Nonparametric predictive inference for ordinal data

F.P.A. Coolen[a,*], P. Coolen-Schrijner[†], T. Coolen-Maturi[b,‡], F.F. Ali[a,§]

[a]Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

[b]Kent Business School, University of Kent, Canterbury CT2 7PE, UK.

Abstract

Nonparametric predictive inference (NPI) is a powerful frequentist statistical framework based only on an exchangeability assumption for future and past observations, made possible by the use of lower and upper probabilities. In this paper, NPI is presented for ordinal data, which are categorical data with an ordering of the categories. 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, and briefly compared to NPI for non-ordered categorical data. As application the comparison of multiple groups of ordinal data is presented.

Keywords: Categorical data, lower and upper probabilities, multiple comparisons, nonparametric predictive inference, ordinal data.

1 Introduction

Nonparametric Predictive Inference (NPI) is a frequentist statistical framework based only on few modelling assumptions, enabled by the use of lower and upper probabilities to quantify uncertainty [4, 10]. In NPI, attention is restricted to one or more future observable random quantities, and Hill's assumption $A_{(n)}$ [19] is used to link these random quantities to data, in

[*]Email: frank.coolen@durham.ac.uk (Corresponding author)

[†]Pauline died in 2008: she had initiated this research and achieved the first results reported in this paper.

[‡]Email: T.Coolen-Maturi@kent.ac.uk

[§]Email: f.f.ali@durham.ac.uk

a way that is closely related to exchangeability [18]. Coolen and Augustin [11, 12] presented NPI for categorical data with no known relationship between the categories, as an alternative to the popular Imprecise Dirichlet Model (IDM) [6, 24]. In many applications the categories are ordered, e.g. different levels of severity of a disease, in which case such data are also known as ordinal data. The IDM can be applied to ordinal data [8] but it does not explicitly use the ordering of the categories. It is important that such knowledge about ordering of categories is taken into account, this paper 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 $A_{(n)}$. Excellent overviews of established statistical methods for ordinal data are presented in [2, 21].

Section 2 of this paper provides a brief introduction to NPI. Section 3 presents NPI for ordinal data. For events which are of most practical interest, closed-form formulae for the NPI lower and upper probabilities are derived, and some properties of these inferences are discussed. In Section 4 these inferences are briefly compared to NPI in case of non-ordered categories. To illustrate the application of this new method, comparison of multiple groups of ordinal data is presented in Section 5. This section includes extensive examples to illustrate these inferences. The paper ends with concluding remarks in Section 6.

2 Nonparametric predictive inference

Nonparametric predictive inference (NPI) [4, 10] is based on the assumption $A_{(n)}$ proposed by Hill [19]. Let $X_1, \ldots, X_n, X_{n+1}$ be real-valued absolutely continuous and exchangeable random quantities. Let the ordered observed values of $X_1, X_2, \ldots, X_n$ be denoted by $x_1 < x_2 < \ldots < x_n$ and let $x_0 = -\infty$ and $x_{n+1} = \infty$ for ease of notation. We assume that no ties occur; ties can be dealt with in NPI [10] but it is not relevant in this paper. For X_{n+1}, representing a future observation, $A_{(n)}$ partially specifies a probability distribution by $P(X_{n+1} \in I_j = (x_{j-1}, x_j)) = \frac{1}{n+1}$ for $j = 1, \ldots, n+1$. $A_{(n)}$ does not assume anything else, and can be considered to be a post-data assumption related to exchangeability [18]. 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' [18], which are lower and upper probabilities in interval probability theory [23, 25, 26].

In NPI, uncertainty about the future observation X_{n+1} is quantified by lower and upper probabilities for events of interest. Lower and upper probabilities generalize classical ('precise') probabilities, and a lower (upper) probability for event A, denoted by $\underline{P}(A)$ ($\overline{P}(A)$), can be interpreted as supremum buying (infimum selling) price for a gamble on the event A [23], or just as the maximum lower (minimum upper) bound for the probability of A that follows from the assumptions made [10]. This latter interpretation is used 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, $\underline{P}(A)$ ($\overline{P}(A)$) can be considered to reflect the evidence in favour of (against) event A.

Augustin and Coolen [4] proved that NPI has strong consistency properties in the theory of interval probability [23, 25, 26], it is also exactly calibrated from frequentist statistics perspective [20]. Direct application of $A_{(n)}$ for inferential problems is only possible for real-valued random quantities. However, by assuming latent variable representations and variations to $A_{(n)}$, NPI has been developed for different situations, including Bernoulli quantities [9]. Defining an assumption related to $A_{(n)}$, but on a circle instead of the real-line, Coolen [10] enabled inference for circular data. This 'circular-$A_{(n)}$' assumption, in combination with a latent variable representation using a probability wheel, enabled NPI for non-ordered categorical data as presented by Coolen and Augustin [12], with as additional attractive feature the possibility to include both defined and undefined new categories in the event of interest [11]. 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 possible 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 [9], and dependence of specific inferences on the choice of the number of future observations was explicitly studied in the context of multiple comparisons [14]. In this paper, attention is restricted to a single future observation, leaving generalization to multiple future observations as an interesting challenge for future research.

3 NPI for ordinal data

In situations with ordinal data, there are $K \geq 2$ categories to which observations belong, and these categories have a natural fixed ordering, hence they can be denoted by $C_1 < C_2 < \ldots < C_K$. It is attractive to base NPI for such data on the naturally related latent variable representation with the real-line partitioned into K categories, with the same ordering, and observations per category represented by corresponding values on the real-line and in the specific category. Assuming that multiple observations in a category are represented by different values in this latent variable representation, the assumption $A_{(n)}$ can be applied for the latent variables. This is now explained in detail, and for two important situations closed-forms for the NPI lower and upper probabilities are derived. We focus mostly on situations with $K \geq 3$, although the arguments also hold for $K = 2$, in which case the NPI method presented in this paper is identical to NPI for Bernoulli data [9].

