations (3.49) and (3.50) by replacing  $M_i$  by  $M_x$  in those equations.

**Example 3.11.** Consider again the data set described in Example 3.7 (Changed SAH data) and given in Table 3.18. The NPI lower and upper probabilities for the event that the next observation from each group in  $S$  is greater than (or equal to) the next observation from each group in  $S^c$  are given in Table 3.30.

Subset ( $S$ )	$\underline{P}_{S^*}^>$	$\overline{P}_{S^*}^>$	$\Delta_{S^*}^>$	$\underline{P}_{S^*}^{\geq}$	$\overline{P}_{S^*}^{\geq}$	$\Delta_{S^*}^{\geq}$
{1}	0.0272	0.0299	0.0027	0.0894	0.0952	0.0058
{2}	0.0293	0.0324	0.0031	0.1082	0.1150	0.0068
{3}	0.2856	0.2952	0.0096	0.6090	0.6195	0.0105
{4}	0.2786	0.2880	0.0094	0.6017	0.6126	0.0109
{1, 2}	0.0025	0.0029	0.0040	0.0172	0.0188	0.0016
{1, 3}	0.0305	0.0340	0.0035	0.1142	0.1215	0.0073
{1, 4}	0.0326	0.0361	0.0035	0.1200	0.1271	0.0071
{2, 3}	0.0396	0.0436	0.0040	0.1430	0.1511	0.0081
{2, 4}	0.0427	0.0467	0.0040	0.1488	0.1568	0.0080
{3, 4}	0.5755	0.5897	0.0142	0.7953	0.8100	0.0147
{1, 2, 3}	0.0100	0.01217	0.0022	0.0537	0.0603	0.0066
{1, 2, 4}	0.0129	0.01504	0.0021	0.0630	0.0694	0.0064
{1, 3, 4}	0.2953	0.3047	0.0094	0.5958	0.6051	0.0093
{2, 3, 4}	0.3391	0.3493	0.0102	0.6442	0.6537	0.0095

**Table 3.30:** NPI lower and upper probabilities:  $\min_{j \in S} X_{n^j+1}^j > (\geq) \max_{l \in S^c} X_{n^l+1}^l$

It can be seen from Table 3.30 that groups 3 and 4 give the largest lower and upper probabilities for the event that the next observation of each group in subset  $S$  is greater than the next observation from each group in the complementary set  $S^c$  ( $\min_{j \in S} X_{n^j+1}^j > \max_{l \in S^c} X_{n^l+1}^l$ ), so any subset containing these groups will be likely to give the maximum next observation. Clearly, subset  $\{3, 4\}$  gives the largest lower and upper probabilities as it has more combinations for the event that the next observation of each group in subset  $S$  is greater than the next observation from each group in the complementary set  $S^c$ .

△

## 3.7 Concluding remarks

In this chapter NPI has been presented for multiple comparisons for ordinal categorical groups. Although the data look like contingency tables, we do not test for the independence but we consider inferences for future observations based on the available data, which is an attractive alternative approach to the classical methods which typically test hypotheses. Our results hold for a single future observation for each category. If interest is in more future observations, the NPI approach is possible too but it has not yet been developed; this is an interesting topic for future research. It is also possible to combine NPI lower and upper probabilities with utilities in a decision theoretic framework, which will often be relevant in applications with ordered categorical data. A further topic of interest is the possibility to use other latent variable representations to model specific relations between different categories. For example, there may be applications where a 2-dimensional latent variable representation is suitable for the categories. This would require the development of NPI for 2-dimensional random quantities, which is also an interesting research challenge.

# Chapter 4

## Accuracy of ordinal diagnostic tests

### 4.1 Introduction

In this chapter we introduce nonparametric predictive inference (NPI) for accuracy of diagnostic tests with ordinal outcomes, with the inferences based on data for a disease group and a non-disease group. The evaluation of the accuracy of a diagnostic test is crucial in many application areas, in particular for medical diagnoses. Traditional statistical methods tend to use concepts like ‘sensitivity’ and ‘specificity’ to express such accuracy, and these are combined into the Receiver Operating Characteristic (ROC) curve, which has become a popular methodological tool which provides insight into the quality of the test. Sensitivity, specificity and the ROC curve are properties of assumed underlying populations and their probability distributions, and are estimated from available data, in line with the traditional frequentist approach to statistics. The data consist of observations of test outcomes for two groups, one consisting of individuals known to have the disease or other condition of interest, the other consisting of non-diseased individuals. The predictive nature of NPI can be attractive for diagnostic tests as one may wish to consider explicitly the quality of the test for one or more future individuals. Recently, Coolen-Maturi et al [27] introduced NPI for accuracy of diagnostic tests which have real-valued outcomes, and also for tests with binary outcomes [28].

In many medical applications, as well as in other areas, diagnostic tests with ordinal outcomes appear, so the test gives a result in one of several ordered categories [1, 7]. In this chapter, NPI for such ordinal diagnostic tests is presented, using the NPI approach for ordinal data which was presented in Section 2.4 [25]. Section 4.2 provides a brief introduction to ordinal diagnostic tests. In Section 4.3 we explain how the usual empirical estimator for the ROC curve for ordinal diagnostic tests can be generalized to empirical lower and upper ROC curves. NPI lower and upper ROC curves for ordinal diagnostic tests are introduced in Section 4.4, where also the use of the Youden index for determining an optimal cut-off point for the diagnostic test is discussed. In Section 4.5 the areas under the NPI lower and upper ROC curves are presented, and it is shown that these are equal to the NPI lower and upper probabilities for correctly ordered future observations from the non-disease and disease groups. Section 4.6 presents two examples to illustrate and discuss the new approach presented in this chapter, and Section 4.7 presents some concluding remarks.

## 4.2 Ordinal diagnostic tests

We consider a diagnostic test with ordinal test results, meaning that 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 disease (or other condition) of interest. We assume throughout this chapter that error-free data are available on individuals in two groups according to known disease status, with presence of the disease ('disease group') indicated by  $D = 1$  and absence of the disease ('non-disease group') by  $D = 0$ . The notation for the numbers of individuals for each combination of disease status and test result is introduced in Table 4.1. Throughout this chapter the disease (non-disease) group is indicated by a superscript 1 (0).

We will assume that there is a  $k \in \{1, \dots, K\}$  such that a test result in categories  $\{C_k, \dots, C_K\}$  is interpreted as indicating presence of the disease ('positive test result') and a test result in categories  $\{C_1, \dots, C_{k-1}\}$  as indicating absence of the disease



	Diagnostic test result						
Condition status	$C_1$	$\dots$	$C_{k-1}$	$C_k$	$\dots$	$C_K$	Total
$D = 0$	$n_1^0$	$\dots$	$n_{k-1}^0$	$n_k^0$	$\dots$	$n_K^0$	$n^0$
$D = 1$	$n_1^1$	$\dots$	$n_{k-1}^1$	$n_k^1$	$\dots$	$n_K^1$	$n^1$
Total	$n_1$	$\dots$	$n_{k-1}$	$n_k$	$\dots$	$n_K$	$n$

**Table 4.1:** Ordinal test data

(‘negative test result’) [69, 73]. Study of an appropriate choice of the value  $k$ , called the ‘cut-off (point)’ or the ‘threshold’, is a main goal for statistical inference in this scenario. The sensitivity (*sens*) of a diagnostic test is the probability that the test result is positive given that the disease is present, it is also called the true positive fraction (*TPF*) (it is also known in the literature as the true positive rate). The specificity (*spec*) of a diagnostic test is the probability that the test result is negative given absence of the disease. The false positive fraction (*FPF*) is the probability of a positive test result for an individual without the disease, so  $spec = 1 - FPF$ . Let random quantity  $X^1$  denote the test result for an individual of the disease group and  $X^0$  the test result for an individual of the non-disease group, then with given cut-off point  $k$  we have  $TPF(k) = P(X^1 \in \{C_k, \dots, C_K\})$  and  $FPF(k) = P(X^0 \in \{C_k, \dots, C_K\})$ .

The unbiased empirical estimators of  $TPF(k)$  and  $FPF(k)$  for  $k \in \{1, \dots, K\}$  are [73]

$$TPF_e(k) = \frac{1}{n^1} \sum_{i=k}^K n_i^1 \quad (4.1)$$

$$FPF_e(k) = \frac{1}{n^0} \sum_{i=k}^K n_i^0 \quad (4.2)$$

### 4.2.1 ROC curve

The receiver operating characteristic (ROC) curve is a popular tool for describing and comparing the accuracy of diagnostic tests. The ROC curve in case of ordinal diagnostic tests applied to a disease group and a non-disease group, is defined as follows.

**Definition 4.1.** *The ROC curve in case of ordinal diagnostic tests is defined by the points  $(FPF(k), TPF(k))$  for  $k \in \{1, \dots, K+1\}$ , where  $(FPF(1), TPF(1)) = (1, 1)$  and  $(FPF(K+1), TPF(K+1))$  is defined as  $(0, 0)$ .*

It should be emphasized that both  $FPF(k)$  and  $TPF(k)$  are decreasing functions of  $k$ . In between these  $K+1$  points the ROC curve can be defined by straight line segments linking all pairs of neighbouring points, which is attractive with regard to properties of the area under the ROC curve as explained below.

The empirical ROC curve [55], based on estimators (4.1) and (4.2), is defined at the points corresponding to  $k = 1, \dots, K+1$ , with the end-points as above, as

$$ROC_e(k) = (FPF_e(k), TPF_e(k)) \quad (4.3)$$

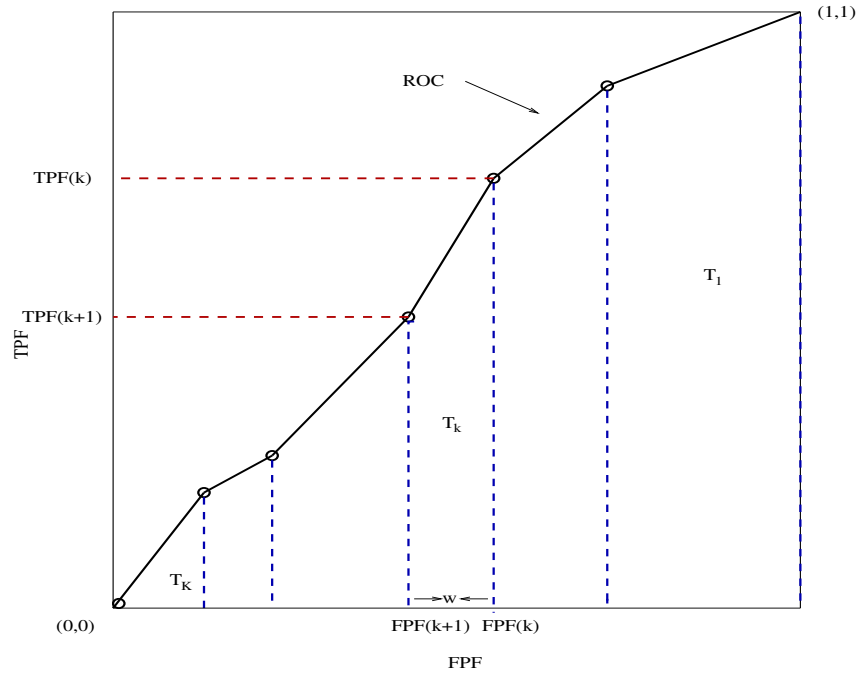
and the full empirical ROC curve is again usually defined by linking these points  $ROC_e(k)$  for neighbouring values of  $k$  by straight line segments.

### 4.2.2 Area under the ROC curve (AUC)

The area under the ROC curve, the AUC, reflects the intrinsic diagnostic capability of a test and it can be estimated nonparametrically or parametrically [73]. An ideal test and an uninformative test have AUC equal to 1 and 0.5, respectively. The AUC is equal to the probability that a randomly chosen individual from the non-disease group has a test result that is lower than that of a randomly selected individual from the disease group [52, 73]. For discrete-valued test results, as is the case with ordinal data, the AUC is equal to [55]

$$AUC = P(X^0 < X^1) + \frac{1}{2}P(X^0 = X^1) \quad (4.4)$$

where  $X^0 < (=)X^1$  denotes that  $X^0$  belongs to a category which is smaller than (equal to) the category to which  $X^1$  belongs, in the ordering of the categories  $C_1 < \dots < C_K$ . The second term on the right-hand side of Equation (4.4) corresponds to the definition of the ROC curve as a straight line segment in between the points corresponding to neighbouring values of  $k$  as explained in the following proof.



**Figure 4.1:** Area under the *ROC* curve

The proof of Equation (4.4) is as follows, this proof is an extended version of the proof provided in [55], which we present because of its connection to the area under the NPI lower and upper ROC curve which will be discussed in Section 4.5. The AUC can be calculated by summing the area of trapezoids which form the ROC curve as shown in Figure 4.1. The area of trapezoid  $T_k$  is equal to  $w \left[ \frac{1}{2} (a + b) \right]$  where  $w$  is the width of  $T_k$  and is equal to

$$w = P(X^0 \in \{C_k, \dots, C_K\}) - P(X^0 \in \{C_{k+1}, \dots, C_K\}) = P(X^0 \in C_k) = \frac{n_k^0}{n^0}$$

where  $P(X^0 \in \{C_k, \dots, C_K\}) = FPF(k)$  and  $P(X^0 \in \{C_{k+1}, \dots, C_K\}) = FPF(k+1)$ . The values  $a$  and  $b$  are the lengths of the parallel sides  $P(X^1 \in \{C_k, \dots, C_K\})$  and  $P(X^1 \in \{C_{k+1}, \dots, C_K\})$ , respectively, where  $P(X^1 \in \{C_k, \dots, C_K\}) = TPF(k)$  and  $P(X^1 \in \{C_{k+1}, \dots, C_K\}) = TPF(k+1)$ . The area of  $T_k$  is equal to

$$\begin{aligned} & P(X^0 \in C_k) \left[ \frac{1}{2} P(X^1 \in \{C_k, \dots, C_K\}) + \frac{1}{2} P(X^1 \in \{C_{k+1}, \dots, C_K\}) \right] \\ &= P(X^0 \in C_k) \left[ \frac{1}{2} P(X^1 \in C_k) + P(X^1 \in \{C_{k+1}, \dots, C_K\}) \right] \end{aligned}$$

By summing the areas of all  $K$  trapezoids we get the area under the ROC curve,

$$AUC = \sum_{k=1}^K P(X^0 \in C_k) P(X^1 \in \{C_{k+1}, \dots, C_K\}) + \frac{1}{2} \sum_{k=1}^K P(X^0 \in C_k) P(X^1 \in C_k)$$

$$AUC = P(X^0 < X^1) + 1/2P(X^0 = X^1)$$

which is Equation (4.4).

In terms of the  $FPF(k)$  and  $TPF(k)$ , the AUC can be expressed as

$$AUC = \sum_{k=1}^K \frac{1}{2} (TPF(k) + TPF(k+1)) \times (FPF(k) - FPF(k+1)) \quad (4.5)$$

An unbiased empirical estimator of the AUC for the ordinal setting discussed in this chapter, which is related to the Mann-Whitney statistic for comparison of two distributions, is [73]

$$AUC_e = \frac{1}{n^0 n^1} \sum_{i=1}^K n_i^0 \left( \frac{n_i^1}{2} + \sum_{j=i+1}^K n_j^1 \right) \quad (4.6)$$

### 4.2.3 Youden's index

To fully define the diagnostic method, a value of the cut-off point  $k$  must be determined such that the test provides good discrimination between the disease and non-disease groups. One popular approach is to maximise Youden's index [39, 72], which is defined by

$$J(k) = sens(k) + spec(k) - 1 = TPF(k) - FPF(k) \quad (4.7)$$

which is the maximum vertical distance from the curve to the diagonal line from  $(0, 0)$  to  $(1, 1)$  [58].

The optimal cut-off point, based on Youden's index, is the value of  $k$  which maximises  $J(k)$ , so which maximises the sum of sensitivity and specificity [40, 57]. An empirical estimator of  $J(k)$  is

$$J_e(k) = TPF_e(k) - FPF_e(k) \quad (4.8)$$

$J_e(k)$  is equal to one if the two groups are completely separated and zero if there is complete overlap between the groups.

Of course, there are other optimality criteria, for example to take into account the severity of consequences of wrong diagnoses, we refer to [73] for more details and leave further investigation of the use of other criteria within the NPI framework as a topic for future research.

### 4.3 Empirical lower and upper ROC curves

As explained in the previous section, the empirical ROC curve for ordinal diagnostic tests is logically defined at  $K + 1$  points, namely by the points  $ROC_e(k) = (FPF_e(k), TPF_e(k))$  for  $k = 1, \dots, K + 1$ . In between these points, however, although the tradition is to define the full ROC curve by connecting the neighbouring points by straight line segments, one could consider another approach that is more in line with the general ideas of lower and upper probabilities, namely by considering the set of all possible ROC curves that go through these  $K + 1$  points and defining the maximum lower bound and minimum upper bound of this set as the empirical lower and upper ROC curves, respectively. This is of interest in its own right, but particularly also for comparison with the NPI lower and upper ROC curves which we will introduce in Section 4.4. This is all illustrated in examples in Section 4.6.

