Nonparametric predictive inference for future order statistics*

Frank P.A. Coolen†
Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK.
Tahani Coolen-Maturi ‡
Durham University Business School, Durham DH1 3LB, UK.
Hana N. Alqifari§
Department of Mathematics, Qassim University, Buraidah, Saudi Arabia.

March 21, 2017

Abstract

This paper presents nonparametric predictive inference for future order statistics. Given data consisting of n real-valued observations, m future observations are considered and predictive probabilities are presented for the r-th ordered future observation. In addition, joint and conditional probabilities for events involving multiple future order statistics are presented. The paper further presents the use of such predictive probabilities for order statistics in statistical inference, in particular considering pairwise and multiple comparisons based on two or more independent groups of data.

Keywords: Future order statistics, lower and upper probabilities, multiple comparisons, nonparametric predictive inference, pairwise comparisons.

1 Introduction

Nonparametric predictive inference (NPI) [6, 8] 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,

*The authors dedicate this paper to Professor Paul van der Laan (Eindhoven University of Technology, The Netherlands), who passed away in July 2016. Paul was PhD supervisor of the first-named author and 'scientific grandfather' of the other two authors. He published on a variety of topics in statistics, including selection methods and nonparametrics. He was an excellent statistician, motivating teacher and supervisor, and a kind and supportive person who is much missed.

†Email: frank.coolen@durham.ac.uk (Corresponding author)
‡Email: tahani.maturi@durham.ac.uk
§Email: h.n.a.alqifari@durham.ac.uk

although multiple future observations have been considered for NPI methods for statistical process control [4, 5] and system reliability [9].

In this paper, we consider m future real-valued observations, given n data observations, and we focus on the order statistics of these m future observations. Initial results were presented before by the first and second authors of this paper, in a conference paper [11], which only included the predictive probability for a single future order statistics (Equation (2) in the current paper) and pairwise comparisons (part of the current Section 3.1). We present the joint probability distribution for any collection of such order statistics over the intervals created by the partition of the real-line formed by the n data observations. We derive some properties for this probability distribution and we present its use for several inferential problems. Without making further assumptions, some of these inferences require the use of lower and upper probabilities, as such this work fits in the theory of imprecise probability [7, 24] and interval probability [25, 26].

Assume that we have real-valued ordered data $x_1 < x_2 < \ldots < x_n$, with $n \geq 1$. For ease of notation, define $x_0 = -\infty$ and $x_{n+1} = \infty$. The n observations create a partition of the real-line into $n+1$ intervals $I_j = (x_{j-1}, x_j)$ for $j = 1, \ldots, n+1$. We assume throughout this paper that ties do not occur. If we wish to allow ties, also between past and future observations, we could e.g. use closed intervals $[x_{j-1}, x_j]$ or half-open intervals 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 $i = 1, \ldots, m$. It should be emphasized that the future observations X_{n+i} are assumed to come from the same data collecting process as the n data observations, the use of the indices $n+i$ does not imply that the X_{n+i} are ordered in any way, so they are also not assumed to exceed the largest data observation x_n. We link the data and future observations via Hill's assumption $A_{(n)}$ [19], or, more precisely, via $A_{(n+m-1)}$ (which implies $A_{(n+k)}$ for all $k = 0, 1, \ldots, m-2$; 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. The $A_{(n)}$ assumptions imply 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 $S_j = \#\{X_{n+i} \in I_j,\ i = 1, \ldots, m\}$, then the $A_{(n)}$ assumptions lead to

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

where s_j are non-negative integers with $\sum_{j=1}^{n+1} s_j = m$. Another convenient way to interpret the $A_{(n)}$ assumptions 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 $A_{(n)}$ assumptions then imply 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 [6, 8].

NPI is close in nature to predictive inference for the low structure stochastic case as briefly outlined by Geisser [18], which is in line with many earlier nonparametric test methods where the interpretation of the inferences is in terms of confidence levels or intervals. In NPI the $A_{(n)}$ assumptions justify the use of these inferences directly as predictive 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 [15], that corresponding optimal bounds can be derived for all events of interest [6]. NPI provides exactly calibrated frequentist inferences [20], and it has strong consistency properties in theory of interval probability [6]. 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. However, the m future observations must result from the same sampling method as the n data observations in order to have full exchangeability. 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 [8].

This paper is organized as follows: In Section 2 we present the probability distributions for any collection of one or more future order statistics over the intervals I_j created by the n data observations, and we derive some properties of these distributions. The use of these distributions for a variety of inferential problems is presented in Section 3, with main focus on pairwise and multiple comparisons. Examples are provided to illustrate the new inferences. The paper ends with some concluding remarks in Section 4.

