On Nonparametric Predictive Inference for Ordinal Data

Frank P.A. Coolen*, Pauline Coolen-Schrijner**, and Tahani A. Maturi

Department of Mathematical Sciences
Durham University, Durham, DH1 3LE, UK
frank.coolen@durham.ac.uk, tahani.maturi@durham.ac.uk
http://npi-statistics.com

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 an example application the comparison of two groups of ordinal data is presented.

Key words: Categorical data; lower and upper probabilities; nonparametric predictive inference; ordinal data; pairwise comparison.

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 [2, 6]. In NPI, attention is restricted to one or more future observable random quantities, and Hill's assumption $A_{(n)}$ [11] is used to link these random quantities to data, in a way that is closely related to exchangeability [10]. Coolen and Augustin [7, 8] presented NPI for categorical data with no known relationship between the categories, as an alternative to the Imprecise Dirichlet Model (IDM) [15]. However, in many practical applications the categories are ordered, 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 paper presents the first NPI results 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 $A_{(n)}$. An excellent recent

* corresponding author

** This work was started by Pauline Coolen-Schrijner, and completed by the other authors after her death (April 2008).

overview of established statistical methods for ordinal data was presented by Liu and Agresti [12]. The IDM can be applied to ordinal data, see e.g. Coolen [4] who applied it to grouped lifetime data including right-censored observations, but it does not naturally use the ordering of the categories.

Section 2 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. These results are briefly compared to NPI for non-ordered categorical data [8] in Section 4. To illustrate the application of this new method to practical problems, comparison of two groups of ordinal data is briefly presented in Section 5. More general results, including multiple comparisons and inferences for multiple future observations, together with more detailed analyses of properties of such methods, will be presented elsewhere.

2 Nonparametric predictive inference

Nonparametric predictive inference [2, 6] is based on Hill's assumption $A_{(n)}$ [11]. 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 [6] but it is not relevant in this paper. For X_{n+1}, representing a future observation, $A_{(n)}$ [11] 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 [10]. 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' [10], which are lower and upper probabilities in interval probability theory [14, 16, 17].

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 [14], or just as the maximum lower (minimum upper) bound for the probability of A that follows from the assumptions made [6]. 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 [2] proved that NPI has strong consistency properties in the theory of interval probability [14, 16, 17]. Direct application of $A_{(n)}$ for inferential problems is only possible for real-valued random quantities. However, by using assumed latent variable representations and variations to $A_{(n)}$, NPI has been developed for different situations, including Bernoulli quantities [5]. Defining an assumption related to $A_{(n)}$, but on a circle instead of the real-line, Coolen [6] 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 [8], with as additional attractive feature the possibility to include both defined and undefined new categories in the event of interest [7]. Whilst 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 a sequential way, taking the inter-dependence of the multiple future observations into account. For example in NPI for Bernoulli quantities this was included throughout [5], and dependence of specific inferences on the choice of the number of future observations was explicitly studied in the context of multiple comparisons [9].

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 several 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 [5]. We restrict attention to a single future observation, the interesting case of ordinal data with multiple future observations will be presented elsewhere.

We assume that n observations are available, with only the number of observations in each category given. Let $n_l \geq 0$ be the number of observations in category C_l, for $l = 1, \ldots, k$, so $\sum_{l=1}^{k} n_l = n$. Let Y_{n+1} denote the random quantity representing the category a future observation will belong to. We wish to derive the NPI lower and upper probabilities for events $Y_{n+1} \in \bigcup_{l \in L} C_l$ 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) [14].

Using the latent variable representation, we assume that category C_l is represented by interval IC_l, with the intervals $IC_1, \ldots, IC_k$ forming a partition of the real-line and logically ordered, that is interval IC_l has neighbouring intervals IC_{l-1} to its left and IC_{l+1} to its right on the real-line (or only one of these neighbours if $l = 1$ or $l = k$, of course). We further assume that the n observations are represented by $x_1 < \ldots < x_n$, of which n_l are in interval IC_l, these are also denoted by x_i^l for $i = 1, \ldots, n_l$. A further latent variable X_{n+1} on the real-line corresponds to the future observation Y_{n+1}, so the event $Y_{n+1} \in C_l$ corresponds to the event $X_{n+1} \in IC_l$. 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 Fig. 1.