When plotting the points  $ROC_e(k) = (FPF_e(k), TPF_e(k))$  for  $k = 1, \dots, K + 1$  in the unit square, any ROC curve that goes through these  $K + 1$  points must be a non-decreasing function, and all such non-decreasing functions can be ROC curves. Therefore, the maximum lower bound, which we call the empirical lower ROC curve for ordinal diagnostic tests and denote by  $\underline{ROC}_e$ , is the step-function with  $TPF$  value equal to  $TPF(k+1)$  for all  $FPF \in [FPF(k+1), FPF(k))$ , and the corresponding empirical upper ROC curve, denoted by  $\overline{ROC}_e$  is the step-function with  $TPF$  value equal to  $TPF(k)$  for all  $FPF \in (FPF(k+1), FPF(k)]$ . These empirical lower and upper ROC curves are equal to the empirical ROC curve at the points  $(FPF_e(k), TPF_e(k))$ .

It is also of interest to consider lower and upper empirical estimators for the AUC, denoted by  $\underline{AUC}_e$  and  $\overline{AUC}_e$ , which are defined to be the areas under the  $\underline{ROC}_e$  and  $\overline{ROC}_e$ , respectively. These are easily proven to be equal to

$$\underline{AUC}_e = \frac{1}{n^0 n^1} \sum_{i=1}^K n_i^0 \left( \sum_{j=i+1}^K n_j^1 \right) \quad (4.9)$$

$$\overline{AUC}_e = \frac{1}{n^0 n^1} \sum_{i=1}^K n_i^0 \left( \sum_{j=i}^K n_j^1 \right) \quad (4.10)$$

These are unbiased estimators for  $P(X^0 < X^1)$  and  $P(X^0 \leq X^1)$ , respectively, and  $AUC_e = (\underline{AUC}_e + \overline{AUC}_e)/2$ .

## 4.4 NPI lower and upper ROC curves

To introduce NPI for the ROC curve for ordinal tests, the same latent variable setting and data notation as in Figure 2.1 in Section 2.4 and Table 4.1 are used. We now consider the outcome of the ordinal diagnostic test applied to one further (‘future’) individual from the disease group, represented by the random quantity  $X_{n^1+1}^1$ , and to one further individual from the non-disease group, with random outcome  $X_{n^0+1}^0$ . For these random quantities we make the assumptions  $A_{(n^1)}$  and  $A_{(n^0)}$ , respectively, for the assumed underlying latent variable representation as discussed in Section 2.4, given the data for each group. We should emphasize here that we assume these two groups to be fully independent, which can be interpreted in the sense that no information with regard to one group contains information about the other group.

To define NPI lower and upper ROC curves, we first define NPI lower and upper true and false positive fractions.

**Definition 4.2.** For  $k \in \{2, \dots, K\}$ , the NPI lower true positive fraction  $\underline{TPF}(k)$  and the corresponding NPI upper true positive fraction  $\overline{TPF}(k)$  are defined by

$$\underline{TPF}(k) = \underline{P}(X_{n^1+1}^1 \in \{C_k, \dots, C_K\}) = \frac{1}{n^1 + 1} \sum_{j=k}^K n_j^1 \quad (4.11)$$

$$\overline{TPF}(k) = \overline{P}(X_{n^1+1}^1 \in \{C_k, \dots, C_K\}) = \frac{1}{n^1 + 1} \left[ \left( \sum_{j=k}^K n_j^1 \right) + 1 \right] \quad (4.12)$$

For  $k = 1$ , we define  $\underline{TPF}(1) = 1$  and  $\overline{TPF}(1) = 1$ .

In addition, for use in the definition of an end-point of the ROC curves, we define  $\underline{TPF}(K + 1) = 0$  and  $\overline{TPF}(K + 1) = 0$ .

**Definition 4.3.** The NPI lower false positive fraction  $\underline{FPF}(k)$  and the corresponding NPI upper false positive fraction  $\overline{FPF}(k)$  are defined, for  $k \in \{2, \dots, K\}$ , by

$$\underline{FPF}(k) = \underline{P}(X_{n^0+1}^0 \in \{C_k, \dots, C_K\}) = \frac{1}{n^0 + 1} \sum_{i=k}^K n_i^0 \quad (4.13)$$

$$\overline{FPF}(k) = \overline{P}(X_{n^0+1}^0 \in \{C_k, \dots, C_K\}) = \frac{1}{n^0 + 1} \left[ \left( \sum_{i=k}^K n_i^0 \right) + 1 \right] \quad (4.14)$$

And further we define  $\underline{FPF}(1) = 1$ ,  $\overline{FPF}(1) = 1$  and  $\underline{FPF}(K+1) = 0$ ,  $\overline{FPF}(K+1) = 0$ . The Equations (4.11) - (4.14) are derived by the first special case (2.6) and (2.7) presented in Section 2.4. It is straightforward to show that, for all  $k \in \{1, \dots, K+1\}$ ,

$$\underline{FPF}(k) \leq FPF_e(k) \leq \overline{FPF}(k) \quad (4.15)$$

and

$$\underline{TPF}(k) \leq TPF_e(k) \leq \overline{TPF}(k) \quad (4.16)$$

ROC curves in the ordinal setting represent the pairs  $(FPF(k), TPF(k))$  for  $k = 1, \dots, K+1$ , all non-decreasing functions that link these points can be ROC curves for such a diagnostic test with ordinal outcomes applied to a disease group and a non-disease group. Using the NPI lower and upper false and true positive fractions, we can consider the set of all ROC curves with  $\underline{FPF}(k) \leq FPF(k) \leq \overline{FPF}(k)$  and  $\underline{TPF}(k) \leq TPF(k) \leq \overline{TPF}(k)$  for  $k = 1, \dots, K+1$ , and links between these points as discussed. The NPI lower and upper ROC curves are defined as the maximum lower bound and the minimum upper bound, respectively, of this set of ROC curves. It is easy to verify that the NPI lower ROC curve  $\underline{ROC}$  goes through the points  $(\overline{FPF}(k), \underline{TPF}(k))$  and has  $TPF = \underline{TPF}(k+1)$  corresponding to all  $FPF \in [\overline{FPF}(k+1), \overline{FPF}(k)]$ . Similarly, the NPI upper ROC curve  $\overline{ROC}$  goes through the points  $(\underline{FPF}(k), \overline{TPF}(k))$  and has  $TPF = \overline{TPF}(k)$  corresponding to all  $FPF \in (\underline{FPF}(k+1), \underline{FPF}(k)]$ . These NPI lower and upper ROC curves are illustrated in Figures 4.2 and 4.3, which support the proofs of the main results in Section 4.5, and also in the examples in Section 4.6. It is also easy to verify that the NPI lower and upper ROC curves bound the empirical lower and upper ROC

curves, which in turn bound the empirical ROC curve. This can be denoted by

$$\underline{ROC} \leq \underline{ROC}_e \leq ROC_e \leq \overline{ROC}_e \leq \overline{ROC} \quad (4.17)$$

where, for example,  $\underline{ROC} \leq \underline{ROC}_e$  denotes that the NPI lower ROC curve  $\underline{ROC}$  lies entirely below or on the empirical lower ROC curve  $\underline{ROC}_e$  in the  $(FPF, TPF)$ -square. This will also be illustrated in the examples in Section 4.6.

Let us introduce Youden's index corresponding to the NPI lower ROC curve, for  $k = 1, \dots, K$ , as

$$\underline{J}(k) = \underline{TPF}(k) - \overline{FPF}(k) \quad (4.18)$$

and Youden's index corresponding to the NPI upper ROC curve, for  $k = 1, \dots, K$ , as

$$\overline{J}(k) = \overline{TPF}(k) - \underline{FPF}(k) \quad (4.19)$$

For  $k = 1$ , we have  $\underline{J}(1) = \overline{J}(1) = 0$ . It follows that, for  $k = 2, \dots, K$ ,

$$\overline{J}(k) = \underline{J}(k) + \frac{1}{n^0 + 1} + \frac{1}{n^1 + 1} \quad (4.20)$$

Relations (4.15) and (4.16) imply, for all  $k = 1, \dots, K$ ,

$$\underline{J}(k) \leq J_e(k) \leq \overline{J}(k) \quad (4.21)$$

Relation (4.20) implies that  $\overline{J}(k)$  and  $\underline{J}(k)$  are maximal for the same value of  $k$ . However, the inequalities (4.21) do not imply that  $J_e(k)$  is also maximal for the same value of  $k$ , although in many applications this is the case.

## 4.5 Area under the NPI lower or upper ROC curve

As discussed before, the area under an ROC curve has a nice interpretation due to its relation to the probability of correct ordering of one random observation from each of the non-disease group and the disease group. For the areas under the NPI lower and upper ROC curves, similar results are derived in this section, which allows a particularly attractive interpretation of these areas due to the predictive nature of NPI, with inferences directly comparing one further individual from each group.



Because NPI is exactly calibrated from frequentist statistics perspective [46], these areas under the NPI lower and upper ROC curves provide meaningful insight into the accuracy of the ordinal diagnostic test if it were applied to one individual from each of the non-disease group and the disease group.

We denote the area under the NPI lower ROC curve  $\underline{ROC}$  by  $\underline{AUC}$  and the area under the NPI upper ROC curve  $\overline{ROC}$  by  $\overline{AUC}$ . Of course, (4.17) implies that

$$\underline{AUC} \leq \underline{AUC}_e \leq AUC_e \leq \overline{AUC}_e \leq \overline{AUC} \quad (4.22)$$

Theorems 4.1 and 4.2 provide the important relations between  $\underline{AUC}$  and the NPI lower probability  $\underline{P}(X_{n^0+1}^0 < X_{n^1+1}^1)$  and between  $\overline{AUC}$  and the NPI upper probability  $\overline{P}(X_{n^0+1}^0 \leq X_{n^1+1}^1)$ . These show that the NPI approach to accuracy of ordinal diagnostic tests via the ROC curves is fully consistent with the established ROC theory. These NPI lower and upper probabilities were presented by [25], it is important to emphasize that these NPI lower and upper probabilities are for two different events, which correspond to the probabilities for which the empirical  $\underline{AUC}_e$  and  $\overline{AUC}_e$ , given by (4.9) and (4.10), are unbiased estimators, as discussed in Section 4.3.

**Theorem 4.1.** The area under the NPI lower ROC curve  $\underline{ROC}$ , denoted by  $\underline{AUC}$ , is equal to the NPI lower probability for the event that the test result  $X_{n^0+1}^0$  for a further individual from the non-disease group falls into a lower ranked category than the test result  $X_{n^1+1}^1$  for a further individual from the disease group,

$$\underline{AUC} = \frac{1}{(n^0 + 1)(n^1 + 1)} \sum_{k=1}^{K-1} \sum_{j=k+1}^K n_k^0 n_j^1 = \underline{P}(X_{n^0+1}^0 < X_{n^1+1}^1) \quad (4.23)$$

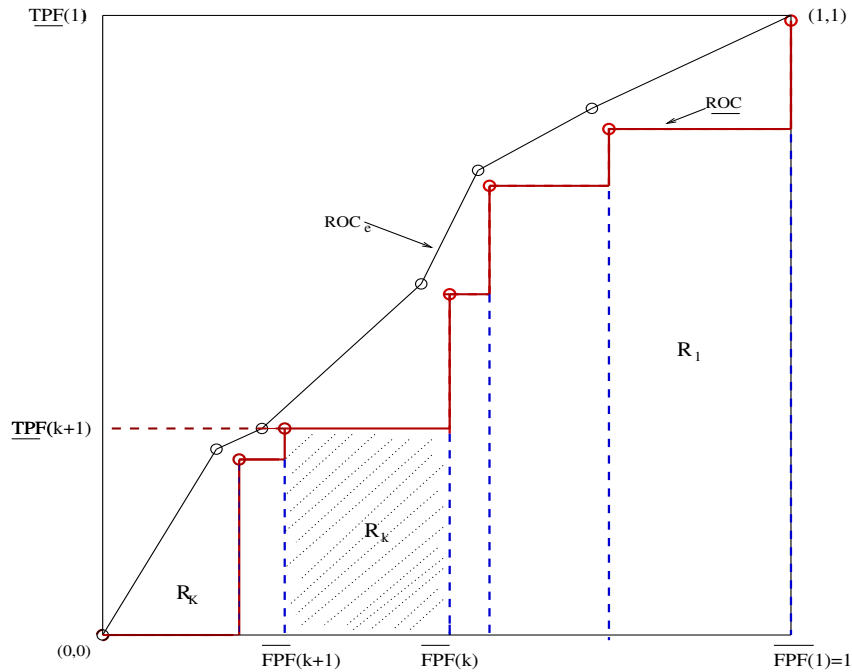
*Proof.* Figure 4.2 shows a typical empirical ROC curve  $ROC_e$  and the corresponding NPI lower ROC curve  $\underline{ROC}$ . The area under the  $\underline{ROC}$  consists of  $K$  rectangular blocks  $R_1$  to  $R_K$ , from right to left as illustrated, of which  $R_K$  actually has area zero. The area  $R_k$  corresponding to category  $k$ ,  $k = 1, \dots, K$ , adds a rectangular area of size  $\frac{n_k^0}{n^0+1} * \underline{TPF}(k+1)$  to the total area under  $\underline{ROC}$ , where  $\frac{n_k^0}{n^0+1}$  is equal to

$$\overline{FPF}(k) - \overline{FPF}(k+1) = \overline{P}(X_{n^0+1}^0 \in \{C_k, \dots, C_K\}) - \overline{P}(X_{n^0+1}^0 \in \{C_{k+1}, \dots, C_K\})$$

and  $\underline{TPF}(k+1) = \underline{P}(X_{n^1+1}^1 \in \{C_{k+1}, \dots, C_K\})$ . Summing these areas for blocks  $R_1, \dots, R_{K-1}$  gives  $\underline{AUC}$  as in Equation (4.23), by

$$\underline{AUC} = \sum_{k=1}^{K-1} \frac{n_k^0}{n^0 + 1} * \underline{TPF}(k+1) = \sum_{k=1}^{K-1} \frac{n_k^0}{n^0 + 1} \sum_{j=k+1}^K \frac{n_j^1}{n^1 + 1} \tag{4.24}$$

The second equality in Equation (4.23) is the first special case in Subsection 3.2.3.



**Figure 4.2:** The area under the NPI lower  $ROC$  curve

□

**Theorem 4.2.** The area under the NPI upper ROC curve  $\overline{ROC}$ , denoted by  $\overline{AUC}$ , is equal to the NPI upper probability for the event that the test result  $X_{n^0+1}^0$  for a further individual from the non-disease group falls into a lower or equally ranked category than the test result  $X_{n^1+1}^1$  for a further individual from the disease group,

$$\overline{AUC} = \frac{1}{(n^0 + 1)(n^1 + 1)} \left[ \sum_{k=1}^K \sum_{j=k}^K n_k^0 n_j^1 + n^0 + n^1 + 1 \right] = \overline{P}(X_{n^0+1}^0 \leq X_{n^1+1}^1) \tag{4.25}$$



$$\begin{aligned}
&= \frac{1}{(n^0 + 1)(n^1 + 1)} \left[ (n_1^0 + 1)(n^1 + 1) + \sum_{k=1}^K n_k^0 \left[ \left( \sum_{j=k}^K n_j^1 \right) + 1 \right] - n_1^0 \left[ \left( \sum_{j=1}^K n_j^1 \right) + 1 \right] \right] \\
&= \frac{1}{(n^0 + 1)(n^1 + 1)} \left[ n_1^0 n^1 + n_1^0 + n^1 + 1 + \sum_{k=1}^K \sum_{j=k}^K n_k^0 n_j^1 + \sum_{k=1}^K n_k^0 - n_1^0 (n^1 + 1) \right] \\
&= \frac{1}{(n^0 + 1)(n^1 + 1)} \left[ \sum_{k=1}^K \sum_{j=k}^K n_k^0 n_j^1 + n^0 + n^1 + 1 \right] \tag{4.26}
\end{aligned}$$

The second equality in Equation (4.25) is the first special case in Subsection 3.2.3.

□

It is important to emphasize that the areas under the NPI lower and upper ROC curves, as given in Theorems 4.1 and 4.2, are indeed equal to the NPI lower and upper probabilities given in these theorems. This is in contrast to the areas under the empirical lower and upper ROC curves, as given by (4.9) and (4.10), which are unbiased estimators of the corresponding probabilities. In the latter case, those probabilities are considered to be properties of assumed underlying populations, while in NPI the inferences are directly predictive, so in terms of the future observations, hence this conceptually important difference in the interpretation of these areas under the respective ROC curves.

