

Nonparametric predictive inference for order statistics of future observations

Frank P.A. Coolen and Tahani A. Maturi

Abstract Nonparametric predictive inference (NPI) is a powerful frequentist statistical framework which uses only few assumptions. Based on a post-data exchangeability assumption, precise probabilities for some events involving one or more future observations are defined, based on which lower and upper probabilities can be derived for all other events of interest. We present NPI for the r -th order statistic of m future real-valued observations and its use for comparison of two groups of data.

1 Introduction

Nonparametric predictive inference (NPI) [3, 5] is a statistical framework which uses few modelling assumptions, with inferences explicitly in terms of future observations. For real-valued random quantities attention has thus far been mostly restricted to a single future observation, although multiple future observations have been considered for some NPI methods for statistical process control [1, 2]. For Bernoulli quantities, NPI has also been presented for $m \geq 1$ future observations [4], with explicit study of the influence of the choice of m for comparison of groups of proportions data [6].

In this paper, we consider m future real-valued observations, given n observations, and as main contribution we focus on the r -th ordered observation of these m future observations, including comparison of two groups of data via comparison of their corresponding r -th ordered future observations. Without making further assumptions, these inferences require the use of lower and upper probabilities for several events of interest, as such this work fits in the theory of imprecise probability [12] and interval probability [13].

Dept. of Mathematical Sciences, Durham University, Durham, United Kingdom
frank.coolen@durham.ac.uk · tahani.maturi@durham.ac.uk

Assume that we have real-valued ordered data $x_1 < x_2 < \dots < x_n$, with $n \geq 1$. We assume that ties do not occur, in Example 2 in Section 3 we explain how to deal with ties. For ease of notation, define $x_0 = -\infty$ and $x_{n+1} = \infty$. The n observations partition the real-line into $n + 1$ intervals $I_j = (x_{j-1}, x_j)$ for $j = 1, \dots, n + 1$. If we wish to allow ties between past and future observations explicitly, we could use closed intervals $[x_{j-1}, x_j]$ instead of these open intervals I_j , the difference is rather minimal and to keep presentation easy we have opted not to do this here. We are interested in $m \geq 1$ future observations, X_{n+i} for $i = 1, \dots, m$. We link the data and future observations via Hill's assumption $A_{(n)}$ [10], or, more precisely, via $A_{(n+m-1)}$ (which implies $A_{(n+k)}$ for all $k = 0, 1, \dots, m - 2$; we will refer to this generically as 'the $A_{(n)}$ assumptions'), which can be considered as a post-data version of a finite exchangeability assumption for $n + m$ random quantities. $A_{(n+m-1)}$ implies that all possible orderings of the n data observations and the m future observations are equally likely, where the n data observations are not distinguished among each other, and neither are the m future observations. Let $S_j = \#\{X_{n+i} \in I_j, i = 1, \dots, m\}$, then assuming $A_{(n+m-1)}$ we have

$$P\left(\bigcap_{j=1}^{n+1} \{S_j = s_j\}\right) = \binom{n+m}{n}^{-1} \quad (1)$$

where s_j are non-negative integers with $\sum_{j=1}^{n+1} s_j = m$. Another convenient way to interpret the $A_{(n+m-1)}$ assumption with n data observations and m future observations is to think that n randomly chosen observations out of all $n + m$ real-valued observations are revealed, following which you wish to make inferences about the m unrevealed observations. The $A_{(n+m-1)}$ assumption then implies that one has no information about whether specific values of neighbouring revealed observations make it less or more likely that a future observation falls in between them. For any event involving the m future observations, (1) implies that we can count the number of such orderings for which this event holds. Generally in NPI a lower probability for the event of interest is derived by counting all orderings for which this event has to hold, while the corresponding upper probability is derived by counting all orderings for which this event can hold [3, 5].