We assume that n observations are available, with the number of observations in each category given. Let $n_k \geq 0$ be the number of observations in category C_k, for $k = 1, \ldots, K$, so $\sum_{k=1}^{K} n_k = n$. Let Y_{n+1} denote the random quantity representing a future observation. We wish to derive the NPI lower and upper probabilities for events $Y_{n+1} \in \bigcup_{k \in L} C_k$ with $L \subset \{1, \ldots, K\}$. These do not follow straightforwardly from the NPI lower and upper probabilities for the events involving single categories as lower (upper) probabilities are super-additive (sub-additive) [23].

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

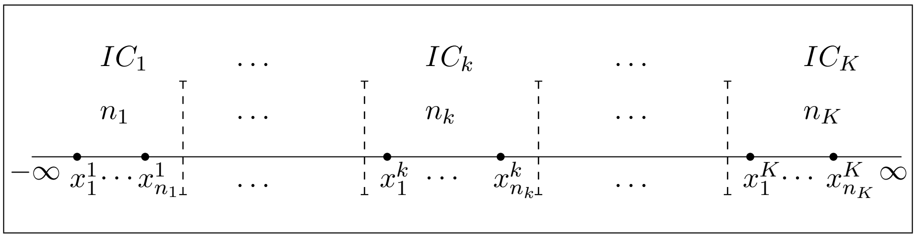

Figure 1: Ordinal data structure

We derive the NPI lower and upper probabilities for general events of the form $Y_{n+1} \in \mathcal{C}_L$, with $\mathcal{C}_L = \bigcup_{k \in L} C_k$ and $L \subset \{1, \ldots, K\}$. We assume that L is a strict subset of $\{1, \ldots, K\}$, 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 1. Assuming $A_{(n)}$ for X_{n+1} in the latent variable representation, each interval I_j has been assigned probability mass $1/(n+1)$ (see Section 2). Although we do not know exactly the values x_j, since they only exist in the latent variable representation, we do know the number of these x_j values in each interval IC_k. It should be emphasized that the intervals I_j are, as before, intervals between consecutive latent points x_{j-1} and x_j, and with the number of such points in each interval IC_k known we therefore also know how many intervals I_j are fully within each IC_k.

To derive the NPI lower probability for the event $Y_{n+1} \in \mathcal{C}_L$, we derive the NPI lower probability for the corresponding latent variable event $X_{n+1} \in \mathcal{IC}_L$, where $\mathcal{IC}_L = \bigcup_{k \in L} IC_k$ and $L \subset \{1, \ldots, K\}$. This lower probability is derived by summing all probability masses assigned to intervals I_j that are fully within $\mathcal{IC}_L$, so in effect we minimise the total probability mass assigned to $\mathcal{IC}_L$. Hence, these NPI lower probabilities are

$$\underline{P}(Y_{n+1} \in \mathcal{C}_L) = \underline{P}(X_{n+1} \in \mathcal{IC}_L) = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbf{1}\{I_j \subset \mathcal{IC}_L\} \tag{1}$$

where $\mathbf{1}\{A\}$ is equal to 1 if A is true and equal to 0 else. As we do not know the exact locations of the intervals IC_k, this may appear to be vague, yet the fact that we know the numbers of x_j

values within each interval IC_k 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 $\mathcal{IC}_L$. Without any further assumptions on the way the probability mass $1/(n+1)$ is spread over an interval I_j, this means that we can include all such probability masses corresponding to intervals I_j that have a non-empty intersection with $\mathcal{IC}_L$. So the NPI upper probabilities are

$$\overline{P}(Y_{n+1} \in \mathcal{C}_L) = \overline{P}(X_{n+1} \in \mathcal{IC}_L) = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbf{1}\{I_j \cap \mathcal{IC}_L \neq \emptyset\} \tag{2}$$

These NPI upper probabilities are also uniquely determined. The construction of these NPI lower and upper probabilities can be presented following Shafer's concept of basic probability assignments [22], but it should be emphasized that the NPI approach does not follow the Dempster-Shafer rule for updating which is often associated with the use of basic probability assignments. From the perspective of frequentist statistics, the NPI lower and upper probabilities (1) and (2) can be considered as 'confidence statements', in the sense that repeated application of this procedure will lead to correct predictions of the event $Y_{n+1} \in \mathcal{C}_L$ in a proportion that, in the limit, will be in the interval $[\underline{P}(Y_{n+1} \in \mathcal{C}_L), \overline{P}(Y_{n+1} \in \mathcal{C}_L)]$. Of course, to achieve this result the method should only be applied in cases where the assumption $A_{(n)}$ is reasonable, so not for example if there are clear patterns in the data set. Next, we present closed-form results for these NPI lower and upper probabilities for two special cases which are of practical interest. Thereafter we briefly discuss some properties of these NPI lower and upper probabilities, and we present an example to illustrate them.