A further interesting fact is that the conjugacy property of NPI lower and upper probabilities, that is  $\underline{P}(A) = 1 - \overline{P}(A^c)$  for any event  $A$  and its complementary event  $A^c$  [4], leads to

$$\underline{P}(X_{n^1+1}^1 < X_{n^0+1}^0) = 1 - \overline{P}(X_{n^0+1}^0 \leq X_{n^1+1}^1)$$

so the lower probability on the left-hand side is actually the area above the NPI upper ROC curve  $\overline{ROC}$  in the unit-square, and similarly

$$\overline{P}(X_{n^1+1}^1 \leq X_{n^0+1}^0) = 1 - \underline{P}(X_{n^0+1}^0 < X_{n^1+1}^1)$$

so the upper probability on the left-hand side is the area above the NPI lower ROC curve  $\underline{ROC}$  in the unit-square.

## 4.6 Examples

In this section, two examples are given to illustrate the NPI lower and upper ROC curves presented in this chapter, together with the empirical lower and upper ROC curves and the areas under these curves. The first example has small numbers of observations per group, the second has relatively large numbers of observations.

**Example 4.1.** Table 4.2 presents data on a diagnostic test with outcomes on an ordinal scale with five categories and involving the use of gallium citrate imaging to identify the focal source of sepsis [48, 73]. The empirical estimates of  $FPF(k)$  and

	Imaging results ( $Y$ )					
Disease status	1	2	3	4	5	Total
No disease ( $D = 0$ )	11	2	3	1	3	20
Disease ( $D = 1$ )	12	6	3	1	18	40

**Table 4.2:** Results of gallium citrate imaging

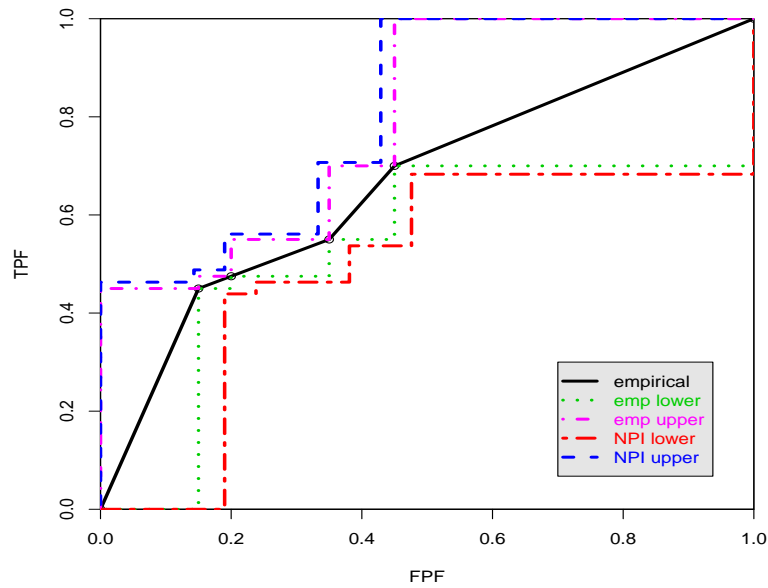
$TPF(k)$  for  $k = 1, \dots, 5$ , given by Equations (4.1) and (4.2), are presented in Table 4.3.

$k$	1	2	3	4	5
$FPF(k)$	1	0.4286	0.3333	0.1905	0.1429
$FPF_e(k)$	1	0.4500	0.3500	0.2000	0.1500
$\overline{FPF}(k)$	1	0.4762	0.3810	0.2381	0.1905
$\underline{TPF}(k)$	1	0.6829	0.5366	0.4634	0.4390
$TPF_e(k)$	1	0.7000	0.5500	0.4750	0.4500
$\overline{TPF}(k)$	1	0.7073	0.5610	0.4878	0.4634

**Table 4.3:** NPI lower, upper and empirical  $FPF$  and  $TPF$

The area under the empirical ROC curve, as given in Equation (4.6), is  $AUC_e = 0.6638$ . Using Equations (4.9) and (4.10), the areas under the empirical lower and upper ROC curves are  $\underline{AUC}_e = 0.5338$  and  $\overline{AUC}_e = 0.7938$ , respectively.

The NPI lower and upper  $FPF$  and  $TPF$ , as given in Equations (4.11) - (4.14), are also presented in Table 4.3. These clearly bound the corresponding empirical  $FPF$  and  $TPF$ . The NPI lower and upper ROC curves,  $\underline{ROC}$  and  $\overline{ROC}$  as introduced in Section 4.4, are presented in Figure 4.4, together with the empirical ROC curve and the empirical lower and upper ROC curves as introduced in Section 4.3. It is clear that these are nested as indicated by Relation (4.17).



**Figure 4.4:** Five (lower and upper) ROC curves

The area under the NPI lower ROC curve is  $\underline{AUC} = 0.4959$ , which is equal to the NPI lower probability  $\underline{P}(X_{21}^0 < X_{41}^1) = 0.4959$ . The area under the NPI upper ROC curve is  $\overline{AUC} = 0.8084$ , which is equal to the NPI upper probability  $\overline{P}(X_{21}^0 \leq X_{41}^1) = 0.8084$ . By the conjugacy property of NPI lower and upper probabilities, these values imply  $\underline{P}(X_{41}^1 < X_{21}^0) = 0.1916$  and  $\overline{P}(X_{41}^1 \leq X_{21}^0) = 0.5041$ . The areas under the five ROC curves in this example illustrate Relation (4.22).

The fact that the NPI lower AUC is just under 0.5 could be interpreted as there not being clear evidence that this diagnostic test is effective, but as clearly almost the entire range of values  $[\underline{AUC}, \overline{AUC}] = [0.4959, 0.8084]$  is larger than 0.5 one would expect the test to be at least of some use. The fact that this range is quite large is due to the relatively few observations in this example, together with the

inclusion of outcomes in the same categories in the NPI upper probability but not in the NPI lower probability, and the fact that there is quite substantial overlap between the data for the non-disease and disease groups.

Let us finally consider Youden's index for the choice of optimal cut-off point  $k$  to define the diagnostic test which best discriminates between the non-disease and disease groups on the basis of these data. Table 4.4 presents the values of Youden's index  $J_e(k)$  for the empirical ROC curve together with Youden's indices corresponding to the NPI lower and upper ROC curves,  $\underline{J}(k)$  and  $\bar{J}(k)$ , respectively. These indices are all maximal for  $k = 5$ , leading to the optimal diagnostic test being such that an outcome in category  $C_5$  indicates disease while categories  $C_1 - C_4$  indicate non-disease.

$k$	1	2	3	4	5
$\underline{J}(k)$	0	0.2067	0.1556	0.2253	0.2485
$J_e(k)$	0	0.2500	0.2000	0.2750	0.3000
$\bar{J}(k)$	0	0.2787	0.2277	0.2973	0.3205

**Table 4.4:** Youden's indices  $J_e(k)$ ,  $\underline{J}(k)$  and  $\bar{J}(k)$

△

**Example 4.2.** This example has been created to briefly illustrate, in comparison to Example 4.1, the effect of more data on the inferences presented in this chapter. The data used for this illustration are presented in Table 4.5.

Disease status	Test results ( $X$ )								Total
	1	2	3	4	5	6	7	8	
No disease ( $D = 0$ )	20	25	30	15	11	10	8	5	124
Disease ( $D = 1$ )	11	7	10	14	13	30	35	40	160

**Table 4.5:** Created data

The NPI lower and upper  $FPF(k)$  and  $TPF(k)$  are presented in Table 4.6, together with empirical values  $FPF_e(k)$  and  $TPF_e(k)$ . The NPI lower and upper

$k$	1	2	3	4	5	6	7	8
$\underline{FPF}(k)$	1	0.8320	0.6320	0.3920	0.2720	0.1840	0.1040	0.0400
$FPF_e(k)$	1	0.8387	0.6371	0.3951	0.2742	0.1855	0.1048	0.0403
$\overline{FPF}(k)$	1	0.8400	0.6400	0.4000	0.2800	0.1920	0.1120	0.0480
$\underline{TPF}(k)$	1	0.9255	0.8820	0.8199	0.7329	0.6522	0.4658	0.2484
$TPF_e(k)$	1	0.9312	0.8875	0.8250	0.7375	0.6562	0.4687	0.2500
$\overline{TPF}(k)$	1	0.9317	0.8882	0.8261	0.7391	0.6584	0.4721	0.2547

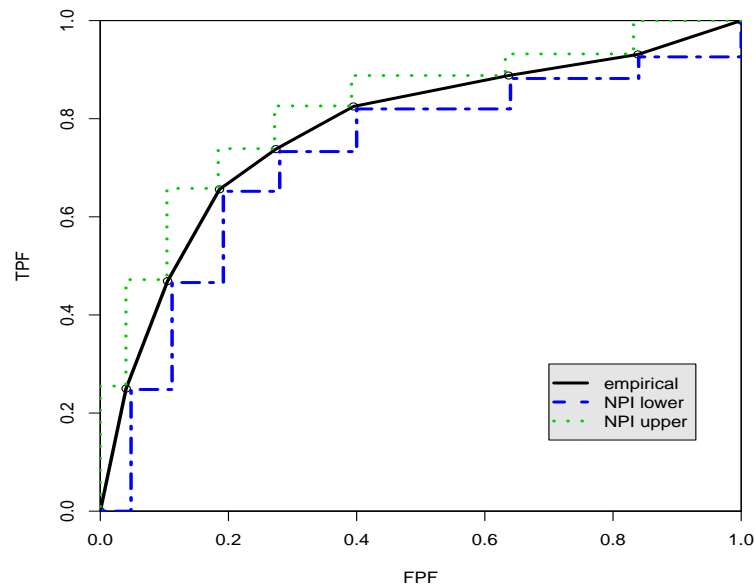
**Table 4.6:** NPI lower, upper and empirical  $FPF$  and  $TPF$  (created data)

ROC curves,  $\underline{ROC}$  and  $\overline{ROC}$ , are given in Figure 4.5, together with the empirical ROC curve. The area under the NPI lower ROC curve is  $\underline{AUC} = 0.7198$ , which is equal to the NPI lower probability  $\underline{P}(X_{125}^0 < X_{161}^1) = 0.7198$ . The area under the NPI upper ROC curve is  $\overline{AUC} = 0.8247$ , which is equal to the NPI upper probability  $\overline{P}(X_{125}^0 \leq X_{161}^1) = 0.8247$ . It is clear that there is substantially less imprecision (i.e. distance between corresponding upper and lower curves and probabilities) in this example than in Example 4.1. This is due to the substantially larger numbers of observations in both the disease and non-disease groups. As the area under the NPI lower ROC curve is greater than 0.5, these two groups can be quite well distinguished. Table 4.7 presents the values of Youden's index  $J_e(k)$  for the empirical ROC curve together with Youden's indices corresponding to the NPI lower and upper ROC curves,  $\underline{J}(k)$  and  $\overline{J}(k)$ , respectively. These indices are all maximal for  $k = 6$ , leading to the optimal diagnostic test being such that an outcome in category  $C_6 - C_8$  indicates disease while categories  $C_1 - C_5$  indicate non-disease.

$k$	1	2	3	4	5	6	7	8
$\underline{J}(k)$	0	0.0855	0.2420	0.4199	0.4529	0.4602	0.3538	0.2004
$J_e(k)$	0	0.0925	0.2504	0.4298	0.4633	0.4708	0.3639	0.2097
$\overline{J}(k)$	0	0.0997	0.2562	0.4341	0.4671	0.4744	0.3681	0.2147

**Table 4.7:** Youden's indices (created data)





**Figure 4.5:** Lower, upper and empirical ROC curves (created data)

## 4.7 Concluding remarks

This chapter has introduced NPI for accuracy of diagnostic tests with ordinal outcomes, with focus on analysis based on the ROC curve. NPI lower and upper ROC curves for this scenario have been introduced, and it was proven that the areas under these curves correspond to NPI lower and upper probabilities in a manner that is consistent with the traditional theory of ROC curves. A general attractive feature of NPI is that it is a frequentist statistics framework which is exactly calibrated [46], for any number of observations, so it can be applied in all scenarios and is particularly useful if there is little information about underlying population distributions, or if one explicitly does not wish to take such information into account. In this chapter only a single further individual from each group has been considered, NPI enables inference on multiple future individuals per group, taking into account the inter-dependence of these individuals [3]. This provides an interesting topic for future research. First NPI for ordinal data must be developed for multiple future observations, this can be followed by application to the scenario considered in this chapter, where the explicit focus on multiple further individuals requires careful

consideration of the events considered to express diagnostic accuracy.

It should be emphasized that the results presented in this chapter are explicitly aimed at a single further patient from each group, so they are not aimed at decisions for multiple future individuals. If one gets new information, for example on further patients to which the test is applied, this can be included in the data set and the analysis can be repeated, clearly providing a statistical method that is fully adaptive to the data. One could argue against the use of inferential methods which are explicitly valid for single future observations only, so here to support the choice of treatment method for a single future patient. But for many practical situations such a method is of direct relevance as one may explicitly have to make a decision for a single future patient. It may also be useful to consider such inferences jointly with the more traditional population-based statistical inferences, which typically require either stronger modelling assumptions or the use of hypotheses tests, the outcomes of which are often quite difficult to interpret.

This introduction of NPI for accuracy of ordinal diagnostic tests has only considered the basic aspects of the theory. In applied clinical experiments there are many factors that must be taken into account in the statistical approach used to analyse the test results [55, 62, 73]. This leads to many research challenges for NPI and its application, with for example the development of NPI for regression models being required to enable co-variates to be taken into account.

# Chapter 5

## Three-group ROC analysis for continuous data

### 5.1 Introduction

In this chapter, the NPI approach is presented for three way Receiver Operating Characteristic (ROC) surfaces. ROC curves have been used to measure the accuracy of diagnostic tests in discriminating between two groups, for example non-disease and disease in medical applications. Recently, ROC methodology was extended to three-group diagnostic problems by constructing a three-dimensional surface [34, 51, 53]. The volume under the ROC surface, VUS, is a direct extension of the area under the ROC curve (AUC). It is a global accuracy measure of a test to classify subjects in three groups. The interpretation of the VUS is similar to the AUC, it is equal to the probability that three random measurements, one from each group, appear in the correct order [52, 71].

Section 5.2 gives an introduction to the concepts of ROC surface for three-group diagnostic tests. The NPI lower and upper ROC surfaces and the volumes under such surfaces are considered and discussed in Section 5.3. In Section 5.4 we present three examples to illustrate and discuss the NPI lower and upper surfaces presented in this chapter. This chapter is finished with some concluding remarks in Section 5.5.

## 5.2 ROC surface for three-group diagnostic tests

In this section, we consider diagnostic accuracy when the diseases have three ordered groups (classes). Such classifications are quite common, for example Wians et al [70] use three groups for iron deficiency related anemia. Non-pregnant women with anemia and a ferritin concentration less than  $20 \mu\text{g}/\text{l}$  were considered to have iron deficiency anemia (*IDA*), the ones with anemia and a ferritin concentration greater than  $240 \mu\text{g}/\text{l}$  were considered to have anemia of chronic disease (*ACD*), while the ones with a ferritin concentration between 20 and  $240 \mu\text{g}/\text{l}$  were considered to belong to the intermediate group [63]. Therefore, it is important to have a good diagnostic test which can discriminate between the disease groups, as patients at different disease groups needs different treatments. In the three-group case, the *ROC* surface is generated in three dimensions by considering all possible diagnostic test values. In what follows, we illustrate the definition and construction of a nonparametric *ROC* surface as introduced by Nakas and Yiannoutsos [52, 53].

Let there be three independent ordered groups of disease, represented  $X$ ,  $Y$  and  $Z$ . Let  $x_1, x_2, \dots, x_{n_x}$  denote the observed test results on  $n_x$  subjects from group  $X$ ;  $y_1, y_2, \dots, y_{n_y}$  the observed test results on  $n_y$  subjects from group  $Y$  and  $z_1, z_2, \dots, z_{n_z}$  the observed test results on  $n_z$  subjects from group  $Z$ . The test results for different groups are assumed to be independent. Suppose that a continuous diagnostic test is used for the discrimination of the patients from these groups, we assume that measurements from group  $X$  tend to be lower than those from group  $Y$ , which in turn tend to be lower than those from group  $Z$ . Let  $F_x, F_y$  and  $F_z$  be the corresponding cumulative distribution functions (CDFs) for the test outcomes of the three groups of disease.

In the case of continuous data, two ordered decision threshold points  $c_1 < c_2$  are needed in order to classify three groups of disease. A diagnostic decision for each patient is based on the following rule:

Denote the whole set of observations for the test result as  $T = T_1, \dots, T_{n_x+n_y+n_z}$ .

1. If  $T_j \leq c_1$  then subject  $j$  is classified into group  $X$ .
2. If  $c_1 < T_j \leq c_2$  then subject  $j$  is classified into group  $Y$ .