2 NPI for future order statistics

This section presents the core probability results on NPI for future order statistics. These will enable statistical inference involving order statistics for m future observations as presented in Section 3, and they also enable development of NPI methods for a range of problems in probability, statistics and related topic areas, as will be explored in future research.

2.1 NPI for the r-th ordered future observation

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

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

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 Equation (2) only specifies the probabilities for the events that $X_{(r)}$ belongs to intervals I_j, it can be considered to provide a partial specification of a probability distribution for $X_{(r)}$, no assumptions are made about the distribution of the probability masses within such intervals I_j.

Analysis of the probability in Equation (2) leads to some interesting results, including the logical 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 $\left(\frac{r-1}{m-1}\right)(n+1) \leq j^* \leq \left(\frac{r-1}{m-1}\right)(n+1)+1$. A further interesting property occurs for the special case where the number of future observations is equal to the number of data observations, so $m = n$. In this case, $P(X_{(r)} < x_r) = P(X_{(r)} > x_r) = 0.5$ holds for all $r = 1, \ldots, m$. This fact can be proven by considering all $\binom{2n}{n}$ equally likely orderings, where clearly in precisely half of these orderings the r-th future observation occurs before the r-th data observation due to the overall exchangeability assumption. The special case $m = n$ is not considered further in this paper, but it plays an important role in analysis of reproducibility of statistical hypothesis tests, for which the explicitly predictive nature of NPI is attractive [10]. Research into such reproducibility of tests using order statistics is currently being undertaken, results will be reported by Alqifari [2].

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, little can be concluded without further assumptions on the probability masses within these end intervals, so outside of the range of observed data.

2.2 Multiple order statistics of m future observations

The joint probability distribution of multiple order statistics of m future observations is of interest and can also be important for statistical inference. By straightforward combinatorial arguments, again counting the number of orderings for which the event of interest holds and using Equation (1), a partial specification of the probability distribution of any subset of the order statistics can be derived. Let $R = \{r_1, \ldots, r_t\} \subset \{1, \ldots, m\}$, with $r_1 < r_2 < \ldots < r_t$ and $1 \leq t \leq m$. For any set $I_R = \{j_{r_1}, \ldots, j_{r_t}\} \subset \{1, \ldots, n+1\}$, with $j_{r_1} \leq j_{r_2} \leq \ldots \leq j_{r_t}$, the $A_{(n)}$ assumptions imply the probabilities

$$P\left(\bigcap_{r \in R} \{X_{(r)} \in I_{j_r}\}\right) =$$
$$\binom{n+m}{n}^{-1} \binom{r_1 + j_{r_1} - 2}{r_1 - 1} \times \prod_{i=2}^{t} \binom{r_i - r_{i-1} - 1 + j_{r_i} - j_{r_{i-1}}}{r_i - r_{i-1} - 1} \times \binom{m - r_t + n - j_{r_t} + 1}{m - r_t} \tag{3}$$

For the special case of two order statistics, using notation $X_{(r)}$ and $X_{(s)}$ with $r < s$ and with $j \leq l$, we have

$$P(X_{(r)} \in I_j, X_{(s)} \in I_l) = \binom{n+m}{n}^{-1} \binom{r+j-2}{r-1} \binom{s-r-1+l-j}{s-r-1} \binom{m-s+n-l+1}{m-s} \tag{4}$$

2.3 Conditional probabilities given some future order statistics

Conditional probabilities on events involving a subset R of the order statistics, given information about another subset D of the future order statistics, is also of interest. Let $D = \{d_1, \ldots, d_v\}$, with $d_1 < d_2 < \ldots < d_v$ with $1 \leq v \leq m - t$ and such that $R \cap D = \emptyset$, and let $I_D = \{j_{d_1}, \ldots, j_{d_v}\} \subset \{1, \ldots, n+1\}$, with $j_{d_1} \leq j_{d_2} \leq \ldots \leq j_{d_v}$. To consider the conditional probability, we need to consider the joint probability for events involving all $X_{(c)}$ with $c \in C = R \cup D$, for which we use the notation $C = \{c_1, \ldots, c_w\}$, where $c_1 < c_2 < \ldots < c_w$ with $w = t + v$, and $I_C = \{j_{c_1}, \ldots, j_{c_w}\} \subset \{1, \ldots, n+1\}$, with $j_{c_1} \leq j_{c_2} \leq \ldots \leq j_{c_w}$.