An important special case of these inferences concerns the event $Y_{n+1} \in \mathcal{C}_L$, with $\mathcal{C}_L$ consisting of adjoining categories, so the corresponding union of intervals $\mathcal{IC}_L$ 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 $L = \{s, \ldots, t\}$, with $s, t \in \{1, \ldots, K\}$, $s \leq t$, excluding the case with $s = 1$ and $t = K$ for which both the NPI lower and upper probabilities are equal to 1. Let $\mathcal{C}_{s,t} = \bigcup_{k=s}^{t} C_k$, $\mathcal{IC}_{s,t} = \bigcup_{k=s}^{t} IC_k$ and let $n_{s,t} = \sum_{k=s}^{t} n_k$. Using the notation $(x)^+ = \max(x, 0)$, the NPI lower and upper probabilities (1) and (2) for such events are

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

$$\overline{P}(Y_{n+1} \in C_{s,t}) = \overline{P}(X_{n+1} \in \mathcal{IC}_{s,t}) = \frac{n_{s,t}+1}{n+1} \quad \text{for } 1 \leq s \leq t \leq K \tag{4}$$

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

A further special case for which closed-form expressions are available for the NPI lower and upper probabilities occurs if $n_k > 0$ for all $k \in \{1, \ldots, K\}$, so there are observations in all K categories. We need to consider if the categories C_1 and C_K are included in $\mathcal{C}_L$ (so IC_1 and IC_K in $\mathcal{IC}_L$) and we need to take account of all pairs of neighbouring categories which are both included in $\mathcal{C}_L$. Let $p_L = \sum_{r=1}^{K-1} \mathbf{1}\{r, r+1 \in L\}$ be the number of neighbouring pairs of categories included in $\mathcal{C}_L$, and let $e_L = \mathbf{1}\{1 \in L\} + \mathbf{1}\{K \in L\} + p_L$. We further introduce the notation s_L for the number of categories in $\mathcal{C}_L$, so $s_L = |L|$, and $n_L = \sum_{k \in L} n_k$. Then the NPI lower probability (1), with L a strict subset of $\{1, \ldots, K\}$, is

$$\underline{P}\left(Y_{n+1} \in \mathcal{C}_L\right) = \underline{P}\left(X_{n+1} \in \mathcal{IC}_L\right) = \frac{\sum_{k \in L}(n_k - 1) + e_L}{n+1} = \frac{n_L - s_L + e_L}{n+1} \tag{5}$$

and the corresponding NPI upper probability (2) is

$$\overline{P}\left(Y_{n+1} \in \mathcal{C}_L\right) = \overline{P}\left(X_{n+1} \in \mathcal{IC}_L\right) = \frac{\sum_{k \in L}(n_k + 1) - p_L}{n+1} = \frac{n_L + s_L - p_L}{n+1} \tag{6}$$

These two special cases are likely to cover many situations of practical interest. The problem in deriving a simple general closed-form expression for the NPI lower and upper probabilities (1) and (2) results from accounting for one or more consecutive categories without any observations in the event of interest, in which case it is important whether or not there are observations in the neighbouring categories.

The NPI lower and upper probabilities (1) and (2) satisfy the conjugacy property $\underline{P}(Y_{n+1} \in \mathcal{C}_L) = 1 - \overline{P}(Y_{n+1} \in \mathcal{C}_{L^c})$ for all $L \subset \{1, \ldots, K\}$ and $L^c = \{1, \ldots, K\} \backslash L$, which follows from

$\mathbf{1}\{I_j \subset \mathcal{IC}_L\} + \mathbf{1}\{I_j \cap \mathcal{IC}_L^c \neq \emptyset\} = 1$ for all $j = 1, \ldots, n+1$. Augustin and Coolen [4] prove stronger consistency properties for NPI lower and upper probabilities for real-valued random quantities within the theory of Weichselberger [25, 26], in particular that they are F-probability. Their results apply directly to the NPI lower and upper probabilities for X_{n+1} in the latent variable representation in this paper, and hence also imply that the NPI lower and upper probabilities (1) and (2) for the categorical random quantity Y_{n+1} are F-probability. This implies the above mentioned conjugacy property, and also coherence of these lower and upper probabilities in the sense of Walley [23]. However, Walley-coherence goes further by also considering such lower and upper probabilities at different moments in time, that is to say with different numbers of observations as is relevant in case of updating. In NPI, updating is performed by just calculating the relevant lower and upper probabilities using all available data, and is not performed via conditioning on prior sets of probabilities [4]. The NPI lower and upper probabilities (1) and (2) bound the corresponding empirical probability for the event of interest, so

$$\underline{P}(Y_{n+1} \in \mathcal{C}_L) \leq \frac{n_L}{n} \leq \overline{P}(Y_{n+1} \in \mathcal{C}_L) \tag{7}$$

Property (7) 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 [10].

Example 1

Suppose there are $K = 5$ ordered categories, $C_1 < \ldots < C_5$, and $n = 11$ observations with $n_1 = 1$, $n_2 = 3$, $n_3 = 1$, $n_4 = 4$ and $n_5 = 2$, so equations (5) and (6) can be used. The NPI lower and upper probabilities for several events $Y_{12} \in \mathcal{C}_L$ are given in Table 1, together with the corresponding empirical probability n_L/n. These lower and upper probabilities illustrate the relation (7), and they also show that the difference between corresponding upper and lower probabilities is not constant. The lower and upper probabilities for the events with L consisting of a single category or a group of adjoining categories also illustrate the lower and upper probabilities (3) and (4) from the first special case discussed above.

L	$\underline{P}$	$\overline{P}$	n_L/n
$\{1\}$	1/12	2/12	1/11
$\{2\}$	2/12	4/12	3/11
$\{3\}$	0	2/12	1/11
$\{4\}$	3/12	5/12	4/11
$\{5\}$	2/12	3/12	2/11
$\{1,2\}$	4/12	5/12	4/11
$\{1,2,3\}$	5/12	6/12	5/11
$\{2,3,4\}$	7/12	9/12	8/11
$\{1,2,4\}$	7/12	10/12	8/11
$\{1,2,4,5\}$	10/12	1	10/11

Table 1: NPI lower and upper probabilities

4 Comparison to NPI for non-ordered categorical data

Coolen and Augustin [12] presented NPI for categorical data with a known number of possible categories yet with no ordering or other known relationship between the categories. Their inferences are based on a latent variable representation using a probability wheel, with each category represented by a single segment of the wheel yet without any assumption about the specific configuration of the wheel. Their NPI lower and upper probabilities with regard to the next observation are further based on a circular version of $A_{(n)}$ [10] and optimisation over all possible configurations of the probability wheel that are possible corresponding to the data and this so-called circular-$A_{(n)}$ assumption. Coolen and Augustin [11] illustrated how this model can also be used in case of an unknown number of possible categories, which is less likely to be of relevance in case of ordinal data hence we have not addressed it here. 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 [12], we just wish to emphasize that the inferences can differ substantially if categories are known to be ordered and therefore the inferences presented here are applied.