3. If  $T_j > c_2$  then then subject  $j$  is classified into group  $Z$ .

This procedure is applied for all subjects, so for all  $j = 1, \dots, n_x + n_y + n_z$ . The correct classification probabilities for the three groups can be computed for a pair of thresholds  $(c_1, c_2)$ ,

1.  $p_1 = P(X \leq c_1) = F_x(c_1)$  is the probability of correct classification of a subject belonging to group  $X$ ,
2.  $p_2 = P(c_1 < Y \leq c_2) = F_y(c_2) - F_y(c_1)$  is the probability of correct classification of a subject belonging to group  $Y$ ,
3.  $p_3 = P(Z > c_2) = 1 - F_z(c_2)$  is the probability of correct classification of a subject belonging to group  $Z$ .

Using these probabilities for correct classification of subjects in the different groups, we can define the Receiver Operating Characteristic (ROC) surfaces as follows.

**Definition 5.1.** *The ROC surface is constructed by plotting the points*

$$P = \{(p_1, p_2, p_3) \mid 0 \leq p_1, p_2, p_3 \leq 1\} \quad (5.1)$$

for all possible real valued  $c_1$  and  $c_2$  with  $c_1 < c_2$

The functional form of the ROC surface can be obtained by writing  $p_2$  as a function of  $p_1$  and  $p_3$  [47, 53, 63], as follows.

**Definition 5.2.** *For  $0 \leq p_1, p_3 \leq 1$ ,*

$$ROC_s(p_1, p_3) = \begin{cases} F_y(F_z^{-1}(1 - p_3)) - F_y(F_x^{-1}(p_1)) & \text{if } F_x^{-1}(p_1) \leq F_z^{-1}(1 - p_3) \\ 0 & \text{otherwise} \end{cases} \quad (5.2)$$

where  $F^{-1}(p)$  is the inverse function of  $F$  for  $p \in [0, 1]$ .

For continuous CDFs,  $F_x, F_y$  and  $F_z$ , these two definitions lead to the same ROC surface [53]. But they do not always lead to the same ROC surface for non-continuous CDFs. Throughout this chapter, we will use Definition 5.2 to develop the

NPI analysis for the ROC surface. Development of NPI corresponding to Definition 5.1, and comparisons of the results with those presented in this thesis, is left as a topic for future research.

A nonparametric estimator of the ROC surface can be obtained by replacing all the distribution functions in (5.2) with their empirical counterparts [9, 47], so that the empirical ROC surface can be constructed as the plot consisting of the points

$$(p_1, \max(0, \widehat{ROC}_s(p_1, p_3)), p_3) \quad (5.3)$$

where  $\widehat{ROC}_s(p_1, p_3) = \widehat{F}_y(\widehat{F}_z^{-1}(1 - p_3)) - \widehat{F}_y(\widehat{F}_x^{-1}(p_1))$  for  $0 \leq p_1, p_3 \leq 1$ , with

$$\widehat{F}_x^{-1}(p) = \begin{cases} -\infty & \text{if } p = 0 \\ x_i & \text{if } p \in (\frac{i-1}{n_x}, \frac{i}{n_x}], \quad i = 1, \dots, n_x \end{cases} \quad (5.4)$$

and

$$\widehat{F}_z^{-1}(p) = \begin{cases} -\infty & \text{if } p = 0 \\ z_l & \text{if } p \in (\frac{l-1}{n_z}, \frac{l}{n_z}], \quad l = 1, \dots, n_z \end{cases} \quad (5.5)$$

where  $x_1 < x_2 < \dots < x_{n_x}$  and  $z_1 < z_2 < \dots < z_{n_z}$ . The empirical cumulative distribution functions  $F_x, F_y$  and  $F_z$  are defined by [52, 54]

$$\widehat{F}_x(c) = \sum_{i=1}^{n_x} \frac{I(x_i \leq c)}{n_x} \quad (5.6)$$

$$\widehat{F}_y(c) = \sum_{j=1}^{n_y} \frac{I(y_j \leq c)}{n_y} \quad (5.7)$$

$$\widehat{F}_z(c) = \sum_{l=1}^{n_z} \frac{I(z_l \leq c)}{n_z} \quad (5.8)$$

where  $I(A)$  is the indicator function, which is equal to one if  $A$  is true and zero otherwise.

### 5.2.1 Volume under the ROC Surface (VUS)

The volumes under the ROC Surface (VUS) has been proposed as a summary measure for the ROC surface and can be used as a global measure of the three-group discriminatory ability of the test under consideration. The VUS equals to the probability that three randomly selected measurements, one from each disease group,

have the correct ordering  $X < Y < Z$ , which is a straightforward generalization of the two-dimensional ROC curve and the AUC [51,52] as reviewed in Chapter 2. An unbiased nonparametric estimator of the VUS is given by [52,53]

$$\widehat{VUS} = \frac{1}{n_x n_y n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{l=1}^{n_z} I(x_i < y_j < z_l) \quad (5.9)$$

Equation (5.9) gives the proportion of all possible three-patient combinations from the data that are ordered correctly. The  $\widehat{VUS}$  can vary from zero to one. It is about 1/6 if the diagnostic test outcomes for the three groups completely overlap, in which case the test is of little use. If there is perfect separation of the test results for the three groups, so  $x_i < y_j < z_l$  for all measurements, then  $\widehat{VUS}=1$ . In practice, ties between measurements may occur, in that case  $I(x, y, z)$  can be defined as follows [52,53]

$$I(x, y, z) = \begin{cases} 1 & \text{if } x < y < z \\ 1/2 & \text{if } x < y = z \text{ or } x = y < z \\ 1/6 & \text{if } x = y = z \\ 0 & \text{otherwise} \end{cases} \quad (5.10)$$

In this chapter, we assume that no ties occur in the data. Ties can be dealt with in NPI by assuming that tied observations differ by small amounts which tend to zero [43].

### 5.2.2 Cut-off point selection in three-group classification

The selection of optimal cut-off points  $c_1$  and  $c_2$ , also called the threshold values, is of course an important aspect of analysis of the quality of a diagnostic test. Several approaches for the selection of the optimal cut-off points for ROC analysis have been considered [40,57], one of these approaches is Youden's index [72] which we will use in order to select the optimal cut-off point. As mentioned in Chapter 4, in case of two groups this index is defined as  $J(c) = \text{sens}(c) + \text{spec}(c) - 1$ , for a given cut-off point  $c$ , and the optimal cut-off point is the value of  $c$  which maximises  $J(c)$  [39,40,54,72].

Nakas et al [54] extended the optimal cut-off point selection method based on Youden's index to the ROC analysis with three groups. The three-group Youden's index given by Nakas et al [54] is

$$\begin{aligned} J_3(c_1, c_2) &= p_1 + p_2 + p_3 = P(X \leq c_1) + P(c_1 < Y \leq c_2) - P(Z \leq c_2) + 1 \\ &= F_x(c_1) + F_y(c_2) - F_y(c_1) - F_z(c_2) + 1 \end{aligned} \quad (5.11)$$

The optimal cut-off points are the values of  $c_1, c_2$  which maximise  $J_3(c_1, c_2)$ , and these can be used for the diagnoses.  $J_3(c_1, c_2)$  is equal to 1 if the three distributions  $F_x, F_y$  and  $F_z$  are identical and  $J_3(c_1, c_2) = 3$  if the distributions  $F_x, F_y$  and  $F_z$  are perfectly discriminated, so if  $P(X < Y < Z) = 1$ . The empirical estimator for  $J_3(c_1, c_2)$ , using the empirical CDFS for  $F_x, F_y$  and  $F_z$ , is equal to

$$\hat{J}_3(c_1, c_2) = \hat{F}_x(c_1) + \hat{F}_y(c_2) - \hat{F}_y(c_1) - \hat{F}_z(c_2) + 1 \quad (5.12)$$

### 5.3 NPI analysis for the ROC surface

Recall that  $X, Y$  and  $Z$  are continuous random quantities representing the test results for the three disease groups, which can e.g. be disease, intermediate and healthy. To develop the nonparametric predictive inference (NPI) approach for this analysis, let  $X_{n_x+1}, Y_{n_y+1}$  and  $Z_{n_z+1}$  be the next future observations from these groups, respectively. We apply  $A_{(n)}$  per group, so we need to introduce the following notation. The ordered observed values of  $X_1, X_2, \dots, X_{n_x}$ , are denoted by  $x_1 < x_2 < \dots < x_{n_x}$ , and let  $x_0 = -\infty$  and  $x_{n_x+1} = \infty$  for ease of notation. For  $X_{n_x+1}$ , representing a future observation from group  $X$ , the assumption  $A_{(n_x)}$  is  $P(X_{n_x+1} \in (x_{i-1}, x_i)) = \frac{1}{n_x+1}$  for  $i = 1, \dots, n_x + 1$ . Similarly, let the ordered observed values of  $Y_1, Y_2, \dots, Y_{n_y}$  be denoted by  $y_1 < y_2 < \dots < y_{n_y}$ , and  $y_0 = -\infty$  and  $y_{n_y+1} = \infty$ . For  $Y_{n_y+1}$ , representing a future observation from group  $Y$ ,  $A_{(n_y)}$  gives  $P(Y_{n_y+1} \in (y_{j-1}, y_j)) = \frac{1}{n_y+1}$  for  $j = 1, \dots, n_y + 1$ . And finally, let the ordered observed values of  $Z_1, Z_2, \dots, Z_{n_z}$  be denoted by  $z_1 < z_2 < \dots < z_{n_z}$ , and  $z_0 = -\infty$  and  $z_{n_z+1} = \infty$ . For  $Z_{n_z+1}$ , representing a future observation from group  $Z$ ,  $A_{(n_z)}$  gives  $P(Z_{n_z+1} \in (z_{l-1}, z_l)) = \frac{1}{n_z+1}$  for  $l = 1, \dots, n_z + 1$ .



Notice that, the NPI lower and upper cumulative distribution functions for  $X_{n_x+1}$ , at  $c \neq x_i$ , for all  $i = 1, \dots, n_x$ , are

$$\underline{F}_x(c) = \underline{P}(X_{n_x+1} \leq c) = \frac{\sum_{i=1}^{n_x} I(x_i \leq c)}{n_x + 1} \quad (5.13)$$

$$\overline{F}_x(c) = \overline{P}(X_{n_x+1} \leq c) = \frac{\sum_{i=1}^{n_x} I(x_i \leq c) + 1}{n_x + 1} \quad (5.14)$$

at  $c = x_i$ ,  $\underline{F}_x(x_i) = \overline{F}_x(x_i) = \frac{i}{n_x+1}$ , so at the data observations the NPI lower and upper CDFs are identical. The NPI lower and upper CDFs for  $Y_{n_y+1}$  and  $Z_{n_z+1}$  are similar and denoted by  $\underline{F}_y, \overline{F}_y, \underline{F}_z$  and  $\overline{F}_z$ .

### 5.3.1 NPI lower and upper ROC surface

To define NPI lower and upper ROC surfaces according to Equation (5.2) in Definition 5.2, we follow the process illustrated in Figure 5.1.

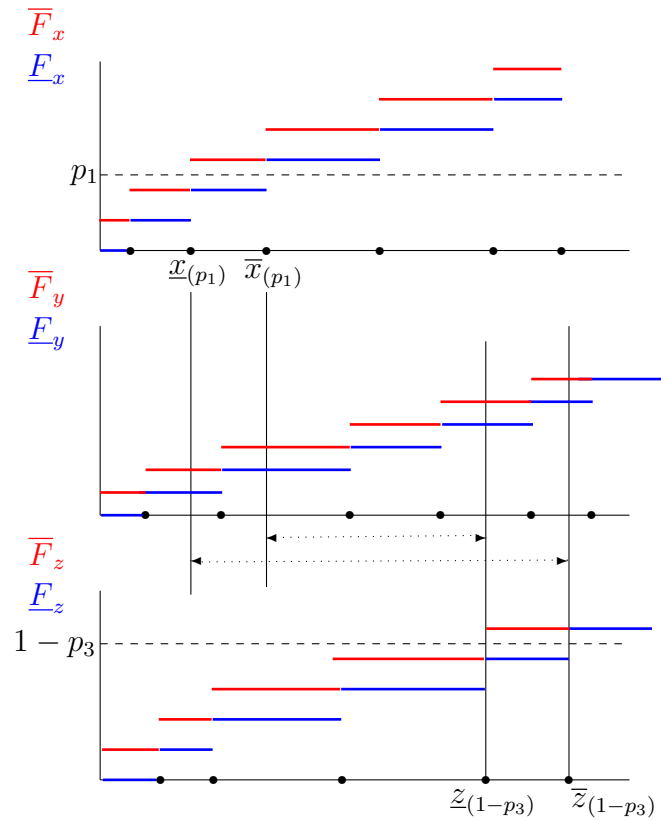


Figure 5.1: Construction of NPI lower and upper ROC surfaces

In NPI we have lower and upper CDFs, so the inverse CDFs in Equation (5.2) are not uniquely defined any more. For a value  $p_1 \in [0, 1]$ , we can define the interval  $(\underline{x}_{(p_1)}, \bar{x}_{(p_1)})$  such that, for every precise CDF which is entirely between  $\underline{F}_x, \bar{F}_x$ , the inverse function value at  $p_1$  falls within this interval. These values  $\underline{x}_{(p_1)}$  and  $\bar{x}_{(p_1)}$  are the points at which the NPI lower and upper CDFs  $\underline{F}_x$  and  $\bar{F}_x$  change value, hence they are equal to the observed values, or possibly  $-\infty$  or  $+\infty$ .

From the first graph in Figure 5.1 it is clear that

$$\underline{F}_x(\underline{x}_{(p_1)}) < p_1 < \bar{F}_x(\bar{x}_{(p_1)}) \quad (5.15)$$

So  $\underline{x}_{(p_1)}$  and  $\bar{x}_{(p_1)}$  take on consecutive the observed values  $x_1, \dots, x_{n_x}$  and  $-\infty$  and  $+\infty$ . While at these points  $\underline{F}_x(x_i) = \bar{F}_x(x_i) = \frac{i}{n_x+1}, i = 1, \dots, n_x + 1$ , so there is precision, for the open intervals  $(x_{i-1}, x_i)$  there is imprecision, so  $\underline{F}_x(x) = \frac{i-1}{n_x+1}$  and  $\bar{F}_x(x) = \frac{i}{n_x+1}$  for  $x \in (x_{i-1}, x_i)$ .

All different  $p_1 \in [0, 1]$ , lead to  $n_x + 1$  pairs  $(\underline{x}_{(p_1)}, \bar{x}_{(p_1)})$ , which are equal to  $(x_{i-1}, x_i)$  for  $p_1 \in (\frac{i-1}{n_x+1}, \frac{i}{n_x+1})$   $i = 1, \dots, n_x + 1$ . If  $p_1 = \frac{i}{n_x+1}$  for  $i = 1, \dots, n_x$ , then  $(\underline{x}_{(p_1)}, \bar{x}_{(p_1)}) = (x_{i-1}, x_{i+1})$ , we neglect this further as it does not influence the volume under the ROC surface.

Similarly, from the third graph in Figure 5.1, it is clear that

$$\underline{F}_z(\underline{z}_{(1-p_3)}) < 1 - p_3 < \bar{F}_z(\bar{z}_{(1-p_3)}) \quad (5.16)$$

Here  $\underline{z}_{(1-p_3)}$  and  $\bar{z}_{(1-p_3)}$  take on consecutive the observed values  $z_1, \dots, z_{n_z}$  and  $-\infty$  and  $+\infty$ , at these points  $\underline{F}_z(z_l) = \bar{F}_z(z_l) = \frac{l}{n_z+1}, l = 1, \dots, n_z + 1$  there is precision but for the open intervals  $(z_{l-1}, z_l)$  there is imprecision, with  $\underline{F}_z(z) = \frac{l-1}{n_z+1}$  and  $\bar{F}_z(z) = \frac{l}{n_z+1}$  for  $z \in (z_{l-1}, z_l)$ .

All different  $1 - p_3 \in [0, 1]$ , lead to  $n_z + 1$  pairs  $(\underline{z}_{(1-p_3)}, \bar{z}_{(1-p_3)})$ , which are equal to  $(z_{l-1}, z_l)$  for  $1 - p_3 \in (\frac{l-1}{n_z+1}, \frac{l}{n_z+1})$   $l = 1, \dots, n_z + 1$ . If  $1 - p_3 = \frac{l}{n_z+1}$  for  $l = 1, \dots, n_z$ , then  $(\underline{z}_{(1-p_3)}, \bar{z}_{(1-p_3)}) = (z_{l-1}, z_{l+1})$ , we neglect this further as it does not influence the volume under the ROC surface.

Now we know the intervals for the inverse values of the NPI lower and upper CDFs for group  $X$  and  $Z$ , we complete the generalization of Equation (5.2) by deriving the corresponding bounds for  $ROC_s(p_1, p_3)$ , where we also take the imprecision

in the lower and upper CDFs  $\underline{F}_y$  and  $\overline{F}_y$  into account. To derive the lower bound  $\underline{ROC}_s(p_1, p_3)$ , we take the minimal range resulting from the imprecise inverse CDFs as discussed above, so  $(\overline{x}_{(p_1)}, \underline{z}_{(1-p_3)})$ , as long as  $\underline{z}_{(1-p_3)} \geq \overline{x}_{(p_1)}$ , and we take the minimal probability mass for  $Y_{n_y+1}$  which has to be in this interval, this leads to Definition 5.3.