The $A_{(n)}$ assumptions lead to the following conditional probabilities

$$P\left(\bigcap_{r \in R} \{X_{(r)} \in I_{j_r}\} \mid \bigcap_{d \in D} \{X_{(d)} \in I_{j_d}\}\right) = \frac{P\left(\bigcap_{r \in R} \{X_{(r)} \in I_{j_r}\} \cap \bigcap_{d \in D} \{X_{(d)} \in I_{j_d}\}\right)}{P\left(\bigcap_{d \in D} \{X_{(d)} \in I_{j_d}\}\right)} =$$
$$\frac{P\left(\bigcap_{c \in C} \{X_{(c)} \in I_{j_c}\}\right)}{P\left(\bigcap_{d \in D} \{X_{(d)} \in I_{j_d}\}\right)} = \frac{\binom{c_1 + j_{c_1} - 2}{c_1 - 1} \times \prod_{i=2}^{w} \binom{c_i - c_{i-1} - 1 + j_{c_i} - j_{c_{i-1}}}{c_i - c_{i-1} - 1} \times \binom{m - c_w + n - j_{c_w} + 1}{m - c_w}}{\binom{d_1 + j_{d_1} - 2}{d_1 - 1} \times \prod_{i=2}^{v} \binom{d_i - d_{i-1} - 1 + j_{d_i} - j_{d_{i-1}}}{d_i - d_{i-1} - 1} \times \binom{m - d_v + n - j_{d_v} + 1}{m - d_v}} \tag{5}$$

In case of interest in one future order statistic $X_{(r)}$ given information about one other future order statistic $X_{(d)}$, so the general case above with $t = v = 1$, this conditional probability for the case $r > d$ with $j \geq l$ is

$$P\left(X_{(r)} \in I_j \mid X_{(d)} \in I_l\right) = \frac{P\left(\{X_{(d)} \in I_l\} \cap \{X_{(r)} \in I_j\}\right)}{P\left(X_{(d)} \in I_l\right)} = \frac{\binom{r-d-1+j-l}{r-d-1} \binom{m-r+n-j+1}{m-r}}{\binom{n-l+1+m-d}{m-d}} \tag{6}$$

and for the case $r < d$ with $j \leq l$ this conditional probability is

$$P\left(X_{(r)} \in I_j \mid X_{(d)} \in I_l\right) = \frac{P\left(\left\{X_{(r)} \in I_j\right\} \cap \left\{X_{(d)} \in I_l\right\}\right)}{P\left(X_{(d)} \in I_l\right)} = \frac{\binom{r+j-2}{r-1}\binom{d-r-1+l-j}{d-r-1}}{\binom{d+l-2}{d-1}} \tag{7}$$

For completeness, the obvious case with $r = d$ gives $P\left(X_{(d)} \in I_j \mid X_{(d)} \in I_l\right)$ is equal to 1 if $j = l$ and is equal to 0 else.

It is straightforward to show that for the general conditional probability (5) the following property holds

$$P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D_R}\left\{X_{(d)} \in I_{j_d}\right\}\right) \tag{8}$$

with $D_R \subset D$ consisting of elements of D which in the combined set $C = R \cup D$ have an element of R as neighbour, so

$$D_R = \{c_i \in C \mid c_i \in D \text{ and } (c_{i-1} \in R \text{ or } c_{i+1} \in R), i \in \{1, \ldots, w\}\}$$

where the 'or' is of course not strict and events concerning the non-existent c_0 or c_{w+1} do not hold. Property (8) is easily shown to hold as factors for any $d \in D$ such that all its neighbouring values in C also belong to D appear in both the numerator and denominator of (5). Although this property is important in general, its main use may well be in predicting later order statistics on the basis of early order statistics [21], in which case it is a Markov property that also holds for order statistics in the classical theory [3, Sect. 2.4]. If $d_v < r_1$ and $j_{d_v} \leq j_{r_1}$ then

$$\begin{aligned} &P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid X_{(d_v)} \in I_{j_{d_v}}\right) \\ &= \frac{\binom{r_1-d_v-1+j_{r_1}-j_{d_v}}{r_1-d_v-1} \times \prod_{i=2}^{t}\binom{r_i-r_{i-1}-1+j_{r_i}-j_{r_{i-1}}}{r_i-r_{i-1}-1} \times \binom{m-r_t+n-j_{r_t}+1}{m-r_t}}{\binom{m-d_v+n-j_{d_v}+1}{m-d_v}} \end{aligned} \tag{9}$$