To illustrate that NPI for non-ordered categorical data and NPI for ordinal data can be very different, consider the following simple example. Suppose we have $K = 6$ ordered categories, $C_1 < \ldots < C_6$, and only $n = 3$ observations, one in each of the first three categories, so

$n_1 = n_2 = n_3 = 1$ and $n_4 = n_5 = n_6 = 0$. Following the results presented in this paper, the NPI lower and upper probabilities for the event $Y_4 \in \{C_1, C_2, C_3\}$ are 3/4 and 1, respectively. If, however, the categories were not assumed to be ordered, then the corresponding NPI lower and upper probabilities for this event would be 0 and 1, respectively [12]. The latter lower probability may be surprising, it results from the possibility that the categories C_1, C_2, C_3 could, in the probability wheel representation, be separated by the other three categories, and from the fact that no single category has been observed more than once. We do not discuss this difference in more detail, but it is important to recognize that the inferences for categorical data can differ substantially if one can use a known ordering of the categories. Due to the different latent variable representations for these two situations, it is not the case that the NPI lower and upper probabilities according to these two models are nested, as could perhaps have been expected. One could consider different structures for the categories and different latent variable representations, this is left as an interesting topic for future research.

5 Multiple comparisons

In this section, we introduce NPI for multiple comparisons of ordinal data. Due to the explicitly predictive nature of NPI, we consider events of the following form. We divide the 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. We also 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. Of course, these results include the important special case of pairwise comparison, in which case S and S^c both contain just a single group, and the case with S consisting of just one group but S^c of multiple groups which is relevant if one is just interested in the group that provides the smallest next observation. Similar NPI methods for multiple comparisons, with some important variations, have been presented for real-valued data [7, 15], for proportions data [13, 14] and for lifetime data including right-censored observations

[16, 17].

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 ($j = 1, \ldots, J$) of which n^j_k are in category C_k, $k = 1, \ldots, K$. So $n = \sum_{j=1}^{J} n^j = \sum_{j=1}^{J} \sum_{k=1}^{K} n^j_k$, and let $n^j_{s,t} = \sum_{k=s}^{t} n^j_k$ where $s < t$. 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. Let $Y^j_{n^j+1}$ denote the next observation from group j and let the corresponding latent variable be denoted by $X^j_{n^j+1}$. We further introduce the notation

$$P^L(Y^j_{n^j+1} \in \mathcal{C}_{s,t}) = P^L(X^j_{n^j+1} \in \mathcal{IC}_{s,t}) = \begin{cases} \dfrac{n^j_{s,t}}{n^j+1} & \text{if} \quad 1 < s \leq t \leq K \\ \dfrac{n^j_{s,t}+1}{n^j+1} & \text{if} \quad s = 1 \end{cases} \tag{8}$$

and

$$P^R(Y^j_{n^j+1} \in \mathcal{C}_{s,t}) = P^R(X^j_{n^j+1} \in \mathcal{IC}_{s,t}) = \begin{cases} \dfrac{n^j_{s,t}}{n^j+1} & \text{if} \quad 1 \leq s \leq t < K \\ \dfrac{n^j_{s,t}+1}{n^j+1} & \text{if} \quad t = K \end{cases} \tag{9}$$

where if $t = s$ $\mathcal{C}_{s,s} = C_s$ and $n^j_{s,s} = n^j_s$. Probability (8) 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 end-point per interval. Similarly, probability (9) corresponds to the situation with all these probability masses put at the right end-point per interval.

We consider the event that for a selected subset of groups, the next observation from each group in the selected subset is less than the next observation from each group outside this selected subset. Let $S = \{j_1, \ldots, j_w\} \subset \{1, \ldots, J\}$ be the selected subset containing w groups, for $1 \leqslant w \leqslant J-1$, and let $S^c = \{1, \ldots, J\} \setminus S$ 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 for the next observation from each group in S is less than the next observation from each group in S^c, denoted by $\underline{P}^<_S = \underline{P}\left(\max_{j \in S} Y^j_{n^j+1} < \min_{l \in S^c} Y^l_{n^l+1}\right)$ and

$\overline{P}_S^{<} = \overline{P}\left(\max_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right)$, respectively, are

$$\underline{P}_S^{<} = \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[\prod_{j\in S} P^R\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^L\left(Y^l_{n^l+1} \in \mathcal{C}_{M_x+1,K}\right)\right] \tag{10}$$

$$\overline{P}_S^{<} = \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[\prod_{j\in S} P^L\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^R\left(Y^l_{n^l+1} \in \mathcal{C}_{M_x+1,K}\right)\right] \tag{11}$$

where $M_x = \max\{k_{j_1}, \ldots, k_{j_w}\}$ and P^L and P^R are as given by (8) and (9).