NPI is close in nature to predictive inference for the low structure stochastic case as briefly outlined by Geisser [9], which is in line with many earlier nonparametric test methods where the interpretation of the inferences is in terms of confidence intervals. In NPI the $A_{(n)}$ assumptions justify the use of these inferences directly as probabilities. Using only precise probabilities or confidence statements, such inferences cannot be used for many events of interest, but in NPI we use the fact, in line with De Finetti's Fundamental Theorem of Probability [7], that corresponding optimal bounds can be derived for all events of interest [3]. NPI provides exactly calibrated frequentist inferences [11], and it has strong consistency properties in theory of interval

probability [3]. In NPI the n observations are explicitly used through the $A_{(n)}$ assumptions, yet as there is no use of conditioning as in the Bayesian framework, we do not use an explicit notation to indicate this use of the data. It is important to emphasize that there is no assumed population from which the n observations were randomly drawn, and hence also no assumptions on the sampling process. NPI is totally based on the $A_{(n)}$ assumptions, which however should be considered with care as they imply e.g. that the specific ordering in which the data appeared is irrelevant, so accepting $A_{(n)}$ implies an exchangeability judgement for the n observations. It is attractive that the appropriateness of this approach can be decided upon after the n observations have become available. NPI is always in line with inferences based on empirical distributions, which is an attractive property when aiming at objectivity [5].

2 NPI for order statistics

Let $X_{(r)}$, for $r = 1, \dots, m$, be the r -th ordered future observation, so $X_{(r)} = X_{n+i}$ for one $i = 1, \dots, m$ and $X_{(1)} < X_{(2)} < \dots < X_{(m)}$. The following probabilities are derived by counting the relevant orderings, and hold for $j = 1, \dots, n+1$, and $r = 1, \dots, m$,

$$P(X_{(r)} \in I_j) = \binom{j+r-2}{j-1} \binom{n-j+1+m-r}{n-j+1} \binom{n+m}{n}^{-1} \quad (2)$$

For this event NPI provides a precise probability, as each of the $\binom{n+m}{n}$ equally likely orderings of n past and m future observations has the r -th ordered future observation in precisely one interval I_j .

As an example, suppose that one is interested in the minimum $X_{(1)}$ of the m future observations. Formula (2) gives $P(X_{(1)} \in I_j) = \binom{n-j+m}{n-j+1} \binom{n+m}{n}^{-1}$, with for example $P(X_{(1)} \in I_1) = \frac{m}{n+m}$. Clearly, the event $X_{(1)} \in I_1$ occurs if the smallest of all $n+m$ observations, so the n data observations and m future observations, is not in the data set, which would occur with probability $\frac{n}{n+m}$. A further special case of interest is $P(X_{(1)} \in I_{n+1}) = \binom{n+m}{n}^{-1}$, following from the fact that there is only one ordering for which all n data observations occur before all m future observations.

In theory of mathematical statistics and probability, much attention is paid to limit results. Many popular statistical methods are justified through limit properties, with limits taken with regard to the number n of data observations, leading to 'large-sample' methods that are often applied in cases with relatively small samples without due consideration of the quality of the approximations involved and lacking clear foundational justification. Considering limits for n going to infinity is not very exciting in NPI as one just

ends up with the empirical distribution function and corresponding inferences. However, in NPI it might be of some interest to consider the limiting behaviour of the predictive probabilities (2) if m goes to infinity, hence if we consider an ever increasing future. Defining $\theta \in [0, 1]$ through the relationship $r = \theta m$ (of course, this only makes sense when θm , and therefore also $(1 - \theta)m$, is integer, we only sketch the argument here without giving the detailed mathematical presentation), the following limiting result is easily proven, for $j = 1, \dots, n + 1$,

$$\lim_{m \rightarrow \infty} P(X_{(\theta m)} \in I_j) = \binom{n}{j-1} \theta^{j-1} (1 - \theta)^{n-j+1} \quad (3)$$

It is important to emphasize the difference with established statistical methods. The θ in (3) is not a characteristic of an assumed population from which the data are sampled, indeed no population assumption is made. Furthermore, (3) is not a probability distribution nor a likelihood function for θ . Instead, θ only serves for notation of this event of interest, and indicates the specific relative (with regard to m) future order statistic of interest. Of course, the actual $A_{(n)}$ assumptions required for this limit imply infinite exchangeability of the future observations, hence De Finetti's Representation Theorem [7] indicates that a parametric representation can be assumed, yet this is different from the explicitly predictive use in NPI, most noticeably through the absence of a probability distribution for θ . The limiting probability (3) can be understood from the consideration that for the event $X_{(\theta m)} \in I_j$ to hold, precisely $j - 1$ of the n data observations must be smaller than $X_{(\theta m)}$, but it must be emphasized again that (3) specified probabilities for $X_{(\theta m)}$, not for any aspect of the observed data for which no concept of randomness, e.g. as following from assumed sampling from a population, is used in NPI. In NPI, the data are given, all randomness is explicitly with regard to the future observations, which nicely reflects where the uncertainty really is in applications.