The backward analogue of this result may also be of use: If $d_1 > r_t$ and $j_{d_1} \geq j_{r_t}$ then

$$\begin{aligned} &P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid X_{(d_1)} \in I_{j_{d_1}}\right) \\ &= \frac{\binom{r_1+j_{r_1}-2}{r_1-1} \times \prod_{i=2}^{t}\binom{r_i-r_{i-1}-1+j_{r_i}-j_{r_{i-1}}}{r_i-r_{i-1}-1} \times \binom{d_1-r_t-1+j_{d_1}-j_{r_t}}{d_1-r_t-1}}{\binom{d_1+j_{d_1}-2}{d_1-1}} \end{aligned} \tag{10}$$

An interesting special case of the probability (6) is inference on a future order statistic $X_{(r)}$ given information about $X_{(r-1)}$, which by the above Markov property also includes the case of additional information on further earlier order statistics. For $j \geq l$,

$$P(X_{(r)} \in I_j \mid X_{(r-1)} \in I_l) = \frac{m-r+1}{n-l+m-r+2} \prod_{k=2}^{m-r+1} \left(\frac{n-j+k}{n-l+k} \right)$$

This is exactly the same as the probability for the event that $X_{(1)} \in I_{j-l+1}$, as given by Equation (2), for the case with $n-l+1$ data observations and $m-r+1$ future observations. A more general form of this result is presented in the following proposition, followed by a special case given as a corollary. The proofs of these properties are straightforward and hence not included.

Proposition 1. *For $r > d$ and $j \geq l$, the NPI probability that the rth future observation belongs to interval I_j, given that the dth future observation belongs to I_l, as given by Equation* (6)*, is equal to the NPI probability for the event that $X_{(r-d)} \in I_{j-l+1}$, as given by Equation* (2)*, for $n-l+1$ data observations and $m-d$ future observations. Similarly, for $r < d$ and $j \leq l$, the NPI probability for the event that the rth future observation belongs to I_j, given that the dth future observation belongs to I_l, as given by Equation* (7)*, is equal to the probability for the event that $X_{(d-r)} \in I_{l-j+1}$, as given by Equation* (2)*, for $l-1$ data observations and $d-1$ future observations.*

Corollary 1. *The conditional probability (6) for $j = l$ and $r > d$,*

$$P(X_{(r)} \in I_j | X_{(d)} \in I_j) = \prod_{k=0}^{r-d-1} \frac{m-r+1+k}{m-r+n-j+2+k} = \prod_{k=0}^{r-d-1} P(X_{(r-k)} \in I_j | X_{(r-k-1)} \in I_j)$$

That is, the probability for the event that the rth future observation belongs to I_j given that the dth future observation belongs to the same interval, is equal to the product of the probabilities for the events $X_{(r-k)} \in I_j$ given that its previous neighbour $X_{(r-k-1)} \in I_j$, for $k = 0, \ldots, r-d-1$. From Proposition 1, this probability is equal to the probability for the event that $X_{(r-d)} \in I_1$, as given by Equation (2), for $n-j+1$ data observations and $m-d$ future observations.

Corollary 2. *The conditional probability (7) for $j = l$ and $r < d$,*

$$P(X_{(r)} \in I_j | X_{(d)} \in I_j) = \prod_{k=0}^{d-r-1} \frac{r+k}{r+j-1+k} = \prod_{k=0}^{d-r-1} P(X_{(r+k)} \in I_j | X_{(r+1+k)} \in I_j)$$

That is, the probability for the event that the rth future observation belongs to I_j given that the dth future observation belongs to the same interval is equal to the product of the probabilities for the events $X_{(r+k)} \in I_j$ given that its next neighbour $X_{(r+1+k)} \in I_j$, for $k = 0, \ldots, d-r-1$. From Proposition 1, this probability is equal to the probability for the event that $X_{(d-r)} \in I_1$, as given by Equation (2), for $j-1$ data observations and $d-1$ future observations.

If one has information about two future order statistics, one either side of the $X_{(r)}$ of interest, than a similar result is presented next.

Proposition 2. *The conditional probability for $X_{(r)} \in I_j$ given that $X_{(d_1)} \in I_{j_{d_1}}$ and $X_{(d_2)} \in I_{j_{d_2}}$, for $d_1 < r < d_2$, is equal to the probability of $X_{(r-d_1)} \in I_{j-j_{d_1}+1}$, as given by Equation (2), for $j_{d_2} - j_{d_1}$ data observations and $d_2 - d_1 - 1$ future observations.*

Proof. The proof is straightforward using Equation (5) for the conditional probability $P(X_{(r)} \in I_j | X_{(d_1)} \in I_{j_{d_1}}, X_{(d_2)} \in I_{j_{d_2}})$ and Equation (2) for the probability for the event that $X_{(r-d_1)} \in I_{j-j_{d_1}+1}$, for $j_{d_2} - j_{d_1}$ data observations and $d_2 - d_1 - 1$ future observations. $\square$