The NPI lower probability (10) is derived as follows

$$P\left(\max_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right) = P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1}\right\}\right)$$
$$= \sum_{k_{j_1}=1}^{K} \cdots \sum_{k_{j_w}=1}^{K} \left[P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1} \,\middle|\, Y^j_{n^j+1} \in C_{k_j}, j\in S\right\}\right) \prod_{j\in S} P\left(Y^j_{n^j+1} \in C_{k_j}\right)\right]$$
$$\geq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1} \,\middle|\, Y^j_{n^j+1} \in C_{k_j}, j\in S\right\}\right) \prod_{j\in S} P^R\left(Y^j_{n^j+1} \in C_{k_j}\right)\right]$$
$$\geq \sum_{k_{j_1}=1}^{K-1} \cdots \sum_{k_{j_w}=1}^{K-1} \left[\prod_{j\in S} P^R\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^L\left(Y^l_{n^l+1} \in \mathcal{C}_{M_x+1,K}\right)\right]$$

In this derivation, we use the assumptions $A_{(n_j)}$ for all groups with assumed independence of the J groups. The lower bound is obtained by putting the probability mass per interval at the right end-points for each group in S and for the other groups at the left end-points. The corresponding NPI upper probability (11) is obtained by putting the probability mass per interval at the opposite end-points, leading to

$$
\begin{aligned}
&P\left(\max_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right) = P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1}\right\}\right)\\
&= \sum_{k_{j_1}=1}^{K}\cdots\sum_{k_{j_w}=1}^{K}\left[P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1}\Big| Y^j_{n^j+1}\in C_{k_j}, j\in S\right\}\right)\prod_{j\in S} P\left(Y^j_{n^j+1}\in C_{k_j}\right)\right]\\
&\leq \sum_{k_{j_1}=1}^{K-1}\cdots\sum_{k_{j_w}=1}^{K-1}\left[P\left(\bigcap_{l\in S^c}\left\{Y^l_{n^l+1} > \max_{j\in S} Y^j_{n^j+1}\Big| Y^j_{n^j+1}\in C_{k_j}, j\in S\right\}\right)\prod_{j\in S} P^L\left(Y^j_{n^j+1}\in C_{k_j}\right)\right]\\
&\leq \sum_{k_{j_1}=1}^{K-1}\cdots\sum_{k_{j_w}=1}^{K-1}\left[\prod_{j\in S} P^L\left(Y^j_{n^j+1}\in C_{k_j}\right)\prod_{l\in S^c} P^R\left(Y^l_{n^l+1}\in \mathcal{C}_{M_x+1,K}\right)\right]
\end{aligned}
$$

Because the expressions in the final lines of these derivations 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.

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 $\underline{P}^{\leqslant}_S = \underline{P}\left(\max_{j\in S} Y^j_{n^j+1} \leqslant \min_{l\in S^c} Y^l_{n^l+1}\right)$ and $\overline{P}^{\leqslant}_S = \overline{P}\left(\max_{j\in S} Y^j_{n^j+1} \leqslant \min_{l\in S^c} Y^l_{n^l+1}\right)$. These NPI lower and upper probabilities can be derived similarly to $\underline{P}^{<}_S$ and $\overline{P}^{<}_S$ as presented above and are given by

$$
\underline{P}^{\leqslant}_S = \sum_{k_{j_1}=1}^{K}\cdots\sum_{k_{j_w}=1}^{K}\left[\prod_{j\in S} P^R\left(Y^j_{n^j+1}\in C_{k_j}\right)\prod_{l\in S^c} P^L\left(Y^l_{n^l+1}\in \mathcal{C}_{M_x,K}\right)\right] \tag{12}
$$

$$
\overline{P}^{\leqslant}_S = \sum_{k_{j_1}=1}^{K}\cdots\sum_{k_{j_w}=1}^{K}\left[\prod_{j\in S} P^L\left(Y^j_{n^j+1}\in C_{k_j}\right)\prod_{l\in S^c} P^R\left(Y^l_{n^l+1}\in \mathcal{C}_{M_x,K}\right)\right] \tag{13}
$$

Note that (12) and (13) only differ from (10) and (11), respectively, by the use of M_x instead of $M_x + 1$ in the events involving the groups in S^c and by the summations including the terms corresponding to category C_K.

A special case of the above results that is often of practical interest is if S contains just a single group, so the next observation from one group is compared to all other groups. The NPI lower and upper probabilities for the event that the next observation $Y^j_{n^j+1}$ from group j $(j = 1, \ldots, J)$ is less than the next observation from each of the other groups, so less than $Y^l_{n^l+1}$ for $l = 1, \ldots, J$, $l \neq j$, follow immediately from (10) and (11) using (8) and (9)

$$\underline{P}^<_j = \underline{P}\left(Y^j_{n^j+1} < \min_{\substack{1 \leq l \leq J \\ l \neq j}} Y^l_{n^l+1}\right) = \frac{1}{\prod_{j=1}^{J}(n^j+1)} \sum_{k=1}^{K-1}\left(n^j_k \prod_{\substack{l=1 \\ l \neq j}}^{J} n^l_{k+1,K}\right) \tag{14}$$

$$\overline{P}^<_j = \overline{P}\left(Y^j_{n^j+1} < \min_{\substack{1 \leq l \leq J \\ l \neq j}} Y^l_{n^l+1}\right) = \frac{1}{\prod_{j=1}^{J}(n^j+1)} \left[\sum_{k=1}^{K-1}\left(n^j_k \prod_{\substack{l=1 \\ l \neq j}}^{J} \left(n^l_{k+1,K}+1\right)\right) + \prod_{\substack{l=1 \\ l \neq j}}^{J}\left(n^l_{2,K}+1\right)\right] \tag{15}$$