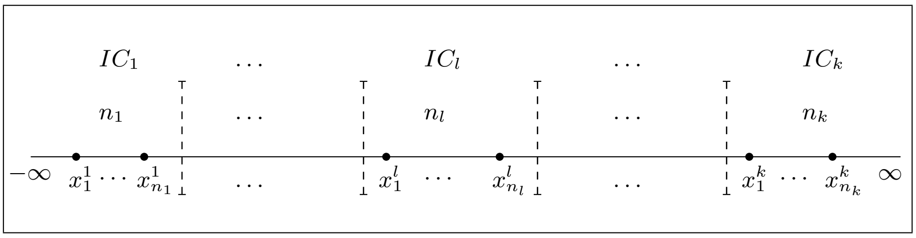

Fig. 1. Ordinal data structure

We now derive the NPI lower and upper probabilities for general events of the form $Y_{n+1} \in \mathcal{C}_L$, with $\mathcal{C}_L = \bigcup_{l \in L} C_l$ 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_l.

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 $\mathcal{IC}_L = \bigcup_{l \in L} IC_l$ 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

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

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_l, this may appear to be vague, yet the fact that we know the numbers of x_j values within each interval IC_l 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 [13], 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. 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.

3.1 Special cases

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 $\mathcal{C}_{s,t} = \bigcup_{l=s}^{t} C_l$, $\mathcal{IC}_{s,t} = \bigcup_{l=s}^{t} IC_l$ and let $n_{s,t} = \sum_{l=s}^{t} n_l$. Using the notation $(x)^+ = \max(x, 0)$, the NPI lower and upper probabilities (1) and (2) for such events are

$$\underline{P}(Y_{n+1} \in C_{s,t}) = \underline{P}(X_{n+1} \in \mathcal{IC}_{s,t}) = \begin{cases} \dfrac{(n_{s,t} - 1)^+}{n+1} & \text{if } 1 < s \leq t < k \\[2ex] \dfrac{n_{s,t}}{n+1} & \text{if } s = 1 \text{ or } t = k \end{cases} \tag{3}$$

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

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 $n_l > 0$ for all $l \in \{1, \ldots, k\}$, 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 $n_L = \sum_{l \in L} n_l$. Then the NPI lower probability (1), with L a strict subset of $\{1, \ldots, k\}$, is

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

and the corresponding NPI upper probability (2) is

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

These two special cases are likely to cover many situations of practical interest. The problem for 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.

3.2 Properties

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 [2] prove stronger consistency properties for NPI lower and upper probabilities for real-valued random quantities within the theory of Weichselberger [16, 17], 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 [14]. 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 [2]. The NPI lower and upper probabilities (1) and (2) bound the corresponding empirical probability for the event of interest, so

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

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 [6].

3.3 Example

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.

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

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.

4 Comparison to NPI for non-ordered categorical data

Coolen and Augustin [8] 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 $A_{(n)}$ [6] 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 [7] 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. For further details of NPI for non-ordered categorical data we refer to Coolen and Augustin [8], 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 [8]. 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 Comparison of two groups

In many applications of statistics, one aims at comparing multiple groups of data. We briefly illustrate how the NPI approach presented in this paper can be used for comparison of two groups of data, detailed justification of these results will be presented elsewhere, together with generalization to comparisons of more than two groups of data. Suppose that, as before, we consider k ordered categories, $C_1 < \ldots < C_k$, but now we have data for two independent groups which we wish to compare. Traditional statistical methods [12] tend to formulate problems of comparison of multiple groups as tests of hypotheses, but in NPI comparisons are necessarily predictive, hence one or more future observations per group are compared. Let us denote the two different groups by A and B, and we add a superscript a or b to our earlier notation to indicate the group. So, the total number of observations for group A (B) is n^a (n^b), of which n^a_j (n^b_j) are in category C_j. To use NPI for the comparison of these two groups, restricting attention to a single future observation per group, we assume $A_{(n^a)}$ for the next observation $Y^a_{n^a+1}$ from group A, and $A_{(n^b)}$ for the next observation $Y^b_{n^b+1}$ from group B, and per group we use the same latent variable representation as before.