The information used in the conditional probability (5) provides for each $X_{(d)}$, with $d \in D$, the interval in the partition created by the n observations in which this future order statistic is. One may wish to consider instead information in the form of precise values for some of the future order statistics. Due to the nature of NPI, where the $A_{(n)}$-based probabilities are assigned to intervals without further assumptions about their distribution within such intervals, such detailed information for some order statistics makes no difference to the probabilities assigned to intervals for other order statistics, except for the obviously required ordering of the order statistics.

Analysis of the conditional probability (7) leads to an interesting property of stochastic ordering for the comparison of two different conditional events for the same random quantities. Let $F_{r|d}(j|l)$ be the conditional cumulative distribution function (cdf) for $X_{(r)}$ given $X_{(d)} \in I_l$, which is defined as

$$F_{r|d}(j|l) = P(X_{(r)} \in \bigcup_{k=1}^{j} I_k | X_{(d)} \in I_l) = \sum_{k=1}^{j} P(X_{(r)} \in I_k | X_{(d)} \in I_l) \quad \text{for } j = 1, \ldots, n+1 \quad (11)$$

where $F_{r|d}(n+1|l) = 1$ and $F_{r|d}(1|1) = 1$.

Theorem 1. *Consider two future order observations $X_{(r)}$ and $X_{(d)}$ with $r < d$, and intervals $I_j = (x_{j-1}, x_j)$ and $I_{l-1} = (x_{l-2}, x_{l-1})$ with $j \leq l-1$. If*

$$F_{r|d}(j|l-1) \geq F_{r|d}(j|l) \qquad \textit{for all } j \qquad (12)$$

where $F_{r|d}(.|.)$ is the conditional cdf, then $X_{(r)} \in I_j | X_{(d)} \in I_{l-1}$ is said to be stochastically smaller than $X_{(r)} \in I_j | X_{(d)} \in I_l$, denoted by $X_{(r)} \in I_j | X_{(d)} \in I_{l-1} \leq_{st} X_{(r)} \in I_j | X_{(d)} \in I_l$. In general, for $l = 1, \ldots, n+1$, we have

$$F_{r|d}(j|n+1) \leq F_{r|d}(j|n) \leq \ldots \leq F_{r|d}(j|2) \leq F_{r|d}(j|1)$$

So $X_{(r)} \in I_j | X_{(d)} \in I_{n+1} \geq_{st} X_{(r)} \in I_j | X_{(d)} \in I_n \geq_{st} \ldots \geq_{st} X_{(r)} \in I_j | X_{(d)} \in I_1$.

The proof of Theorem 1 is given in the appendix.

2.4 NPI for $X_{(r)} \in S_l$ with S_l any subset of the real values

Thus far, we have considered probabilities for events $X_{(r)} \in I_j$, and related joint and conditional events. For all these, the $A_{(n)}$ assumptions provide precise probabilities. More generally, interest may be in the event $X_{(r)} \in S_l$ with S_l any subset of the real values, for example an interval not equal to one of the I_j created by the data. Generally, NPI provides bounds for the probability for such an event, where the maximum lower bound and minimum upper bound are lower and upper probabilities, respectively [6, 7, 24, 25, 26]. This can be regarded as an application of De Finetti's 'Fundamental Theorem of Probability' [15]. For any subset S_l of the real values, we can derive the NPI lower probability

$$\underline{P}\left(X_{(r)} \in S_l\right) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \subseteq S_l\} P\left(X_{(r)} \in I_j\right) \tag{13}$$

and the corresponding NPI upper probability

$$\overline{P}\left(X_{(r)} \in S_l\right) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \cap S_l \neq \phi\} P\left(X_{(r)} \in I_j\right) \tag{14}$$

3 Statistical inference

In this section, we present the application of NPI for future order statistics to statistical inference problems. We mainly focus on pairwise and multiple comparisons, and briefly outline some further possible inferences. As the classical theory of order statistics [3] has many applications to important statistical inference problems, there are many possible further applications that can be developed as topics for future research.