**Definition 5.3.** *The NPI lower ROC surface is defined by*

$$\underline{ROC}_s(p_1, p_3) = \begin{cases} \underline{F}_y(\underline{z}_{(1-p_3)}) - \overline{F}_y(\overline{x}_{(p_1)}) & \text{if } \underline{z}_{(1-p_3)} \geq \overline{x}_{(p_1)} \\ 0 & \text{otherwise} \end{cases} \quad (5.17)$$

To derive the upper bound  $\overline{ROC}_s(p_1, p_3)$ , we take the maximal range resulting from the imprecise inverse CDFs, so  $(\underline{x}_{(p_1)}, \overline{z}_{(1-p_3)})$ , as long as  $\overline{z}_{(1-p_3)} \geq \underline{x}_{(p_1)}$ , and we take the maximal probability mass for  $Y_{n_y+1}$  which has to be in this interval, this leads to Definition 5.4.

**Definition 5.4.** *The NPI upper ROC surface is defined by*

$$\overline{ROC}_s(p_1, p_3) = \begin{cases} \overline{F}_y(\overline{z}_{(1-p_3)}) - \underline{F}_y(\underline{x}_{(p_1)}) & \text{if } \underline{x}_{(p_1)} \leq \overline{z}_{(1-p_3)} \\ 0 & \text{otherwise} \end{cases} \quad (5.18)$$

These functions  $\underline{ROC}_s(p_1, p_3)$  and  $\overline{ROC}_s(p_1, p_3)$  are the tightest bounds can be derived and justified for the ROC surface for the three-group diagnostic tests in the NPI framework.

### 5.3.2 Volumes under the NPI lower and upper ROC surfaces

The volumes under the NPI lower and upper ROC surfaces are denoted by  $\underline{VUS}$  and  $\overline{VUS}$ , respectively, and these are given in Theorem 5.1.

**Theorem 5.1.** The volumes under the NPI lower and upper ROC surfaces are

$$\underline{VUS} = A \left[ \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(x_i < y_{j-1} \wedge y_j < z_{l-1}) \right] \quad (5.19)$$

$$\overline{VUS} = A \left[ \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(x_{i-1} < y_j \wedge x_{i-1} < z_l \wedge y_{j-1} < z_l) \right] \quad (5.20)$$

where  $A = \frac{1}{(n_x+1)(n_y+1)(n_z+1)}$ .

*Proof.* The volume under the NPI lower ROC surface is derived as follows, where  $\{A\}^+ = \max\{A, 0\}$  and  $\sum_{p_1} \sum_{p_3}$  indicate the sums over pairs of values for  $p_1$  and  $p_3$  such that one value for  $p_1$  is taken from each interval  $(\frac{i-1}{n_x+1}, \frac{i}{n_x+1})$  for  $i = 1, \dots, n_x+1$  and one value for  $p_3$  from each interval  $(\frac{l-1}{n_z+1}, \frac{l}{n_z+1})$   $l = 1, \dots, n_z+1$ . As the function  $\underline{ROC}_s(p_1, p_3)$  is constant for all values  $p_1 \in (\frac{i-1}{n_x+1}, \frac{i}{n_x+1})$  and  $p_3 \in (\frac{l-1}{n_z+1}, \frac{l}{n_z+1})$ , it does not matter which values for  $p_1$  and  $p_3$  within these intervals are actually used in the calculations.

$$\underline{VUS} = \frac{1}{(n_x+1)(n_z+1)} \sum_{p_1} \sum_{p_3} \underline{ROC}_s(p_1, p_3) = \quad (5.21)$$

$$\frac{1}{(n_x+1)(n_z+1)} \sum_{p_1} \sum_{p_3} \{F_y(z_{(1-p_3)}) - \bar{F}_y(\bar{x}_{(p_1)})\}^+ =$$

$$\frac{1}{(n_x+1)(n_z+1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} \{F_y(z_{l-1}) - \bar{F}_y(x_i)\}^+ =$$

$$\frac{1}{(n_x+1)(n_z+1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} \left\{ \frac{\sum_{j=1}^{n_y} I(y_j \leq z_{l-1})}{n_y+1} - \frac{\sum_{j=1}^{n_y} I(y_j \leq x_i) + 1}{n_y+1} \right\}^+ =$$

$$\frac{1}{(n_x+1)(n_z+1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} \left\{ \frac{\sum_{j=1}^{n_y+1} I(y_j \leq z_{l-1})}{n_y+1} - \frac{\sum_{j=1}^{n_y+1} I(y_{j-1} \leq x_i)}{n_y+1} \right\}^+ =$$

$$\frac{1}{(n_x+1)(n_y+1)(n_z+1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} \{I(y_j \leq z_{l-1}) - I(y_{j-1} \leq x_i)\}^+ =$$

$$A \left[ \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(y_j \leq z_{l-1} \wedge y_{j-1} > x_i) \right] \quad (5.22)$$

So, in calculating the volume under the NPI lower ROC surface, we only take into account the intervals  $(x_{i-1}, x_i)$ ,  $(y_{j-1}, y_j)$  and  $(z_{l-1}, z_l)$ , which are totally separated, with  $x_i < y_{j-1} < y_j < z_{l-1}$ . Equation (5.21) follows from the definition of the

volume under the surface. Equation (5.22) follows from the fact that for two events  $E_1$  and  $E_2$ , the equality  $E_1 - E_2 = E_1 \wedge E_2^c$  holds. Similarly, the volume under the NPI upper ROC surface is

$$\begin{aligned}
\overline{VUS} &= \frac{1}{(n_x + 1)(n_z + 1)} \sum_{p_1} \sum_{p_3} \overline{ROC}_s(p_1, p_3) = \\
&= \frac{1}{(n_x + 1)(n_z + 1)} \sum_{p_1} \sum_{p_3} I(\underline{x}_{(p_1)} \leq \bar{z}_{(1-p_3)}) \{ \bar{F}_y(\bar{z}_{(1-p_3)}) - \underline{F}_y(\underline{x}_{(p_1)}) \} = \\
&= \frac{1}{(n_x + 1)(n_z + 1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} I(x_{i-1} \leq z_l) \{ \bar{F}_y(z_l) - \underline{F}_y(x_{i-1}) \} = \\
&= \frac{1}{(n_x + 1)(n_z + 1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} I(x_{i-1} \leq z_l) \left\{ \frac{\sum_{j=1}^{n_y} I(y_j \leq z_l) + 1}{n_y + 1} - \frac{\sum_{j=1}^{n_y} I(y_j \leq x_{i-1})}{n_y + 1} \right\} = \\
&= \frac{1}{(n_x + 1)(n_z + 1)} \sum_{i=1}^{n_x+1} \sum_{l=1}^{n_z+1} I(x_{i-1} \leq z_l) \left\{ \frac{\sum_{j=1}^{n_y+1} I(y_{j-1} \leq z_l)}{n_y + 1} - \frac{\sum_{j=1}^{n_y+1} I(y_j \leq x_{i-1})}{n_y + 1} \right\} = \\
&= \frac{1}{(n_x + 1)(n_y + 1)(n_z + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(x_{i-1} \leq z_l) \{ I(y_{j-1} \leq z_l) - I(y_j \leq x_{i-1}) \} = \\
&= \frac{1}{(n_x + 1)(n_y + 1)(n_z + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(x_{i-1} \leq z_l) \{ I(y_{j-1} \leq z_l \wedge y_j > x_{i-1}) \} = \\
&= A \left[ \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \sum_{l=1}^{n_z+1} I(x_{i-1} < y_j \wedge x_{i-1} \leq z_l \wedge y_{j-1} \leq z_l) \right]
\end{aligned}$$

So, to get the volume under the NPI upper ROC surface, we count all combinations of the intervals  $(x_{i-1}, x_i)$ ,  $(y_{j-1}, y_j)$  and  $(z_{l-1}, z_l)$ , for which there are  $x \in (x_{i-1}, x_i)$ ,  $y \in (y_{j-1}, y_j)$  and  $z \in (z_{l-1}, z_l)$ , such that  $x < y < z$ .

□

### 5.3.3 NPI cut-off points selection

Corresponding to Youden's index, as presented in Equation (5.11) in Subsection 5.2.2, we introduce the NPI cut-off points selection. In order to do this, we use the NPI lower and upper probabilities for correct classification into the three groups. The NPI lower probability for correct classification into the first group is

$$\underline{P}(X_{n_x+1} \leq c_1) = \underline{F}_x(c_1) = \frac{\sum_{i=1}^{n_x} I(x_i \leq c_1)}{n_x + 1} \quad (5.23)$$

and the NPI upper probability for correct classification into the first group is

$$\bar{P}(X_{n_x+1} \leq c_1) = \bar{F}_x(c_1) = \frac{\sum_{i=1}^{n_x} I(x_i \leq c_1) + 1}{n_x + 1} \quad (5.24)$$

Let  $n_y(c_1, c_2) = \sum_{j=1}^{n_y} I(c_1 < y_j \leq c_2)$ . The NPI lower probability for correct classification into the second group is

$$\underline{P}(c_1 < Y_{n_y+1} \leq c_2) = \frac{\{n_y(c_1, c_2) - 1\}^+}{n_y + 1} = \{\underline{F}_y(c_2) - \bar{F}_y(c_1)\}^+ \quad (5.25)$$

The corresponding NPI upper probability is

$$\bar{P}(c_1 < Y_{n_y+1} \leq c_2) = \frac{n_y(c_1, c_2) + 1}{n_y + 1} = \bar{F}_y(c_2) - \underline{F}_y(c_1) \quad (5.26)$$

Finally, the NPI lower probability for correct classification into the third group is

$$\underline{P}(Z_{n_z+1} > c_2) = 1 - \bar{P}(Z_{n_z+1} \leq c_2) = 1 - \bar{F}_z(c_2) = 1 - \frac{\sum_{l=1}^{n_z} I(z_l \leq c_2) + 1}{n_z + 1} \quad (5.27)$$

and the corresponding NPI upper probability is

$$\bar{P}(Z_{n_z+1} > c_2) = 1 - \underline{P}(Z_{n_z+1} \leq c_2) = 1 - \underline{F}_z(c_2) = 1 - \frac{\sum_{j=1}^{n_z} I(z_j \leq c_2)}{n_z + 1} \quad (5.28)$$

We introduce the NPI lower value for Youden's index as

$$\begin{aligned} \underline{J}_3(c_1, c_2) &= \underline{P}(X_{n_x+1} \leq c_1) + \underline{P}(c_1 < Y_{n_y+1} \leq c_2) + \underline{P}(Z_{n_z+1} > c_2) \\ &= \underline{F}_x(c_1) + \{\underline{F}_y(c_2) - \bar{F}_y(c_1)\}^+ + 1 - \bar{F}_z(c_2) \end{aligned} \quad (5.29)$$

and the corresponding NPI upper value for Youden's index as

$$\begin{aligned} \bar{J}_3(c_1, c_2) &= \bar{P}(X_{n_x+1} \leq c_1) + \bar{P}(c_1 < Y_{n_y+1} \leq c_2) + \bar{P}(Z_{n_z+1} > c_2) \\ &= \bar{F}_x(c_1) + \bar{F}_y(c_2) - \underline{F}_y(c_1) + 1 - \underline{F}_z(c_2) \end{aligned} \quad (5.30)$$

From Equations (5.23)-(5.28) it straightforwardly follows that, for  $c_1$  and  $c_2$  not equal to any of the observations  $x_1, \dots, x_{n_x}$ ,  $y_1, \dots, y_{n_y}$ ,  $z_1, \dots, z_{n_z}$ , and assuming that there is at least one observation  $y_j$  in  $(c_1, c_2)$ ,

$$\bar{J}_3(c_1, c_2) = \underline{J}_3(c_1, c_2) + \frac{1}{n_x + 1} + \frac{2}{n_y + 1} + \frac{1}{n_z + 1} \quad (5.31)$$

Therefore, maximising,  $\underline{J}_3(c_1, c_2)$  and  $\overline{J}_3(c_1, c_2)$  over such values of  $c_1, c_2$  gives the same solution. Because the NPI lower and upper probabilities within the equations (5.29) and (5.30) are step functions, with steps only occurring at the observed data values, considering  $\underline{J}_3(c_1, c_2)$  and  $\overline{J}_3(c_1, c_2)$  at values  $c_1, c_2$  with at least one of these being equal to one of the data observations is easy to do. It is easy to show for all  $c_1, c_2$ ,

$$\underline{J}_3(c_1, c_2) \leq \hat{J}_3(c_1, c_2) \leq \overline{J}_3(c_1, c_2) \quad (5.32)$$

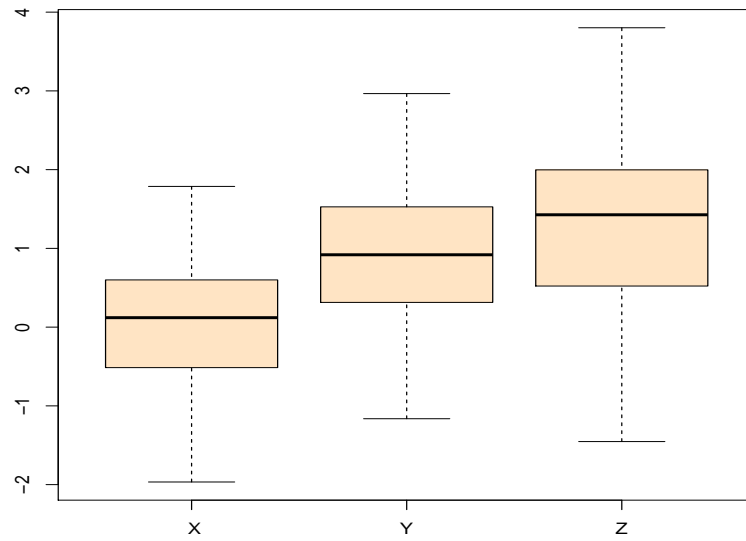
with empirical Youden's index  $\hat{J}_3(c_1, c_2)$  as given in Equation (5.12).

## 5.4 Examples

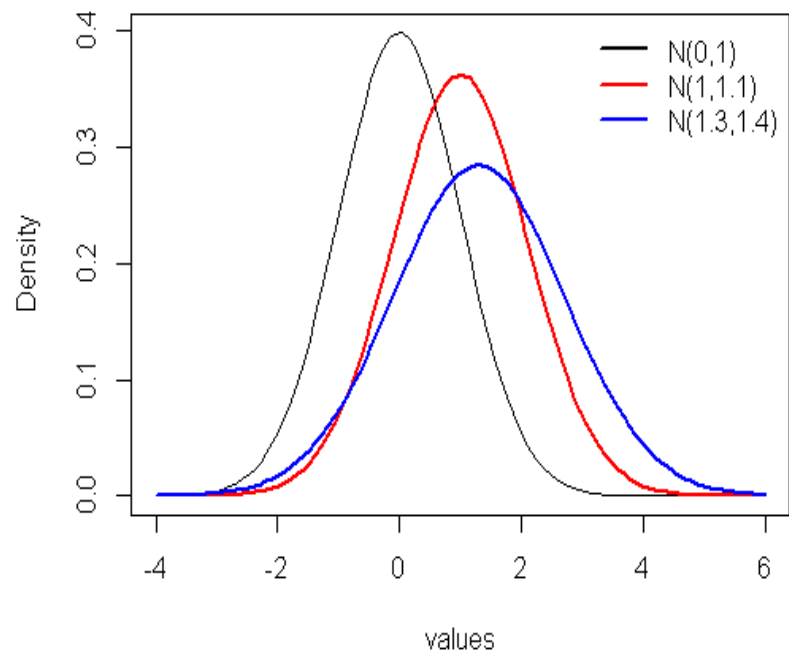
In this section, three examples are given to illustrate the NPI lower and upper ROC surfaces and the volume under these surfaces, as presented in this chapter. The data for the first example are simulated from normal distributions. The second example uses data from the literature. Throughout this section, we use the notation  $\underline{S}_f$  for  $\underline{ROC}_s(p_1, p_3)$ ,  $S_f$  for  $\widehat{ROC}_s(p_1, p_3)$  and  $\overline{S}_f$  for  $\overline{ROC}_s(p_1, p_3)$  in the Figures.

**Example 5.1.** In this example we simulated data from normal distributions as follows: 20 measurements from  $X \sim N(0, 1)$ , 24 measurements from  $Y \sim N(1, 1.1)$  and 22 measurements from  $Z \sim N(1.3, 1.4)$ . The boxplots of these data are presented in Figure 5.2 and the probability density functions of these three distributions are plotted in Figure 5.3.