Whilst ordinal data do not normally have meaningful associated location summaries (e.g. mean or median), due to the natural ordering of the categories it is meaningful to consider the events $Y^a_{n^a+1} < Y^b_{n^b+1}$ and $Y^a_{n^a+1} \leq Y^b_{n^b+1}$ for comparison of the two groups. For the corresponding underlying latent variables, this then follows NPI comparison of two groups of real-valued data as presented by Coolen [3], with the added complication that no actual observations are available for the latent variables and hence there is no knowledge about the ordering of values of the two groups within a category. Hence, the NPI lower and upper probabilities for these events are derived by minimisation and maximisation, respectively, of corresponding lower and upper probabilities for all possible orderings of the latent variables per category. This leads to the following NPI lower and upper probabilities, with $\gamma = ((n^a+1)(n^b+1))^{-1}$,

$$\underline{P}(Y^a_{n^a+1} < Y^b_{n^b+1}) = \gamma \sum_{v=2}^{k} \sum_{w=1}^{v-1} n^a_w n^b_v \tag{8}$$

$$\overline{P}(Y^a_{n^a+1} < Y^b_{n^b+1}) = \gamma \left(\sum_{v=2}^{k} \sum_{w=1}^{v-1} n^a_w n^b_v + n^b - n^b_1 + n^a - n^a_k + 1 \right) \tag{9}$$

and

$$\underline{P}(Y^a_{n^a+1} \leq Y^b_{n^b+1}) = \gamma \left(\sum_{v=1}^{k} \sum_{w=1}^{v} n^a_w n^b_v + n^a_1 + n^b_k \right) \tag{10}$$

$$\overline{P}(Y^a_{n^a+1} \leq Y^b_{n^b+1}) = \gamma \left(\sum_{v=1}^{k} \sum_{w=1}^{v} n^a_w n^b_v + n^a + n^b + 1 \right) \tag{11}$$

We illustrate such comparison of two ordinal data sets, using these NPI lower and upper probabilities, by considering 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 A) and 72 carriers (Group B) 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.

	C_1	C_2	C_3
Noncarriers (A)	497	560	269
Carriers (B)	19	29	24

Table 2. Data: size of tonsils

The NPI lower and upper probabilities (8)-(11) for these data are $\underline{P}(Y^a_{1327} < Y^b_{73}) = \frac{39781}{1327\times 73} = 0.4107$, $\overline{P}(Y^a_{1327} < Y^b_{73}) = \frac{40892}{1327\times 73} = 0.4221$, $\underline{P}(Y^a_{1327} \leq Y^b_{73}) = \frac{72441}{1327\times 73} = 0.7478$ and $\overline{P}(Y^a_{1327} \leq Y^b_{73}) = \frac{73319}{1327\times 73} = 0.7569$. 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. 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. However, this NPI approach remains valid and keeps its attractive frequentist properties for all sizes of data sets, so inferences are not only valid for large samples as is often the case in more established frequentist statistical methods.

References

1. Agresti, A.: Generalized odds ratios for ordinal data. Biometrics 36, 59–67 (1980)
2. Augustin, T., Coolen, F.P.A.: Nonparametric predictive inference and interval probability. Journal of Statistical Planning and Inference 124, 251–272 (2004)
3. Coolen, F.P.A.: Comparing two populations based on low stochastic structure assumptions. Statistics & Probability Letters 29, 297–305 (1996)
4. 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)
5. Coolen, F.P.A.: Low structure imprecise predictive inference for Bayes' problem. Statistics & Probability Letters 36, 349–357 (1998)
6. Coolen, F.P.A.: On nonparametric predictive inference and objective Bayesianism. Journal of Logic, Language and Information 15, 21–47 (2006)
7. 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)
8. 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)
9. Coolen, F.P.A., Coolen-Schrijner, P.: Nonparametric predictive comparison of proportions. Journal of Statistical Planning and Inference 137, 23–33 (2007)
10. De Finetti, B.: Theory of Probability: a Critical Introductory Treatment. Wiley, London (1974)
11. Hill, B.M.: Posterior distribution of percentiles: Bayes' theorem for sampling from a population. Journal of the American Statistical Association 63, 677–691 (1968)
12. 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)
13. Shafer, G.A.: Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
14. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, London (1991)
15. 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).
16. Weichselberger, K.: The theory of interval-probability as a unifying concept for uncertainty. International Journal of Approximate Reasoning 24, 149–170 (2000)
17. Weichselberger, K.: Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg (2001)