3.1 Pairwise comparisons

Suppose we have two independent groups of real-valued observations, X and Y, their ordered observed values are $x_1 < x_2 < \ldots < x_{n_x}$ and $y_1 < y_2 < \ldots < y_{n_y}$. For ease of notation, let $x_0 = y_0 = -\infty$ and $x_{n_x+1} = y_{n_y+1} = \infty$. Let $I^x_{j_x} = (x_{j_x-1}, x_{j_x})$ and $I^y_{j_y} = (y_{j_y-1}, y_{j_y})$. We focus attention on $m \geq 1$ future observations from each group (i.e. $m_x = m_y = m$), so in X_{n_x+i} and Y_{n_y+i} for $i = 1, \ldots, m$. The theory presented in this paper does not require limitation to the case $m_x = m_y$, but it seems to be quite logical when comparing future order statistics to use the same number of future observations for each group; generalization to different values for m_x and m_y is straightforward. Suppose that we wish to compare the r-th ordered future observation from group X to the s-th ordered future observation from group Y, by considering the event $X_{(r)} < Y_{(s)}$. The corresponding NPI lower and upper probabilities, based on the $A_{(n_x)}$

and $A_{(n_y)}$ assumptions per group, are derived by

$$\underline{P}(X_{(r)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x} < y_{j_y-1}\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(s)} \in I_{j_y}^y) \tag{15}$$

$$\overline{P}(X_{(r)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x-1} < y_{j_y}\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(s)} \in I_{j_y}^y) \tag{16}$$

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 $Y_{(s)}$ corresponding to the intervals $I_{j_y}^y = (y_{j_y-1}, y_{j_y})$, $j_y = 1, \ldots, n_y + 1$, to the left (right) end points of these intervals, and by putting all probability masses for $X_{(r)}$ corresponding to the intervals $I_{j_x}^x = (x_{j_x-1}, x_{j_x})$, $j_x = 1, \ldots, n_x + 1$, to the right (left) end points of these intervals.

One may wish to compare two groups by taking multiple future order statistics into account, this can be done using the probabilities presented in the previous section. As an example, suppose that we are interested in comparing two independent groups X and Y by simultaneously considering the rth and the sth future order statistics from each group. We can use the joint probability given by Equation (4) for any event involving the rth and sth future observations per group. Suppose that we are interested in the event $(X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)})$, with $r < s$, which can give insight into the spread of the future observations for the two groups. The NPI lower and upper probabilities for this event are given in the following theorem. Of course, such results for different events of interest are derived similarly.

Theorem 2. *The NPI lower and upper probabilities for the event $X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}$ are*

$$\begin{aligned}\underline{P}(X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}) = &\sum_{j_x=1}^{l_x-1} \sum_{l_x=j_x+1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \sum_{l_y=j_y}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}, x_{l_x-1} > y_{l_y}\} \\ &\times P(X_{(r)} \in I_{j_x}^x, X_{(s)} \in I_{l_x}^x) \times P(Y_{(r)} \in I_{j_y}^y, Y_{(s)} \in I_{l_y}^y)\end{aligned} \tag{17}$$

$$\begin{aligned}\overline{P}(X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}) = &\sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} \sum_{j_y=1}^{l_y-1} \sum_{l_y=j_y+1}^{n_y+1} 1\{x_{j_x-1} < y_{j_y}, x_{l_x} > y_{l_y-1}\} \\ &\times P(X_{(r)} \in I_{j_x}^x, X_{(s)} \in I_{l_x}^x) \times P(Y_{(r)} \in I_{j_y}^y, Y_{(s)} \in I_{l_y}^y) \\ &+ \max\left[P_{(y_{j_y}^*)}\right]\end{aligned} \tag{18}$$

where

$$P_{(y_{j_y}^*)} = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} \sum_{l_y=j_y=1}^{n_y+1} 1\{x_{j_x-1} < y_{j_y}^*, x_{l_x} > y_{j_y}^*\} P(X_{(r)} \in I_{j_x}^x, X_{(s)} \in I_{l_x}^x) P(Y_{(r)} \in I_{j_y}^y, Y_{(s)} \in I_{j_y}^y) \tag{19}$$

The maximisation remaining in Equation (18) is over all $y_{j_y}^* \in I_{j_y}^y$.

The proof of Theorem 2 is given in the appendix, where also the remaining maximisation for the derivation of the upper probability is discussed. We illustrate such NPI pairwise comparisons based on future order statistics in the following example.

Example 1. We consider the data set of a study of the effect of ozone environment on rats growth [16, p.170]. One group of 22 rats were kept in an ozone containing 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 1.