The NPI lower ROC surface, as defined in Section 5.3 is plotted in Figure 5.4(c). The volume under this NPI lower ROC surface is equal to 0.3091. The corresponding NPI upper ROC surface is plotted in Figure 5.4(a). The volume under this NPI upper ROC surface is equal to 0.4267. Figure 5.4(b) present the empirical ROC surface, the volume under it is 0.3854. These values of the volumes under these ROC surfaces indicate that the test is of some use, as they are greater than 1/6, but the test does not manage to achieve correct classification for a quite large proportion of the subjects. It is clear from Figures 5.2 and 5.3 that particularly the data for  $Y$  and  $Z$  groups have substantial overlap, which will lead to substantial risk of



**Figure 5.2:** Boxplots for the data in Example 5.1



**Figure 5.3:** Probability density functions for simulated data (Example 5.1)

misclassification for both groups, reflected in the empirical and NPI lower and upper ROC surfaces.

Although it is not easy to see in Figure 5.4, the empirical ROC surface is every-

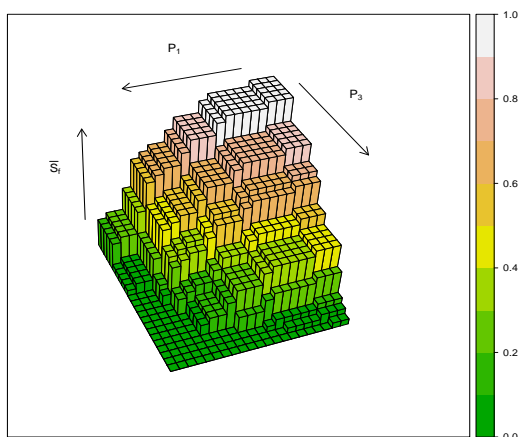


where between the NPI lower and upper ROC surfaces.

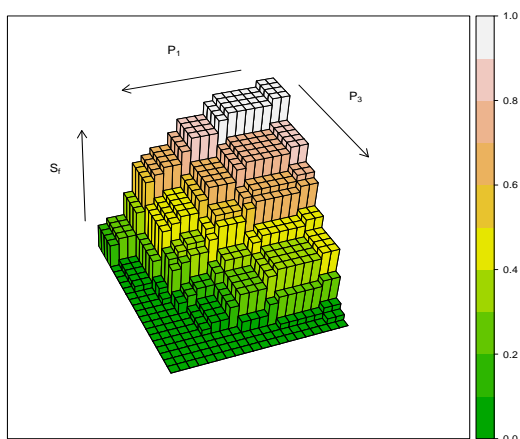
The plots of ROC surfaces in this section are given with  $p_1$  and  $p_3$  both equal to 0 at the right-back corner, while at the left-front corner, both  $p_1$  and  $p_3$  are equal to 1. This is indicated by arrows in the figures.

The maximum value of Youden's index for the empirical ROC surface in this example is equal to 1.6258, which occurs for  $(c_1, c_2) = (0.1293, 1.7715)$  and also for  $(c_1, c_2) = (0.7014, 1.7715)$ . The corresponding NPI lower and upper Youden's indices  $\underline{J}_3(c_1, c_2)$  and  $\bar{J}_3(c_1, c_2)$ , are equal to 0.1293 and 1.6862, respectively, which both occur for  $(c_1, c_2) = (0.1293, 1.7715)$ . Using these values for  $c_1$  and  $c_2$ , a test result of less than or equal to 0.1293 (or 0.7014) leads to the decision to classify the patient to group  $X$ . If the test result is greater than 0.1293 (or 0.7014) and less than or equal to 1.7715 then the decision is group  $Y$ . Finally, if the test result exceeds 1.7715 then the decision is group  $Z$ .

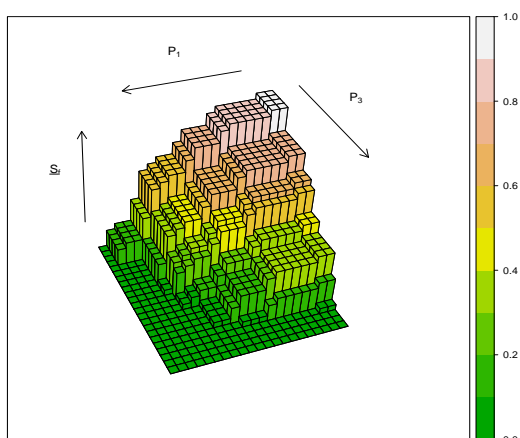
△



(a) Upper ROC surface



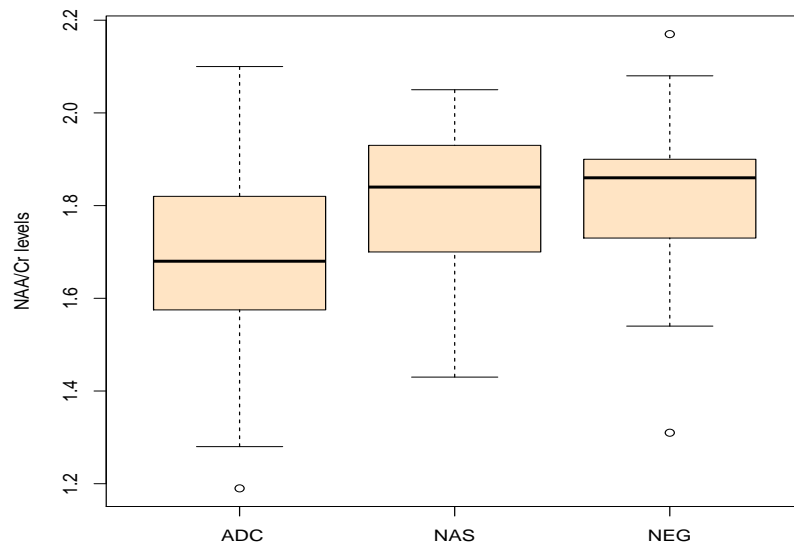
(b) Empirical ROC surface



(c) Lower ROC surface

**Figure 5.4:** Upper, empirical and lower ROC surfaces for Example 5.1

**Example 5.2.** The n-acetyl aspartate over creatinine ( $NAA/Cr$ ) is a diagnostic test used to discriminate between different levels of human immunodeficiency virus ( $HIV$ ) in patients [14, 54, 71]. Nakas et al [47, 54] used the  $NAA/Cr$  test to distinguish between  $HIV$  patients. The  $NAA/Cr$  levels were available in a study of 135 patients, of whom 59 were  $HIV$ -positive with  $AIDS$  dementia complex ( $ADC$ ), 39 were  $HIV$ -positive non-symptomatic subjects ( $NAS$ ), and the other 37 were  $HIV$ -negative individuals ( $NEG$ ). The  $NAA/Cr$  levels are expected to be lowest among the  $ADC$  group and highest among the  $NEG$  group, with the  $NAS$  group being intermediate to the other two, so the anticipated ordering is  $ADC < NAS < NEG$  [14]. Figure 5.5 shows the boxplots of these data. These



**Figure 5.5:** Boxplots of  $NAA/Cr$  levels of  $ADC$ ,  $NAS$  and  $NEG$  groups

boxplots show strong overlap between the different groups, particularly between the  $NAS$  and  $NEG$  groups (groups  $Y$  and  $Z$ , respectively). We consider these data particularly because they were also used by Nakas et al [47, 54] to illustrate ROC surfaces, but the main overlap is likely to lead to quite poor diagnostic accuracy. Example 5.3 will illustrate the application of our new approach to a situation with groups that are less overlapping. The NPI lower ROC surface is plotted in Figure 5.6(c). The volume under this surface is  $\underline{VUS} = 0.2524$ . The corresponding

NPI upper ROC surface is plotted in Figure 5.6(a), the volume under this surface is  $\overline{VUS} = 0.3131$ . Figure 5.6(b) presents the empirical ROC surface, the volume under it is 0.2879. These values of the volumes under these ROC surfaces indicate also that the test is of some use, as they are greater than  $1/6$ , but the test does not manage to achieve correct classification for a quite large proportion of the subjects. It is clear from Figure 5.5 that particularly the data for the  $Y$  and  $Z$  groups have substantial overlap, which will lead to substantial risk of misclassification for both groups, reflected in the empirical and NPI lower and upper ROC surfaces.

The maximum value of Youden's index for the empirical ROC surface in this example is equal to 1.4362, which occurs for  $(c_1, c_2) = (1.76, 2.05)$ . The corresponding NPI lower and upper Youden's indices values are  $\underline{J}_3(c_1, c_2) = 1.3803$  and  $\overline{J}_3(c_1, c_2) = 1.4732$ , which both occur for the same values of  $c_1$  and  $c_2$  as for the empirical ROC surface. So, using these values for  $c_1$  and  $c_2$ , a patient with test result less than or equal to 1.76 is classified into the *ADC* group, while a test result over 2.05 leads to classification into *NEG* group. Test results in between these two values leads to classification into *NAS* group. In this example there is small imprecision due to the large data sets.

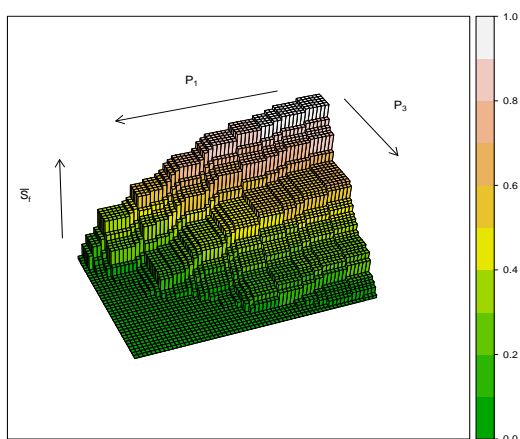
△

As Examples 5.1 and 5.2 both involved data that overlapped quite a bit, we now give a small example with data that are more separated.

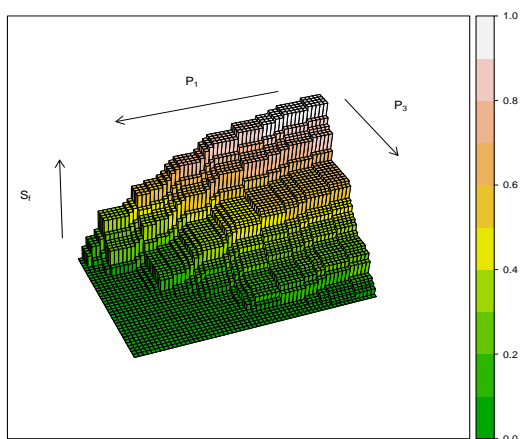
**Example 5.3.** This example consider data in the same range as for Example 5.2, with  $n_x = n_y = n_z = 10$ , where the data are more separated. The data are given in Table 5.1.

	NAA/Cr levels									
$X$	1.28	1.43	1.52	1.53	1.55	1.57	1.6	1.63	1.66	1.64
$Y$	1.65	1.68	1.71	1.76	1.78	1.79	1.80	1.85	1.86	1.87
$Z$	1.83	1.84	1.88	1.89	1.90	1.93	1.96	1.99	2.06	2.08

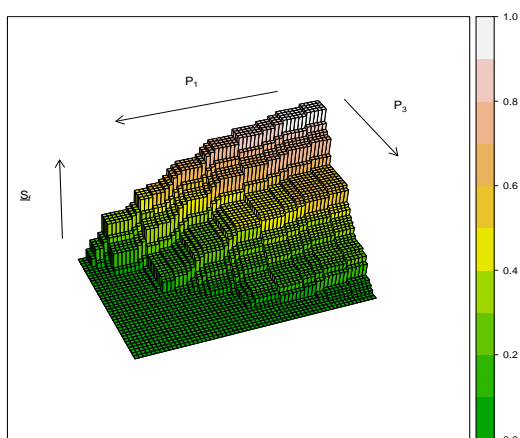
**Table 5.1:** Separated data



(a) Upper ROC surface



(b) Empirical ROC surface



(c) Lower ROC surface

**Figure 5.6:** Upper, empirical and lower ROC surface for Example 5.2

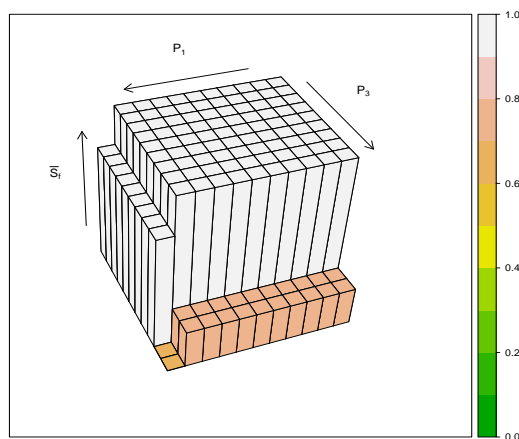
The NPI lower ROC surface is plotted in Figure 5.7(c). The volume under this surface is  $\underline{VUS} = 0.6235$ . The corresponding NPI upper ROC surface is plotted in Figure 5.7(a), the volume under this surface is  $\overline{VUS} = 0.9421$ . Figure 5.7(b) presents the empirical ROC surface, the volume under it is 0.93. These values of the volumes under these ROC surfaces indicate that the test is good for discriminating between the three groups.

The maximum value of Youden's index for the empirical ROC surface in this example is equal to 2.7, which occurs for  $(c_1, c_2) = (1.64, 1.78)$  and also for  $(c_1, c_2) = (1.66, 1.78)$ . The corresponding NPI lower and upper Youden's indices  $\underline{J}_3(c_1, c_2)$  and  $\overline{J}_3(c_1, c_2)$ , are equal to 2.3636 and 2.7273, respectively, both occur for the same values of  $c_1$  and  $c_2$  as for the empirical ROC surface. Therefore, a test result of less than or equal to 1.64 (or 1.66) leads to the decision to classify the patient to group  $X$ . If the test result is greater than 1.64 (or 1.66) and less than or equal to 1.78 then the decision is group  $Y$ . Finally, if the test result exceeds 1.78 then the decision is group  $Z$ . The small data sets in this example, leads to large imprecision than in Examples 5.1 and 5.2.

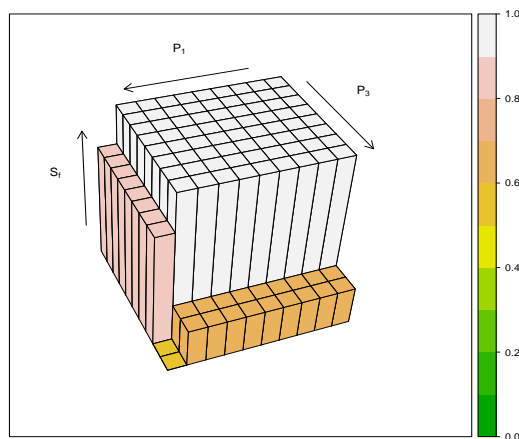
△

## 5.5 Concluding remarks

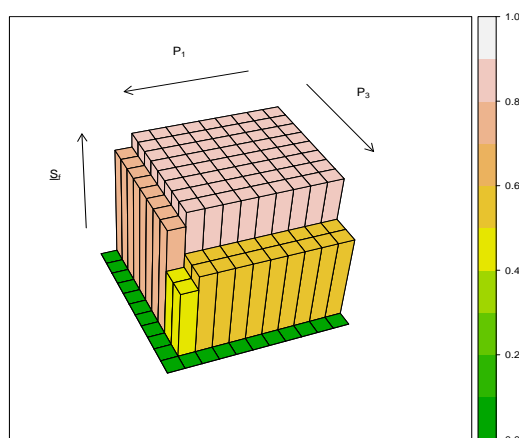
NPI lower and upper ROC surfaces are presented in this chapter for the evaluation of the diagnostic accuracy in three-class classification problems. The corresponding volumes under the NPI lower and upper ROC surfaces have been derived. The choice of cut-off points for such classifications by maximisation of Youden's index is presented. Attention has been restricted to Definition 5.2 for the ROC surface, however, for non-continuous cumulative distribution functions, as is the case for the empirical and NPI ROC analysis, this definition does not always lead to the same results as Definition 5.1. We aim to develop the NPI approach also related to Definition 5.1, and to compare this with the results presented in this chapter. Only real-valued data were considered in this chapter, we also aim to develop such ROC surfaces for ordinal data, this is an interesting topic for future research.



(a) Upper ROC surface



(b) Empirical ROC surface



(c) Lower ROC surface

**Figure 5.7:** Upper, empirical and lower ROC surface for Example 5.3

# Chapter 6

## Concluding remarks

This chapter provides a brief summary of the main results presented in this thesis and some important challenges for future research. In this thesis we have presented Nonparametric Predictive Inference (NPI) for ordinal data and for accuracy of diagnostic tests.

In Chapter 3, multiple comparisons and subset selection for ordinal categorical data were presented. The method uses a latent variable representation of the observations and categories on the real line. NPI lower and upper probabilities are presented for several events involving the next observation, including selecting one group, the subset of best groups, and the subset including the best group.