Analysis of the probability (2) leads to some interesting results, including the obvious symmetry $P(X_{(r)} \in I_j) = P(X_{(m+1-r)} \in I_{n+2-j})$. For all r , the probability for $X_{(r)} \in I_j$ is unimodal in j , with the maximum probability assigned to interval I_{j^*} with $\left(\frac{r-1}{m-1}\right)(n+1) \leq j^* \leq \left(\frac{r-1}{m-1}\right)(n+1) + 1$. This carries through to the limiting situation (3), where for given θ the maximum probability is assigned to interval I_{j^*} with $(n+1)\theta \leq j^* \leq (n+1)\theta + 1$. It is worth commenting on extreme values, in particular inference involving $X_{(1)}$ or $X_{(m)}$ for m large compared to the value of n . In these cases, NPI assigns large probabilities to the intervals I_1 or I_{n+1} , respectively, which are outside the range of the observed data and unbounded unless the random quantities of interest are logically bounded (e.g. zero as lower bound for lifetime data). This indicates that, for such inferences, very little can be concluded without further assumptions on the probability masses within these end intervals beyond the observed data. This will be illustrated in the examples in Section

3. There are several inferential problems where one is explicitly interested in such a future order statistic $X_{(r)}$. It may be of explicit interest to compare different groups or treatments by comparing particular future order statistics, this is presented in Section 3.

3 Comparing two groups

Suppose we have two independent groups of real-valued observations, X and Y , their ordered observed values are $x_1 < x_2 < \dots < x_{n_x}$ and $y_1 < y_2 < \dots < y_{n_y}$. For ease of notation, let $x_0 = y_0 = -\infty$ and $x_{n_x+1} = y_{n_y+1} = \infty$. And let $I_{j_x}^x = (x_{j_x-1}, x_{j_x})$ and $I_{j_y}^y = (y_{j_y-1}, y_{j_y})$. We are interested in $m \geq 1$ future observations from each group (i.e. $m_x = m_y = m$), so in X_{n_x+i} and Y_{n_y+i} for $i = 1, \dots, m$. We wish to compare the r -th future order statistics from these two groups by considering the event $X_{(r)} < Y_{(r)}$, for which the NPI lower and upper probabilities, based on the $A_{(n_x)}$ and $A_{(n_y)}$ assumptions per group, are derived by

$$\underline{P}(X_{(r)} < Y_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x} < y_{j_y-1}\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(r)} \in I_{j_y}^y) \quad (4)$$

$$\overline{P}(X_{(r)} < Y_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x-1} < y_{j_y}\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(r)} \in I_{j_y}^y) \quad (5)$$