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	

Table 1: Rats weight gain data

The NPI lower and upper probabilities (15) and (16) for the events $X_{(r)} < Y_{(r)}$, $r = 1, \dots, m$, are displayed in Figure 1, where the first row gives figures 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 where 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. We should emphasize that the suggested manner for dealing with ties in NPI, discussed in Section 1, by replacing open intervals by closed intervals for the $A_{(n)}$ assumption, could also have been used here, it would have led to the same results as the simple method of breaking the ties we employed here, because the ties only occur within the same groups and the inferences presented only depend on the rankings of the X and Y group observations among eachother, which are not affected by very small additional values to break the ties nor by change to closed intervals in $A_{(n)}$.

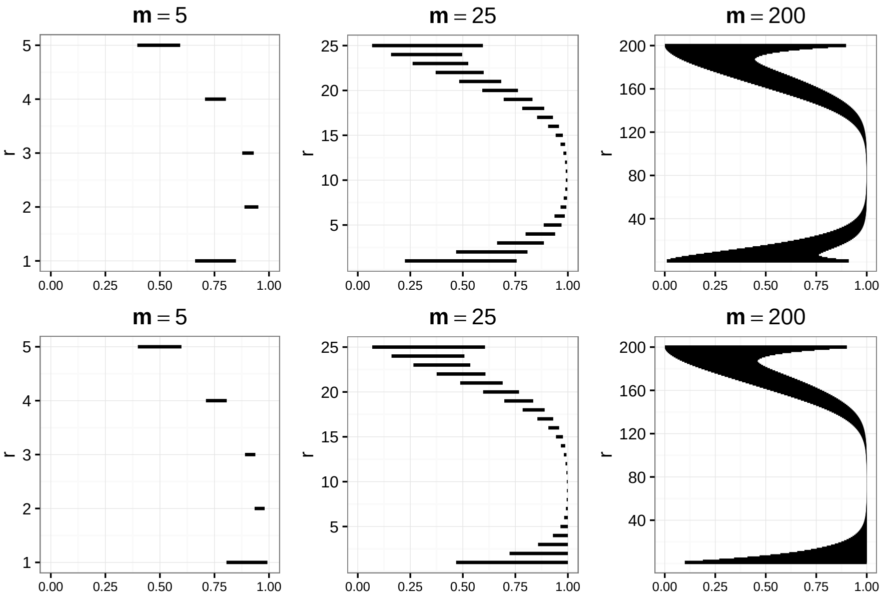

Figure 1: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)})$

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 very large when m is greater than n, due to the fact that those future order statistics are quite likely to both fall in the first or last interval, in which case very little can be said about the comparison of their values. Deleting the smallest Y value from the data, as shown in the second row in this figure, has quite some effect on inferences for small values of r, as the lower parts of the plots in rows 1 and 2 in Figure 1 clearly illustrate, but deleting this possible outlier does not have a noticable effect when larger values of r are used for the pairwise comparison.

To illustrate pairwise comparison using different order statistics for the two groups, we consider the case with $m = 200$ and interest in events $X_{(r)} < Y_{(s)}$. Figure 2 presents the NPI lower and upper probability for these events for the values $r = 1, 50, 100, 150, 200$ and for all $s = 1, \ldots, m$. Note that here the smallest Y observation, -16.9, has been deleted from the data. For $r = 1$ it is very likely that $Y_{(s)} > X_{(1)}$ for nearly all s, apart from the smallest values of s for which we get almost vacuous lower and upper probabilities for this event, that means upper probability of about 1 and lower probability of about 0, so imprecision (difference between the upper and lower probabilities for an event) close to 1. This reflects that the X data set contains quite a few observations which are smaller than all Y data values, and also the earlier discussed fact that one gets much imprecision for extreme future order observations if m is substantially greater than n_x and n_y. Note that for $r = 200$ the effect is very similar, due to the X group data containing the two overall largest observations. The plot for $r = 150$ may well be most informative, with e.g. the event $X_{(150)} < Y_{(s)}$ having lower probability greater than 0.5 already for s from just below 60 onwards.