In Chapter 4, we introduced NPI for accuracy of diagnostic tests with ordinal outcomes, with the inferences based on data for a disease group and a non-disease group. The empirical and NPI lower and upper Receiver Operating Characteristic (ROC) curves and the corresponding areas under the curves were presented and it was proven that these are nested, with the latter equal to the NPI lower and upper probabilities for correctly ordered future observations from the non-disease and disease groups. The use of the Youden index related to the NPI lower and upper ROC curves was discussed in order to determine the optimal cut-off point for the test.

When diagnostic tests have categorical outcomes, these are not always ranked as assumed in Chapter 4. For unordered categorical data, NPI has been developed [22]



and it has already been applied to category selection [6]. There is an interesting opportunity to apply that method to problems involving diagnostic tests, particularly if a loss function can be identified that quantifies the consequences of wrong diagnoses. Such a loss function can also be useful for deciding the optimal cut-off value for the ordinal tests in Chapter 4, based on other criteria than the Youden's index. It seems particularly attractive that such a loss function would explicitly be for the possible consequences to a further individual from either the disease or non-disease group, which may be easier to elicit from medical experts than in terms of proportions of assumed underlying populations.

In Chapter 5, NPI lower and upper ROC surfaces have been proposed for the evaluation of the diagnostic accuracy in ordered three-class classification problems as a direct generalization of the NPI lower and upper ROC curve for two groups. The volumes under the NPI lower and upper ROC surfaces have been considered. Only real-valued data were considered, we aim to develop such ROC surfaces for ordinal data in the future.

A more established approach for nonparametric inference is bootstrapping, which can of course be applied for the inferential problems considered in this thesis. However, if the data contain categories with zero observations, which may well occur for smaller data sets, such categories will also never appear in the bootstrap samples, while in the NPI approach presented here the positive upper probabilities for such outcomes reflect that they are not deemed impossible on the basis of the data. At the same time, the NPI lower probabilities for such categories are zero, reflecting that the data do not provide strong evidence in favour of such observations to occur. Not surprisingly, in case of large data sets the NPI and bootstrap methods will give very similar conclusions when looking a single observation ahead for each group. However, if one is interested in multiple future observations, which is as mentioned above left as a topic for future research, then the NPI method treats these as mutually dependent while bootstrap methods typically assume these to be mutually independent, both given the data. An NPI alternative to bootstrapping is currently being studied with focus on real-valued data. It will be interesting to also

develop such NPI-bootstrapping methods for ordinal data, and apply it to problems of multiple comparisons and diagnostic accuracy.

The NPI approach has the obvious limitation that a substantial number of observations are required in order to get small imprecision, yet the fact that larger imprecision reflects limited availability of data is a natural reflection of the fact that the approach makes only few modelling assumptions. If there are only few observations and the NPI approach does not indicate a clear answer to an inferential problem due to large imprecision, this is a strong indication that other methods, which use stronger modelling assumptions, may be needed, or one should attempt to collect more data.

The inferences presented in this thesis used only a single future observation from each group. NPI enables inference on multiple future observations per group, developing this for ordinal data with similar application to multiple comparisons, is an interesting topic for future research. Once NPI for ordinal data has been developed for multiple future observations then further study of methods for diagnostic accuracy for multiple future individuals per group is an interesting topic for future research. Also for accuracy of diagnostic tests with real-valued data, with either 2 or 3 groups, consideration of multiple future observations provides a nice research challenge.

# Bibliography

- [1] Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley, Hoboken, New Jersey.
- [2] Agresti, A. and Brent, A.C. (2002). The analysis of contingency tables under inequality constraints. *Journal of Statistical Planning and Inference*, **107**, 45-73.
- [3] Arts, G.R J. Coolen, F. P.A. and van der Laan, P. (2004). Nonparametric predictive inference in statistical process control. *Quality Technology and Quantitative Management*, **1**, 201-216.
- [4] Augustin, T. and Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, **124**, 251-272.
- [5] Baker, R.M. (2010). *Multinomial nonparametric predictive inference: Selection, Classification and Subcategory Data*. Ph.D Thesis, Durham University, Durham, UK. Available from [www.npi-statistics.com](http://www.npi-statistics.com).
- [6] Baker, R.M. and Coolen, F.P.A. (2010). Nonparametric predictive category selection for multinomial data. *Journal of Statistical Theory and Practice*, **4**, 509-526.
- [7] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology*, **12**, 387-415.

- [8] Bechhofer, R.E., Santner, T.J. and Goldsman, D.M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. Wiley, New York.
- [9] Beck, A.C. (2005). *Receiver Operating Characteristic Surfaces: Inference and Applications*. Ph.D Thesis, University of Rochester, Rochester, New York.
- [10] Bernard, J.M. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, **39**, 123-150.
- [11] Boole, G. (1854). *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*. Walton & Maberley, London.
- [12] Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, 1145-1159.
- [13] Chateauneuf, A., Cohen, M. and Meilijson, I. (1997). New tools to better model behavior under risk and uncertainty: An overview, *Finance*, **18**, 25-46.
- [14] Chang, L., Lee, P.L., Yiannoutsos, C.T., Ernst, T., Marra, C.M., Richards, T., Kolson, D., Schifitto, G., Jarvik, J.G., Miller, E.N., Lenkinski, R., Gonzalez, G. and Navia, B.A. (2004). A multicenter in vivo proton-MRS study of HIV-associated dementia and its relationship to age. *NeuroImage*, **23**, 1336-1347.
- [15] Chen, C., Cohen, A. and Sackrowitz, H. (2009). Multiple testing for ordinal data. *Electronic Journal of Statistics*, **3**, 912-931.
- [16] Coolen, F.P.A. (1996). Comparing two populations based on low stochastic structure assumptions. *Statistics & Probability Letters*, **29**, 297-305.
- [17] Coolen, F.P.A. (1997). An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, **56**, 61-68.

- [18] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes problem. *Statistics & Probability Letters*, **36**, 349-357.
- [19] Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, **15**, 21-47.
- [20] Coolen, F.P.A. (2011). Nonparametric Predictive Inference. In: M. Lovric (ed.), *International Encyclopaedia of Statistical Sciences*. Springer, Berlin, 968-970.
- [21] Coolen, F.P.A. and Augustin, T. (2005). Learning from multinomial data: a nonparametric predictive alternative to the imprecise dirichlet model. In: Cozman, F.G., Nau, R., Seidenfeld, T. (eds.), ISIPTA'05. *Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications*, pp. 125-134.
- [22] 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**, 217-230.
- [23] Coolen, F.P.A. and Coolen-Schrijner, P. (2006). Nonparametric predictive subset selection for proportions. *Statistics & Probability Letters*, **76**, 1675-1684.
- [24] Coolen, F.P.A. and Coolen-Schrijner, P. (2007). Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference*, **137**, 23-33.
- [25] Coolen, F.P.A., Coolen-Schrijner, P., Coolen-Maturi, T. and Elkhafifi, F.F. Nonparametric predictive inference for ordinal data. *Communications in Statistics-Theory and Methods*, to appear.
- [26] Coolen, F.P.A., Coolen-Schrijner, P. and Maturi, T. (2010). On Nonparametric Predictive Inference for Ordinal Data. In: E. Hullermeier et al (eds) *Computational Intelligence for Knowledge-Based Systems Design, Proceedings 13th International Conference on Information Processing and Management of Uncertainty*, Dortmund (Germany). Springer, Berlin, 188-197.

- [27] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2012a). Nonparametric Predictive Inference for Diagnostic Accuracy. *Journal of Statistical Planning and Inference*, **142**, 1141-1150.
- [28] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2012b). Nonparametric Predictive Inference for binary diagnostic tests. *Journal of Statistical Theory and Practice*, to appear.
- [29] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2011) Nonparametric predictive selection with early termination of experiments. *Journal of Statistical Planning and Inference*, **141**, 1403-1421.
- [30] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. Nonparametric predictive multiple comparisons of lifetime data. *Communications in Statistics - Theory and Methods*, to appear.
- [31] Coolen, F.P.A, Troffaes, M.C. and Augustin, T. (2011). Imprecise probability. In: M. Lovric (ed.), *International Encyclopaedia of Statistical Sciences*. Springer, Berlin, 645-648.
- [32] Coolen, F.P.A. and van der Laan, P. (2001). Imprecise predictive selection bases on low structure assumptions. *Journal of Statistical Planning and Inference*, **98**, 259-277.
- [33] De Finetti, B. (1974). *Theory of Probability*. Wiley, London.
- [34] Dreiseitl, S., Ohno-Machado, L. and Binder, M. (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making*, **20**, 323-331.
- [35] Elkhafifi, F.F. and Coolen, F.P.A. Nonparametric predictive inference for accuracy of ordinal diagnostic tests. *Journal of Statistical Theory and Practice*, to appear.
- [36] Elkhafifi, F.F., Coolen-Maturi, T. and Coolen, F.P.A. (2011). Nonparametric predictive inference to assess diagnostic tests (Abstract). *In Proceedings Inter-*

- national conference of the ERCIM WG on Computing and Statistics*. London, United Kingdom.
- [37] Everitt, B.S. (1992). *The Analysis of Contingency Tables*. Chapman and Hall, London.
- [38] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861-874
- [39] Fluss, R., Faraggi, D. and Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. *Biometrical Journal*, **47**, 458-472.
- [40] Greiner, M., Pfeiffer, D. and Smith, R.D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, **45**, 23-41.
- [41] Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225-245.
- [42] Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, **63**, 677-691.
- [43] Hill, B.M. (1988). De Finetti's theorem, induction, and  $A_{(n)}$ , or Bayesian nonparametric predictive inference (with discussion). In *Bayesian Statistics 3*, (Editors) J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A. Smith, pp.211-241. Oxford University Press.
- [44] Krzanowski, W.J. and Hand, D.J. (2009). *ROC curves for continuous data*. CRC Press, Boca Raton.
- [45] Liu, I. and Agresti, A. (2005). The analysis of ordered categorical data: an overview and a survey of recent developments (with discussion). *TEST*, **14**, 1-73.
- [46] Lawless, J.F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika*, **92**, 529-542.

- [47] Inácio, V. Turkman, A.A., Nakas, C.T and Alonzo, T.A. (2011). Nonparametric Bayesian estimation of the the three-way receiver operating Characteristic surface, *Biometrical Journal* , **53**, 1011-1024
- [48] McNeil, B.J. and Hanley, J.A. (1984). Statistical approaches to the analysis of receiver operating characteristic (ROC) curve. *Medical Decision Making*, **4**, 137-150.
- [49] Metz, C.E. (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, **21**, 720-33.
- [50] Metz, C.E. (2008). ROC analysis in midical imaging: a tutorial review of the literature. *Radiological Physics and Technology*, **1**, 2-12.
- [51] Mossman, D. (1999). Three-way ROCs. *Medical Decision Making*, **19**, 78-89.
- [52] Nakas, C.T. and Yiannoutsos, C. T. (2006). Ordered Multiple Class Receiver Operating Characteristic (*ROC*) Analysis. *Encyclopedia of Biopharmaceutical Statistics*, Taylor and Francis, Oxford.
- [53] Nakas, C.T. and Yiannoutsos, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, **23**, 3437-3449.
- [54] Nakas, C.T., Alonzob, 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**, 2946-2955.
- [55] Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, Oxford.
- [56] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- [57] Schafer, H. (1989). Constructing a cut-off point for a quantitative diagnostic test. *Statistics in Medicine*, **8**, 1381-1391



- [58] Schisterman, E.F., Perkins, N.J., Liu, A. and Bondell, H. (2005). Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology*, **16**, 73-81.
- [59] Simonoff, J.S. (2003). *Analyzing Categorical Data*. Springer Verlag, New York.
- [60] Spackman, K.A. (1989). Signal detection theory: Valuable tools for evaluating inductive learning. *In: Proceedings Sixth International Workshop on Machine Learning*. Morgan Kaufman, San Mateo, CA, 160-163.
- [61] Spiegelhalter, D.J., Aylin, P., Best, N.G., Evans, S.J.W. and Murray, G.D. (2002). Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society. Series A*, **165**, 191-231.
- [62] Sukhatme, S. and Beam, C.A. (1994). Stratification in nonparametric ROC studies. *Journal of Biometrics*, **50**, 149-163.
- [63] Tian, L., Xiong, C., Lai, Y. and Vexler, A. (2011). Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *Journal of Statistical Planning and Inference*, **141**, 549-558.
- [64] Verheijen, J.H.M., Coolen, F. P. A. and van der Laan, P. (1997). Combining two classical approaches for statistical selection. *Communications in Statistics - Theory and Methods*, **26**, 1291-1312.
- [65] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.
- [66] Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles (with discussion). *Journal of the Royal Statistical Society Series B*, **58**, 3-57
- [67] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, **24**, 149-170.

- [68] Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung: I. Intervallwahrscheinlichkeit als umfassendes Konzept* (in German). Physika, Heidelberg.
- [69] Weinstein, S., Obuchowski, N.A. and Lieber, M.L (2005). Clinical Evaluation of Diagnostic Tests. *American Journal of Roentgenology* , **184**, 14-19.
- [70] Wians Jr, F.H., Urban, J.E., Keffer, J.H. and Kroft, S.H. (2001). Discriminating between iron deficiency anemia and anemia of chronic disease using traditional indices of iron status vs transferrin receptor concentration. *American Journal of Clinical Hematopathology*, **115**, 112-118.
- [71] Yiannoutsos, C.T., Nakas, C.T. and Navia, B.A. (2008). Assessing multiple-group diagnostic problems with multi-dimensional receiver operating characteristic surfaces: Application to proton MR Spectroscopy (MRS) in HIV-related neurological injury. *Neuroimage*, **40**, 248-255.
- [72] Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, **3**, 32-35.
- [73] Zhou, X.H. and Obuchowski, N.A. and McClish, D.K. (2002). *Statistical Methods in Diagnostic Medicine*. Wiley, New York.

# Appendix A

## Multiple comparisons

### A.1 Proof of Theorem 3.2

*Proof.* The NPI lower probability is derived as follows:

$$\begin{aligned}
P\left(X_{n^{j+1}}^j = \min_{l=1, \dots, J} X_{n^{l+1}}^l\right) &= P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{j+1}}^j \leq X_{n^{l+1}}^l\}\right) \\
&= \sum_{k=1}^K P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P(X_{n^{j+1}}^j \in C_k) \\
&\geq \sum_{k=1}^{K-1} P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k\right) \times P^R(X_{n^{j+1}}^j \in C_k) \\
&\quad + P\left(\bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_K\right) \times P^R(X_{n^{j+1}}^j \in C_K) \\
&= \sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in C_{k,K}) + \frac{n_K^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in C_K) \\
&\geq \sum_{k=1}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P^L(X_{n^{l+1}}^l \in C_{k,K}) + \frac{n_K^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J P^L(X_{n^{l+1}}^l \in C_K) \\
&= \frac{n_1^j}{n^j + 1} \prod_{l=1, l \neq j}^J P^L(X_{n^{l+1}}^l \in C_{1,K}) + \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left(\sum_{f=k}^K \frac{n_f^l}{n^l + 1}\right) \\
&\quad + \frac{n_K^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J \left(\frac{n_K^l}{n^l + 1}\right) \tag{A.1.1}
\end{aligned}$$

$$\begin{aligned}
&= \frac{n_1^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \sum_{f=2}^K \frac{n_f^l}{n^l + 1} + \frac{n_1^l + 1}{n^l + 1} \right) + \sum_{k=2}^{K-1} \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K \frac{n_f^l}{n^l + 1} \right) \\
&\quad + \frac{n_K^j + 1}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \frac{n_K^l}{n^l + 1} \right) \tag{A.1.2}
\end{aligned}$$

$$\begin{aligned}
&= A \left[ n_1^j \prod_{l=1, l \neq j}^J \left( \sum_{f=1}^K n_f^l + 1 \right) + \sum_{k=2}^{K-1} n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l \right) + (n_K^j + 1) \prod_{l=1, l \neq j}^J (n_K^l) \right] \\
\underline{P}_j^{\leq} &= A \left[ n_1^j \prod_{l=1, l \neq j}^J (n^l + 1) + \sum_{k=2}^K \left( n_k^j \prod_{l=1, l \neq j}^J n_{k,K}^l \right) + \prod_{l=1, l \neq j}^J n_K^l \right] \tag{A.1.3}
\end{aligned}$$