The NPI lower and upper probabilities for the event that the next observation from group j is less than or equal to the next observation from each of the other groups follow from (12) and (13),

$$\underline{P}^\leqslant_j = \underline{P}\left(Y^j_{n^j+1} = \min_{1 \leq l \leq J} Y^l_{n^l+1}\right) = \frac{1}{\prod_{j=1}^{J}(n^j+1)} \left[n^j_1 \prod_{\substack{l=1 \\ l \neq j}}^{J}\left(n^l+1\right) + \sum_{k=2}^{K}\left(n^j_k \prod_{\substack{l=1 \\ l \neq j}}^{J} n^l_{k,K}\right) + \prod_{\substack{l=1 \\ l \neq j}}^{J} n^l_K\right] \tag{16}$$

$$\overline{P}^\leqslant_j = \overline{P}\left(Y^j_{n^j+1} = \min_{1 \leq l \leq J} Y^l_{n^l+1}\right) = \frac{1}{\prod_{j=1}^{J}(n^j+1)} \left[\prod_{\substack{l=1 \\ l \neq j}}^{J}\left(n^l+1\right) + \sum_{k=1}^{K}\left(n^j_k \prod_{\substack{l=1 \\ l \neq j}}^{J}\left(n^l_{k,K}+1\right)\right)\right] \tag{17}$$

If there are only two categories, $K = 2$, the NPI lower and upper probabilities presented above become as follows, with $A = \left(\prod_{j=1}^{J}\left(n^j+1\right)\right)^{-1}$,

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

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

These are identical to the corresponding NPI lower and upper probability for comparison of proportions presented by Coolen and Coolen-Schrijner [14].

A further special case of interest is if there are only two groups to be compared, so $J = 2$, in which case (14)-(17) lead to, with $\gamma = ((n^1+1)(n^2+1))^{-1}$,

$$\underline{P}(Y^1_{n^1+1} < Y^2_{n^2+1}) = \gamma \sum_{v=2}^{K}\sum_{w=1}^{v-1} n_w^1 n_v^2 \tag{18}$$

$$\overline{P}(Y^1_{n^1+1} < Y^2_{n^2+1}) = \gamma \left(\sum_{v=2}^{K}\sum_{w=1}^{v-1} n_w^1 n_v^2 + n^2 - n_1^2 + n^1 - n_K^1 + 1\right) \tag{19}$$

$$\underline{P}(Y^1_{n^1+1} \leq Y^2_{n^2+1}) = \gamma \left(\sum_{v=1}^{K}\sum_{w=1}^{v} n_w^1 n_v^2 + n_1^1 + n_K^2\right) \tag{20}$$

$$\overline{P}(Y^1_{n^1+1} \leq Y^2_{n^2+1}) = \gamma \left(\sum_{v=1}^{K}\sum_{w=1}^{v} n_w^1 n_v^2 + n^1 + n^2 + 1\right) \tag{21}$$

An alternative event that may be of interest in multiple comparisons is that the next observation of at least one group in S ($S = \{j_1, \ldots, j_w\} \subset \{1, \ldots, J\}$) 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

$$\underline{P}^{<}_{S^1} = \underline{P}\left(\min_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right) = \sum_{k_{j_1}=1}^{K} \cdots \sum_{k_{j_w}=1}^{K} \left[\prod_{j\in S} P^R\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^L\left(Y^l_{n^l+1} \in \mathcal{C}_{M_i+1,K}\right)\right] \tag{22}$$

$$\overline{P}^{<}_{S^1} = \overline{P}\left(\min_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right) = \sum_{k_{j_1}=1}^{K} \cdots \sum_{k_{j_w}=1}^{K} \left[\prod_{j\in S} P^L\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^R\left(Y^l_{n^l+1} \in \mathcal{C}_{M_i+1,K}\right)\right] \tag{23}$$

where $M_i = \min\{k_{j_1}, \ldots, k_{j_w}\}$ and P^L and P^R are again as given by (8) and (9), respectively. These NPI lower and upper probabilities are derived similarly to (10) and (11), with 'max' replaced by 'min' everywhere and 'M_x' replaced by 'M_i'. 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 $\underline{P}^{\leqslant}_{S^1}$ and $\overline{P}^{\leqslant}_{S^1}$, respectively, are similar to (22) and (23) but with $M_i + 1$ replaced by M_i in the events involving the groups in S^c in the probabilities on the right-hand side.

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

	C_1	C_2	C_3
Noncarriers (1)	497	560	269
Carriers (2)	19	29	24

Table 2: Tonsils data

S is greater than the next observation from each group in S^c are

$$\underline{P}\left(\min_{j\in S} Y^j_{n^j+1} > \max_{l\in S^c} Y^l_{n^l+1}\right) = \sum_{k_{j_1}=2}^{K} \cdots \sum_{k_{j_w}=2}^{K} \left[\prod_{j\in S} P^L\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^R\left(Y^l_{n^l+1} \in \mathcal{C}_{1,M_i-1}\right)\right] \tag{24}$$

$$\overline{P}\left(\min_{j\in S} Y^j_{n^j+1} > \max_{l\in S^c} Y^l_{n^l+1}\right) = \sum_{k_{j_1}=2}^{K} \cdots \sum_{k_{j_w}=2}^{K} \left[\prod_{j\in S} P^R\left(Y^j_{n^j+1} \in C_{k_j}\right) \prod_{l\in S^c} P^L\left(Y^l_{n^l+1} \in \mathcal{C}_{1,M_i-1}\right)\right] \tag{25}$$

The NPI lower and upper probabilities for the corresponding event with 'greater than or equal to' are derived by replacing $M_i - 1$ by M_i in equations (24) and (25). 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, $\max_{j\in S} Y^j_{n^j+1} > (\geq) \max_{l\in S^c} Y^l_{n^l+1}$ follow from equations (24) and (25) (and the similar formulae for $\underline{P}^{\geqslant}_S$ and $\overline{P}^{\geqslant}_S$ discussed above) by replacing M_i by M_x in those equations.