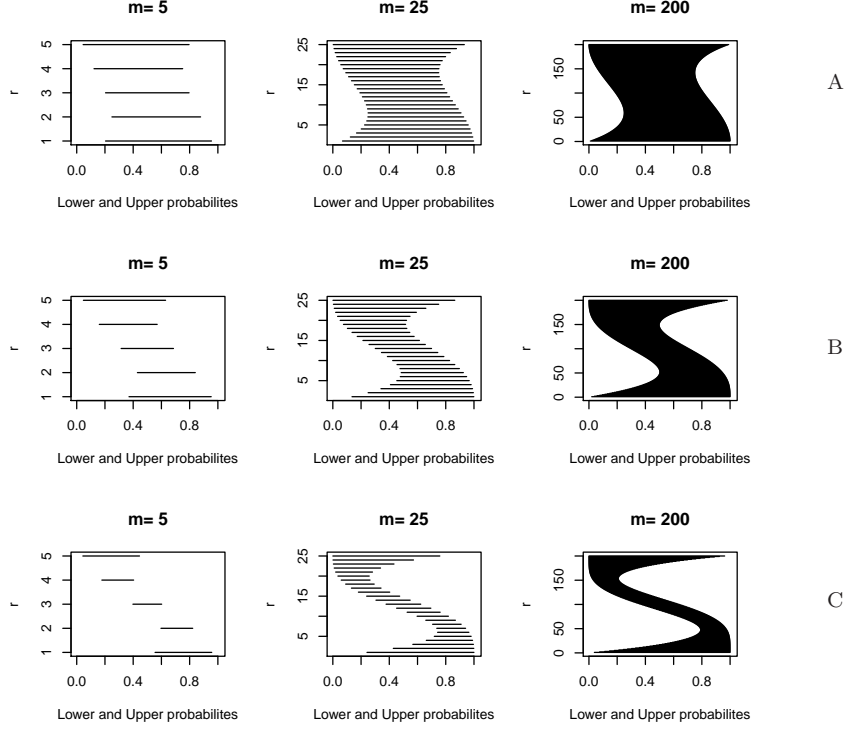
where $\mathbf{1}\{E\}$ is an indicator function which is equal to 1 if event E occurs and 0 else. This NPI lower (upper) probability follows by putting all probability masses for $Y_{(r)}$ corresponding to the intervals $I_{j_y}^y = (y_{j_y-1}, y_{j_y})$, $j_y = 1, \dots, n_y + 1$, to the left (right) end points of these intervals, and by putting all probability masses for $X_{(r)}$ corresponding to the intervals $I_{j_x}^x = (x_{j_x-1}, x_{j_x})$, $j_x = 1, \dots, n_x + 1$, to the right (left) end points of these intervals. We illustrate this NPI method for comparison of two groups based on the r -th future order statistic in two examples, first a small artificial example followed by one considering a real data set.

Example 1. To get a basic feeling for these inferences, we consider three small artificial data sets (cases A,B,C) as given in Table 1. For $m = 5, 25, 200$, the NPI lower and upper probabilities for the events $X_{(r)} < Y_{(r)}$ for all $r = 1, \dots, m$ are displayed in Fig. 1, with row 1,2,3 corresponding to cases A,B,C. Actually, the plotted lines per value of r represent the intervals bounded by the corresponding lower and upper probabilities, so the length of each line is the imprecision for that event.

These results illustrate clearly the effect of increased sample sizes, leading to decreasing imprecision for future order statistics that are most likely to fall within the observed data range. For extreme future order statistics,

Table 1 Data sets, Example 1

A	X: 1 4	Y: 2 3
B	X: 1 2 7 8	Y: 3 4 5 6
C	X: 1 2 3 4 13 14 15 16	Y: 5 6 7 8 9 10 11 12

**Fig. 1** NPI lower and upper probabilities, Example 1

imprecision remains high as no assumptions are made about the spread of probability mass within any interval $I_{j_x}^x$ or $I_{j_y}^y$, so also not in the end intervals. This makes clear that, without additional assumptions, no strong inferences can be achieved for events involving extreme future order statistics if m is substantially larger than n .

Example 2. We consider the data set of a study of the effect of ozone environment on rats growth [8, p.170]. One group of 22 rats were kept in an ozone environment and the second group of 23 similar rats were kept in an ozone-free environment. Both groups were kept for 7 days and their weight gains are given in Table 2.

The NPI lower and upper probabilities (4) and (5) for the events $X_{(r)} < Y_{(r)}$, $r = 1, \dots, m$, are displayed in Fig. 2, where the first row gives figures

Table 2 Rats' weight gains data, Example 2

Ozone group (X)						Ozone-free group (Y)					
-15.9	-14.7	-12.9	-9.9	-9.0	-9.0	-16.9	13.1	15.4	17.4	17.7	18.3
6.1	6.6	6.8	7.3	10.1	12.1	19.2	21.4	21.8	21.9	22.4	22.7
14.0	14.3	15.5	15.7	17.9	20.4	24.4	25.9	26.0	26.0	26.6	27.3
28.2	39.9	44.1	54.6			27.4	28.5	29.4	38.4	41.0	