Equation (A.1.3) can be written as

$$\frac{n_1^j}{n^j + 1} + A \left[ \sum_{k=2}^K \left( n_k^j \prod_{l=1, l \neq j}^J n_{k,K}^l \right) + \prod_{l=1, l \neq j}^J n_K^l \right] \tag{A.1.4}$$

where (A.1.4) will be the result if we put probability  $P^L(X_{n^{l+1}}^l \in \mathcal{C}_{1,K})$  equal to one in (A.1.1). The corresponding NPI upper probability is derived as follows:

$$\begin{aligned}
P \left( X_{n^{j+1}}^j = \min_{l=1, \dots, J} X_{n^{l+1}}^l \right) &= P \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{j+1}}^j \leq X_{n^{l+1}}^l\} \right) \\
&= \sum_{k=1}^K P \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k \right) \times P(X_{n^{j+1}}^j \in C_k) \\
&\leq \sum_{k=1}^K P \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k \right) \times P^L(X_{n^{j+1}}^j \in C_k) \\
&= \sum_{k=2}^K \frac{n_k^j}{n^j + 1} \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_k \right) \\
&\quad + \left( \frac{n_1^j + 1}{n^j + 1} \right) \left( \bigcap_{l=1, l \neq j}^J \{X_{n^{l+1}}^l \geq X_{n^{j+1}}^j\} \mid X_{n^{j+1}}^j \in C_1 \right) \\
&= \sum_{k=2}^K \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in \mathcal{C}_{k,K}) + \left( \frac{n_1^j + 1}{n^j + 1} \right) \prod_{l=1, l \neq j}^J P(X_{n^{l+1}}^l \in \mathcal{C}_{1,K}) \\
&\leq \sum_{k=2}^K \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J P^R(X_{n^{l+1}}^l \in \mathcal{C}_{k,K}) + \left( \frac{n_1^j + 1}{n^j + 1} \right) \prod_{l=1, l \neq j}^J P^R(X_{n^{l+1}}^l \in \mathcal{C}_{1,K}) \tag{A.1.5}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=2}^K \frac{n_k^j}{n^j + 1} \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^{K-1} \frac{n_f^l}{n^l + 1} + \frac{n_K^1 + 1}{n_l + 1} \right) \\
&\quad + \left( \frac{n_1^j + 1}{n^j + 1} \right) \prod_{l=1, l \neq j}^J \left( \sum_{f=1}^{K-1} \frac{n_f^l}{n^l + 1} + \frac{n_K^1 + 1}{n_l + 1} \right) \\
&= A \left[ \sum_{k=2}^K n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l + 1 \right) + (n_1^j + 1) \prod_{l=1, l \neq j}^J \left( \sum_{f=1}^K n_f^l + 1 \right) \right] \\
&= A \left[ \sum_{k=1}^K n_k^j \prod_{l=1, l \neq j}^J \left( \sum_{f=k}^K n_f^l + 1 \right) + \prod_{l=1, l \neq j}^J \left( \sum_{f=1}^K n_f^l + 1 \right) \right] \\
\bar{P}_j^{\leq} &= A \left[ \sum_{k=1}^K \left( n_k^j \prod_{l=1, l \neq j}^J (n_{k,K}^l + 1) \right) + \prod_{l=1, l \neq j}^J (n^l + 1) \right] \tag{A.1.6}
\end{aligned}$$

Equation (A.1.6) can be written as

$$A \left[ \sum_{k=2}^K n_k^j \prod_{l=1, l \neq j}^J (n_{k,K}^l + 1) \right] + \left( \frac{n_1^j + 1}{n^j + 1} \right) \tag{A.1.7}$$

(A.1.7) will be the result if we put probability  $P^R (X_{n^l+1}^l \in \mathcal{C}_{1,K}) = 1$  in (A.1.5).  $\square$

## A.2 Proof of Theorem 3.4 in case of four groups

Suppose we have four groups and we are interested in selecting the subset which contains two groups 1 and 3, so  $S = \{1, 3\}$  and  $S^c = \{2, 4\}$ . We derive the NPI lower and upper probabilities for the event that the next observations of group 1 and of group 3 are both less than or equal to the next observation from group 2 and from group 4. The NPI lower and upper probabilities for the event  $\max_{j=1,3} X_{n^j+1}^j \leq \min_{l=2,4} X_{n^l+1}^l$  are derived similarly to the lower and upper probabilities in Equations (3.35) and (3.38) but with  $<$  everywhere replaced by  $\leq$ . The lower probability is

derived as follows:

$$\begin{aligned}
P\left(\max_{j=1,3} X_{n^{j+1}}^j \leq \min_{l=2,4} X_{n^{l+1}}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{\max_{j \in S} X_{n^{j+1}}^j \leq X_{n^{l+1}}^l\right\}\right) \\
&= P\left(\left\{X_{n^{2+1}}^2 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^{2+1}}^2 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^{j+1}}^j \in C_{k_j}\}\right) \\
&\quad \times P\left(X_{n^{1+1}}^1 \in C_{k_1}, X_{n^{3+1}}^3 \in C_{k_3}\right) \\
&\geq \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^{2+1}}^2 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 \geq \max_{j=1,3} X_{n^{j+1}}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^{j+1}}^j \in C_{k_j}\}\right) \\
&\quad \times P^R\left(X_{n^{1+1}}^1 \in C_{k_1}, X_{n^{3+1}}^3 \in C_{k_3}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P\left(X_{n^{2+1}}^2 \in \mathcal{C}_{M_W, K}\right) P\left(X_{n^{4+1}}^4 \in \mathcal{C}_{M_W, K}\right) \\
&\geq \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L\left(X_{n^{2+1}}^2 \in \mathcal{C}_{M_W, K}\right) P^L\left(X_{n^{4+1}}^4 \in \mathcal{C}_{M_W, K}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^{k_1} \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L\left(X_{n^{2+1}}^2 \in \mathcal{C}_{k_1, K}\right) P^L\left(X_{n^{4+1}}^4 \in \mathcal{C}_{k_1, K}\right) \\
&\quad + \sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L\left(X_{n^{2+1}}^2 \in \mathcal{C}_{k_3, K}\right) P^L\left(X_{n^{4+1}}^4 \in \mathcal{C}_{k_3, K}\right) \\
&= \left[\sum_{k_1=1}^K \sum_{k_3=1}^{k_1} \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_1}^K \frac{L_f^l}{n^l + 1}\right) \\
&\quad + \left[\sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_3}^K \frac{L_f^l}{n^l + 1}\right) \\
\underline{P}_{\{1,3\}}^{\leq} &= A \left[\sum_{k_1=1}^K \sum_{k_3=1}^{k_1} \prod_{j=1,3} \left(R_{k_j}^j\right) \prod_{l=2,4} \left(\sum_{f=k_1}^K L_f^l\right)\right. \\
&\quad \left. + \sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\sum_{f=k_3}^K R_{k_j}^j\right) \prod_{l=2,4} \left(\sum_{f=k_3}^K L_f^l\right)\right] \tag{A.2.8}
\end{aligned}$$

where  $M_W = \max\{k_1, k_3\}$  and  $A = \left[\prod_{j=1}^4 (n^j + 1)\right]^{-1}$ . The corresponding upper probability is derived as follows:

$$\begin{aligned}
P\left(\max_{j=1,3} X_{n^{j+1}}^j \leq \min_{l=2,4} X_{n^l+1}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{\max_{j \in S} X_{n^j+1}^j \leq X_{n^l+1}^l\right\}\right) \\
&= P\left(\left\{X_{n^2+1}^2 \geq \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 \geq \max_{j=1,3} X_{n^j+1}^j\right\}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^2+1}^2 \geq \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 \geq \max_{j=1,3} X_{n^j+1}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\}\right) \\
&\quad \times P(X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}) \\
&\leq \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^2+1}^2 \geq \max_{j=1,3} X_{n^j+1}^j\right\} \cap \left\{X_{n^4+1}^4 \geq \max_{j=1,3} X_{n^j+1}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^j+1}^j \in C_{k_j}\}\right) \\
&\quad \times P^L(X_{n^1+1}^1 \in C_{k_1}, X_{n^3+1}^3 \in C_{k_3}) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right) P(X_{n^2+1}^2 \in \mathcal{C}_{M_W, K}) P(X_{n^4+1}^4 \in \mathcal{C}_{M_W, K}) \\
&\leq \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right) P^R(X_{n^2+1}^2 \in \mathcal{C}_{M_W, K}) P^R(X_{n^4+1}^4 \in \mathcal{C}_{M_W, K}) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^k \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right) P^R(X_{n^2+1}^2 \in \mathcal{C}_{k_1, K}) P^R(X_{n^4+1}^4 \in \mathcal{C}_{k_1, K}) \\
&\quad + \sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right) P^R(X_{n^2+1}^2 \in \mathcal{C}_{k_3, K}) P^R(X_{n^4+1}^4 \in \mathcal{C}_{k_3, K}) \\
&= \left[\sum_{k_1=1}^K \sum_{k_3=1}^k \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_1}^K \frac{R_f^l}{n^l + 1}\right) \\
&\quad + \left[\sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\frac{L_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_3}^K \frac{R_f^l}{n^l + 1}\right) \\
\bar{P}_{\{1,3\}}^{\leq} &= A \left[\sum_{k_1=1}^K \sum_{k_3=1}^{k_1} \prod_{j=1,3} (L_{k_j}^j) \prod_{l=2,4} \left(\sum_{f=k_1}^K R_f^l\right)\right. \\
&\quad \left.+ \sum_{k_1=1}^K \sum_{k_3=k_1+1}^K \prod_{j=1,3} \left(\sum_{f=k_3}^K L_{k_j}^j\right) \prod_{l=2,4} \left(\sum_{f=k_3}^K R_f^l\right)\right] \tag{A.2.9}
\end{aligned}$$

where  $M_W = \max\{k_1, k_3\}$  and  $A = \left[\prod_{j=1}^4 (n^j + 1)\right]^{-1}$ .

### A.3 Proof of Theorem 3.5 in case of four groups

We derive the NPI lower and upper probabilities for the event  $\min_{j=1,3} X_{n^{j+1}}^j < \min_{l=2,4} X_{n^{l+1}}^l$ .

The lower probability is derived as follows:

$$\begin{aligned}
P\left(\min_{j=1,3} X_{n^{j+1}}^j < \min_{l=2,4} X_{n^{l+1}}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{\min_{j \in S} X_{n^{j+1}}^j < X_{n^{l+1}}^l\right\}\right) \\
&= P\left(\left\{X_{n^{2+1}}^2 > \min_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 > \min_{j=1,3} X_{n^{j+1}}^j\right\}\right) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^{2+1}}^2 > \min_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 > \min_{j=1,3} X_{n^{j+1}}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^{j+1}}^j \in C_{k_j}\}\right) \\
&\quad \times P(X_{n^{1+1}}^1 \in C_{k_1}, X_{n^{3+1}}^3 \in C_{k_3}) \\
&\geq \sum_{k_1=1}^K \sum_{k_3=1}^K P\left(\left\{X_{n^{2+1}}^2 > \min_{j=1,3} X_{n^{j+1}}^j\right\} \cap \left\{X_{n^{4+1}}^4 > \min_{j=1,3} X_{n^{j+1}}^j\right\} \mid \bigcap_{j=1,3} \{X_{n^{j+1}}^j \in C_{k_j}\}\right) \\
&\quad \times P^R(X_{n^{1+1}}^1 \in C_{k_1}, X_{n^{3+1}}^3 \in C_{k_3}) \\
&= \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P(X_{n^{2+1}}^2 \in \mathcal{C}_{M_w+1,K}) P(X_{n^{4+1}}^4 \in \mathcal{C}_{M_w+1,K}) \\
&\geq \sum_{k_1=1}^K \sum_{k_3=1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L(X_{n^{2+1}}^2 \in \mathcal{C}_{M_w+1,K}) P^L(X_{n^{4+1}}^4 \in \mathcal{C}_{M_w+1,K}) \\
&= \sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L(X_{n^{2+1}}^2 \in \mathcal{C}_{k_1+1,K}) P^L(X_{n^{4+1}}^4 \in \mathcal{C}_{k_1+1,K}) \\
&\quad + \sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right) P^L(X_{n^{2+1}}^2 \in \mathcal{C}_{k_3+1,K}) P^L(X_{n^{4+1}}^4 \in \mathcal{C}_{k_3+1,K}) \\
&= \left[\sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_1+1}^K \frac{L_f^l}{n^l + 1}\right) \\
&\quad + \left[\sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} \left(\frac{R_{k_j}^j}{n^j + 1}\right)\right] \prod_{l=2,4} \left(\sum_{f=k_3+1}^K \frac{L_f^l}{n^l + 1}\right) \\
P_{\{1,3\}}^< &= A \left[\sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} (R_{k_j}^j) \left[\prod_{l=2,4} \left(\sum_{f=k_1+1}^K L_f^l\right)\right]\right] \\
&\quad + \sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} (R_{k_j}^j) \prod_{l=2,4} \left(\sum_{f=k_3+1}^K L_f^l\right) \tag{A.3.10}
\end{aligned}$$

where  $M_w = \min\{k_1, k_3\}$  and  $A = \left[\prod_{j=1}^4 (n^j + 1)\right]^{-1}$ . The corresponding NPI lower probability for the event that  $S$  contains a group for which the next obser-



vation is less than or equal to the next observations from all not selected groups

$\min_{j=1,3} X_{n^{j+1}}^j \leq \min_{l=2,4} X_{n^{l+1}}^l$  is

$$\begin{aligned} \underline{P}_{\{1,3\}}^{\leq} = A & \left[ \sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} \left( R_{k_j}^j \right) \left[ \prod_{l=2,4} \left( \sum_{f=k_1}^K L_f^l \right) \right] \right. \\ & \left. + \sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} \left( R_{k_j}^j \right) \prod_{l=2,4} \left( \sum_{f=k_3}^K L_f^l \right) \right] \end{aligned} \quad (\text{A.3.11})$$

This lower probability is similarly derived and justified as in the lower probability (A.3.10). Secondly, the upper probability is again similarly derived by replacing every  $L$  by  $R$  in the proof of the lower probability (A.3.10), so it is equal to

$$\begin{aligned} \overline{P}_{\{1,3\}}^{\leq} = A & \left[ \sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} \left( L_{k_j}^j \right) \left[ \prod_{l=2,4} \left( \sum_{f=k_1+1}^K R_f^l \right) \right] \right. \\ & \left. + \sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} \left( L_{k_j}^j \right) \prod_{l=2,4} \left( \sum_{f=k_3+1}^K R_f^l \right) \right] \end{aligned} \quad (\text{A.3.12})$$

The corresponding NPI-based upper probability for the event  $\min_{j=1,3} X_{n^{j+1}}^j \leq \min_{l=2,4} X_{n^{l+1}}^l$  is again similarly derived and it is equal to

$$\begin{aligned} \overline{P}_{\{1,3\}}^{\leq} = A & \left[ \sum_{k_1=1}^K \sum_{k_3=k_1}^K \prod_{j=1,3} \left( L_{k_j}^j \right) \left[ \prod_{l=2,4} \left( \sum_{f=k_1}^K R_f^l \right) \right] \right. \\ & \left. + \sum_{k_3=1}^K \sum_{k_1=k_3+1}^K \prod_{j=1,3} \left( L_{k_j}^j \right) \prod_{l=2,4} \left( \sum_{f=k_3}^K R_f^l \right) \right] \end{aligned} \quad (\text{A.3.13})$$

## A.4 Maximum next observation

The lower probability (3.49) can be derive as follows:

$$\begin{aligned}
P\left(\min_{j \in S} X_{n^j+1}^j > \max_{l \in S^c} X_{n^l+1}^l\right) &= P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l < \min_{j \in S} X_{n^j+1}^j\right\}\right) \\
&= \sum_{k_{j_1}=1}^K \cdots \sum_{k_{j_w}=1}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l < \min_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \right. \\
&\quad \left. \prod_{j \in S} P(X_{n^j+1}^j \in C_{k_j}) \right] \tag{A.4.14}
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \left[ P\left(\bigcap_{l \in S^c} \left\{X_{n^l+1}^l < \min_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \right. \\
&\quad \left. \prod_{j \in S} P^L(X_{n^j+1}^j \in C_{k_j}) \right] \tag{A.4.15}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j+1} \prod_{l \in S^c} P\left(\left\{X_{n^l+1}^l < \min_{j \in S} X_{n^j+1}^j\right\} \mid X_{n^j+1}^j \in C_{k_j}, j \in S\right) \\
&= \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j+1} \prod_{l \in S^c} P(X_{n^l+1}^l \in \mathcal{C}_{1, M_w-1}) \\
&\geq \sum_{k_{j_1}=2}^K \cdots \sum_{k_{j_w}=2}^K \prod_{j \in S} \frac{L_{k_j}^j}{n^j+1} \prod_{l \in S^c} P^R(X_{n^l+1}^l \in \mathcal{C}_{1, M_w-1}) \tag{A.4.16}
\end{aligned}$$

where  $M_w = \min\{k_{j_1}, \dots, k_{j_w}\}$ . The corresponding NPI upper probability is derived similarly by putting the probability masses per interval at the endpoints, for the groups  $j \in S$  at the right endpoints and for all other groups  $l \in S^c$  at the left endpoints in the latent variable representation. These lower and upper probabilities are similarly derived and justified as Equations (3.32) and (3.33). In these formulas the summations start from the second category as the probability of the event  $\min_{j \in S} X_{n^j+1}^j > \max_{l \in S^c} X_{n^l+1}^l$  is zero when one or more of  $X_{n^j+1}^j, j \in S$  are in the first category.