The multiple comparison inferences discussed in this section are illustrated using two examples. Example 2 considers the case with $J = 2$ groups, which provides a suitable starting point for illustration and discussion of the NPI approach for comparison of multiple groups of ordinal data. Example 3 presents an extensive comparison of $J = 4$ groups.

Example 2

We illustrate NPI comparison of two ordinal data sets using the data presented in Table 2, which were also used by Agresti [1] who provides further references to the origins of this data set. The data consider tonsil size for two groups of children, namely 1326 noncarriers (Group 1) and 72 carriers (Group 2) of streptococcus pyogenes. An observation in category C_1 implies that tonsils are present but not enlarged, C_2 that tonsils are enlarged and C_3 that tonsils are greatly enlarged.

The NPI lower and upper probabilities (18)-(21) for these data are $\underline{P}(Y^1_{1327} < Y^2_{73}) = \frac{39781}{1327\times73} = 0.4107$, $\overline{P}(Y^1_{1327} < Y^2_{73}) = \frac{40892}{1327\times73} = 0.4221$, $\underline{P}(Y^1_{1327} \leq Y^2_{73}) = \frac{72441}{1327\times73} = 0.7478$ and $\overline{P}(Y^1_{1327} \leq Y^2_{73}) = \frac{73319}{1327\times73} = 0.7569$. Agresti [1] considered all $1326 \times 72 = 95472$ different carrier-noncarrier pairs that can be put together from these children, of which for $19(560 + 269) + 29(269) = 23552$ pairs the noncarrier has larger tonsils than the carrier, hence for 71920 pairs the carrier's tonsils are as least as large as those of the noncarrier, and for 39781 pairs the carrier has the larger tonsils. Notice that the relative frequencies corresponding to these pairs, $\frac{39781}{95472} = 0.4167$ and $\frac{71920}{95472} = 0.7533$ are bounded by the corresponding NPI lower and upper probabilities. The NPI lower and upper probabilities presented in this paper all bound the corresponding relative frequencies in this way, which can easily be proven from the formulae presented and is an attractive property of NPI. In this example, the differences between corresponding NPI upper and lower probabilities are small, due to the large numbers of observations. Clearly, if one considers groups with fewer observations, there will be more imprecision. This NPI approach remains valid and keeps its attractive frequentist properties for all sizes of data sets, so inferences are not only valid by approximation for large samples as is often the case in more established frequentist statistical methods.

Example 3

The data in Table 3, taken from [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 ($K = 5$). We use this data set to illustrate the NPI lower and upper probabilities for the events discussed in this paper.

	Glasgow outcome scale					
Treatment Group (j)	Death C_1	Vegetative State C_2	Major Disability C_3	Minor Disability C_4	Good Recovery C_5	Total
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: SAH Data

S	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^\leqslant$	$\overline{P}_S^\leqslant$	$\Delta_S^\leqslant$
$\{1\}$	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 4: SAH Data: $\max_{j\in S} Y^j_{n^j+1} < (\leqslant) \min_{l\in S^c} Y^l_{n^l+1}$

Table 4 presents the NPI lower and upper probabilities (10) to (13) 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. For S containing only a single group, Group 4 has the smallest lower and upper probabilities of providing the minimal next observation (so worst outcome) while Group 1 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 $\left(\max_{j\in S} Y^j_{n^j+1} < \min_{l\in S^c} Y^l_{n^l+1}\right)$ implies the event $\left(\max_{j\in S} Y^j_{n^j+1} \leqslant \min_{l\in S^c} Y^l_{n^l+1}\right)$ and because there are multiple observations of each group in each category, the NPI lower and upper probabilities $\underline{P}_S^\leqslant$ and $\overline{P}_S^\leqslant$ are greater than $\underline{P}_S^<$ and $\overline{P}_S^<$. The imprecision is less than 0.01 for all events considered, reflecting the substantial numbers of data for all four groups.

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 5 has the same numbers per group as the original data in Table 5, but the numbers per category are changed such that for Groups 1 and 2 the numbers in C_1 and C_2 have substantially increased while those in C_4 and C_5 have decreased, and for Groups 3 and 4 this change is the other way around. This leads to the outcomes being far worse 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 6. 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 closer to either 0 or 1, the imprecision tends to be smaller than for lower and upper probabilities which are close to 0.5.

One detail of Table 6 that is of interest is seen by comparing the lower and upper probabilities for subsets including either Group 3 or Group 4. Considering S of size 1 or 2, the impression is that the values with Group 3 included in S 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, and care must be taken when determining the actual event(s) of interest.

	Glasgow outcome scale					
Group (j)	C_1	C_2	C_3	C_4	C_5	Total
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 5: Changed SAH Data

S	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^\leqslant$	$\overline{P}_S^\leqslant$	$\Delta_S^\leqslant$
$\{1\}$	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	0.0040	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 6: Changed SAH Data: $\max_{j \in S} Y^j_{n^j+1} < (\leqslant) \min_{l \in S^c} Y^l_{n^l+1}$

Table 7 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. Table 7 corresponds to the original SAH data in Table 3, Table 8 presents the NPI lower and upper probabilities for the same events but with the changed SAH data from Table 5. 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 7 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 these tables are equal to the corresponding values in Tables 4 and 6 as the events are identical. The change in the data is clearly reflected in the different values in these two tables.