corresponding to the full data for the cases with $m = 5, 25, 200$, while the second row gives the corresponding figures but with the observation -16.9 removed from group Y . This is done as this value could perhaps be considered to be an outlier, hence it might be interesting to see its influence on these inferences. Note that the data for group X and for group Y both contain two tied observations, at -9.0 and 26.0 , respectively. As tied observations are within the same group, we just add a very small amount to one of them, not affecting their rankings within the group nor with the data for both groups combined and therefore not affecting the inferences. This can be interpreted as assuming that these values actually differ in a further decimal, not reported due to rounding. If observations were tied among the two groups, the same breaking of ties could be performed, with the NPI method presented in this paper applied to all possible ways to do so, and the smallest (largest) of the corresponding lower (upper) probabilities for the event of interest would be used as the NPI lower (upper) probability. The possibility to break ties in this manner is an attractive feature of statistical methods using lower and upper probabilities, as it does not require further assumptions for such tied values.

This example shows that these data strongly support the event $X_{(r)} < Y_{(r)}$ for future order statistics that are likely to be in the middle area of the data ranges, with the values of the NPI lower and upper probabilities reflecting the amount of overlap in the observed data for groups X and Y . For extreme future order statistics the imprecision is again very large, and the effect of deleting the smallest Y value from the data has caused quite a difference between the inferences for small values of r , as the lower ends of the plots in rows 1 and 2 in Fig. 2 clearly illustrate.

References

1. Arts GRJ, Coolen FPA, van der Laan P (2004) Nonparametric predictive inference in statistical process control. *Quality Technology and Quantitative Management* 1:201–216
2. Arts GRJ, Coolen FPA (2008) Two nonparametric predictive control charts. *Journal of Statistical Theory and Practice* 2:499–512

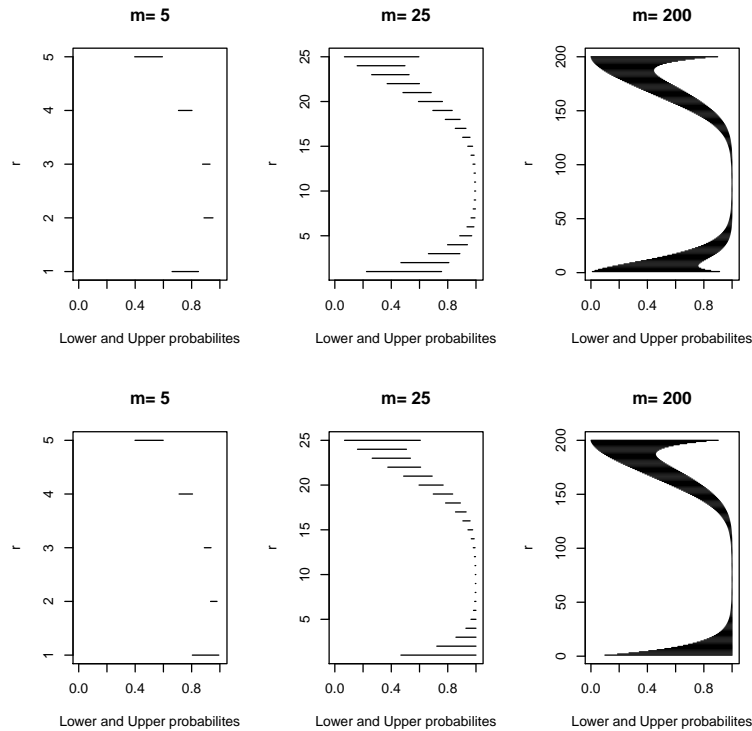


Fig. 2 NPI lower and upper probabilities, Example 2

3. Augustin T, Coolen FPA (2004) Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* 124:251–272
4. Coolen FPA (1998) Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters* 36:349–357
5. Coolen FPA (2006) On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information* 15:21–47
6. Coolen FPA, Coolen-Schrijner P (2007) Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference* 137:23–33
7. De Finetti B (1974) *Theory of probability* (2 volumes). Wiley, London
8. Desu MM, Raghavarao D (2004) *Nonparametric statistical methods for complete and censored data*. Chapman and Hall, Boca Raton
9. Geisser S (1993) *Predictive inference: an introduction*. Chapman and Hall, London
10. Hill BM (1968) Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association* 63:677–691
11. Lawless JF, Fredette M (2005) Frequentist prediction intervals and predictive distributions. *Biometrika* 92: 529–542
12. Walley P (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London
13. Weichselberger K (2001) *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept*. Physika, Heidelberg