Subset (S)	$\underline{P}^{<}_{S^1}$	$\overline{P}^{<}_{S^1}$	$\Delta^{<}_{S^1}$	$\underline{P}^{\leqslant}_{S^1}$	$\overline{P}^{\leqslant}_{S^1}$	$\Delta^{\leqslant}_{S^1}$
$\{1\}$	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 7: SAH Data: $\min_{j \in S} Y^j_{n^j+1} < (\leqslant) \min_{l \in S^c} Y^l_{n^l+1}$

The NPI lower and upper probabilities for the events considered in this paper, as illustrated in this example, can also be of use for subset selection, which is a topic that has received much interest in the statistics literature. For further discussion of this possible use we refer to Coolen

Subset (S)	$\underline{P}^{<}_{S^1}$	$\overline{P}^{<}_{S^1}$	$\Delta^{<}_{S^1}$	$\underline{P}^{\leqslant}_{S^1}$	$\overline{P}^{\leqslant}_{S^1}$	$\Delta^{\leqslant}_{S^1}$
$\{1\}$	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 8: Changed SAH Data: $\min_{j\in S} Y^j_{n^j+1} < (\leqslant) \min_{l\in S^c} Y^l_{n^l+1}$

and van der Laan [15] who presented NPI subset selection for real-valued observations, and to Coolen-Maturi *et al* [16, 17] who presented such selection for right-censored data.

We briefly consider what happens if, with the same data, some of the groups or some of the categories are combined. In Table 9 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 10. The main difference with the original results in Table 4 is that the differences between the $\underline{P}^{<}_S$ and $\overline{P}^{<}_S$ and the corresponding $\underline{P}^{\leqslant}_S$ and $\overline{P}^{\leqslant}_S$ 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 11 presents the NPI lower and upper probabilities for the comparison of

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

Table 9: Combined SAH Data

S	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^\leqslant$	$\overline{P}_S^\leqslant$	$\Delta_S^\leqslant$
$\{1\}$	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 10: Combined SAH Data: $\max_{j\in S} Y^j_{n^j+1} < (\leqslant) \min_{l\in S^c} Y^l_{n^l+1}$

Groups 1 to 3 after Group 4 is removed, using the data from Table 3. 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 11 to Table 4 that deleting Group 4 leads to increased NPI lower and upper probabilities for the events represented, which is logical as Group 4 not 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 [15].

S	$\underline{P}_S^<$	$\overline{P}_S^<$	$\Delta_S^<$	$\underline{P}_S^\leqslant$	$\overline{P}_S^\leqslant$	$\Delta_S^\leqslant$
$\{1\}$	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 11: SAH Data without Group 4: $\max_{j\in S} Y^j_{n^j+1} < (\leqslant) \min_{l\in S^c} Y^l_{n^l+1}$

6 Concluding remarks

The inferences presented in this paper are important as ordinal data occur in many application areas. Although the emphasis has been on multiple comparisons, many other inferences can be formulated in terms of future observations and hence the corresponding NPI lower and upper probabilities can be derived. 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. We have not yet expanded this approach to multiple future observations, which is an important topic for future research. 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 also require the development of NPI for 2-dimensional random quantities, which is an interesting and important research challenge.

References

[1] Agresti, A.: Generalized odds ratios for ordinal data. Biometrics 36, 59–67 (1980).

[2] Agresti, A.: Analysis of Ordinal Categorical Data (2nd ed.). Wiley, Hoboken, New Jersey. (2010).

[3] Arts, G.R.J., Coolen, F.P.A., van der Laan, P: Nonparametric predictive inference in statistical process control. Quality Technology and Quantitative Management 1, 201–216 (2004).

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

[5] Baker, R.M.: Multinomial nonparametric predictive inference: selection, classification and subcategory data. PhD Thesis, Durham University (www.npi-statistics.com) (2010).

[6] Bernard, J.M.: An introduction to the imprecise Dirichlet model for multinomial data. International Journal of Approximate Reasoning 39, 123-150 (2005).

[7] Coolen, F.P.A.: Comparing two populations based on low stochastic structure assumptions. Statistics & Probability Letters 29, 297–305 (1996).

[8] Coolen, F.P.A.: An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. Reliability Engineering and System Safety 56, 61–68 (1997).

[9] Coolen, F.P.A.: Low structure imprecise predictive inference for Bayes' problem. Statistics & Probability Letters 36, 349–357 (1998).

[10] Coolen, F.P.A.: On nonparametric predictive inference and objective Bayesianism. Journal of Logic, Language and Information 15, 21–47 (2006).

[11] Coolen, F.P.A., Augustin, T.: 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 (2005).

[12] Coolen, F.P.A., Augustin, T.: 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 (2009).

[13] Coolen, F.P.A., Coolen-Schrijner, P.: Nonparametric predictive subset selection for proportions. Statistics & Probability Letters 76, 1675–1684 (2006).

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

[15] Coolen, F.P.A., van der Laan, P.: Imprecise predictive selection based on low structure assumptions. Journal of Statistical Planning and Inference 98, 259–277 (2001).

[16] Coolen-Maturi, T., Coolen-Schrijner, P., Coolen, F.P.A.: Nonparametric predictive selection with early termination of experiments. Journal of Statistical Planning and Inference 141, 1403-1421 (2011).

[17] Coolen-Maturi, T., Coolen-Schrijner, P., Coolen, F.P.A.: Nonparametric predictive multiple comparisons of lifetime data. Communications in Statistics - Theory and Methods, to appear.

[18] De Finetti, B.: Theory of Probability. Wiley, London (1974).

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

[20] Lawless, J.F., Fredette, M.: Frequentist prediction intervals and predictive distributions. Biometrika 92, 529–542 (2005).

[21] Liu, I., Agresti, A.: The analysis of ordered categorical data: an overview and a survey of recent developments (with discussion). TEST 14, 1–73 (2005).

[22] Shafer, G.A.: Mathematical Theory of Evidence. Princeton University Press (1976).

[23] Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, London (1991).

[24] Walley, P.: Inferences from multinomial data: learning about a bag of marbles (with discussion). Journal of the Royal Statistical Society Series B 58, 3–57 (1996).

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

[26] Weichselberger, K.: Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg (2001).