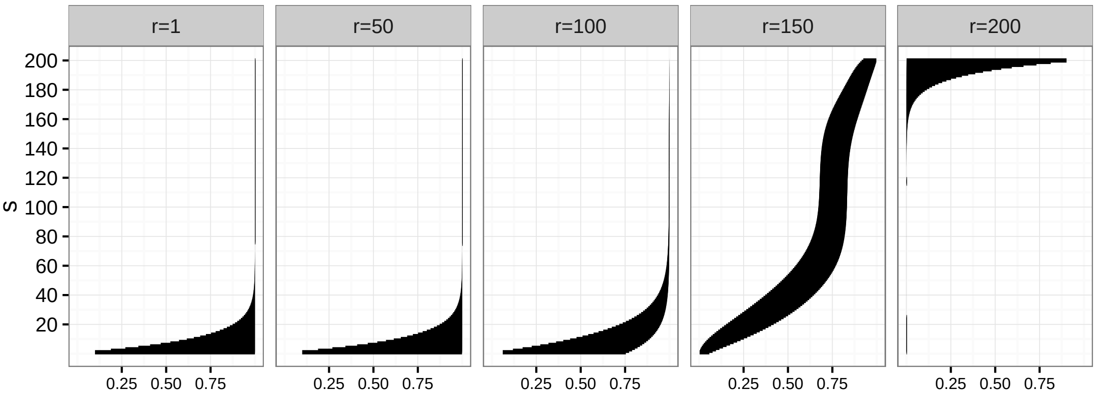

Figure 2: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(s)})$ for $m = 200$

The NPI lower and upper probabilities for the events $(X_{(r)} < Y_{(r)}, Y_{(s)} < X_{(s)})$, with $r < s$, are presented for these data in Figure 3, for the case with $m = 100$ future observations for both groups X and Y. Note that again the smallest Y observation, -16.9, has been deleted. The presented cases are for $r = 5, 10, 25, 50, 75$, and for all $s = r + 1, \ldots, m$. So this event of interest is whether the values $Y_{(r)}$ and $Y_{(s)}$ will both be in the interval $(X_{(r)}, X_{(s)})$. For small values of r it is likely that $X_{(r)} < Y_{(r)}$, as the X data contain the smallest overall observations. So the results for the case $r = 5$ are largely influenced by the event $Y_{(s)} < X_{(s)}$, which for most values of s is quite unlikely to happen, yet for large values of s it becomes well possible, reflecting that the two largest overall data observations belong to group X. Again we see much imprecision for the extreme order statistics.

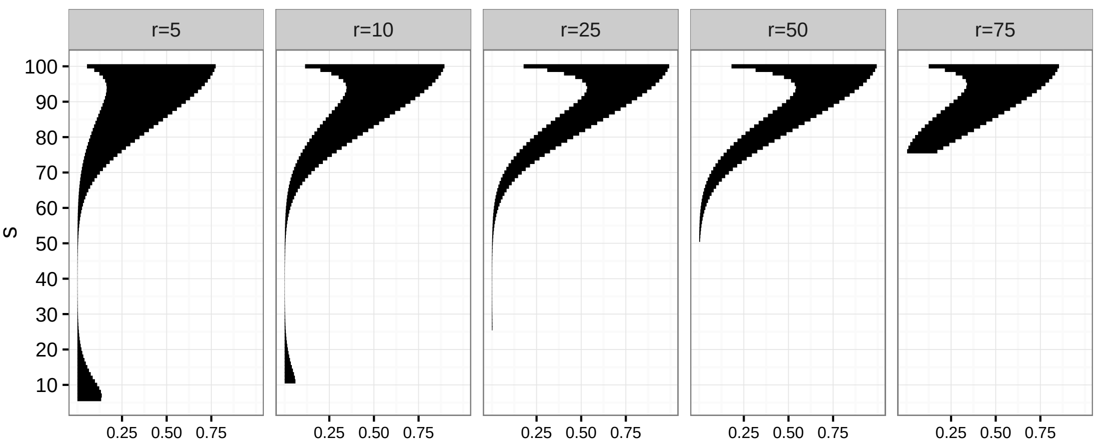

Figure 3: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)}, Y_{(s)} < X_{(s)})$ for $r < s$ and $m = 100$

3.2 Multiple comparisons

Coolen and van der Laan [12] presented NPI methods for comparisons of multiple groups, with different events of interest formulated in terms of the next future observation from each group, selecting the best group, the subset of best groups, and the subset that includes the best group. In this section we present NPI multiple comparisons methods based on order statistics of multiple future observations.

First we consider the selection of the best group based on the value of a single future order statistic. Suppose that there are $k \geq 2$ independent groups of real-valued observations, $X^1, X^2, \dots, X^k$, their ordered observed values are $x_1^g < x_2^g < \dots < x_{n_g}^g$ for each group $g = 1, \dots, k$. For ease of notation let $x_0^g = -\infty$ and $x_{n_g+1}^g = \infty$, and let $I_{j_g}^g = (x_{j_g-1}^g, x_{j_g}^g)$. We are interested in $m \geq 1$ future observations from each group, so in $X_{n_g+i}^g$ for $i = 1, \dots, m$, $g = 1 \dots, k$. As before, we consider inference based on the $A_{(n)}$ assumptions for each group. We are interested in the event that a specific $X_{(r)}^l$ is the maximum of all future observations $X_{(r)}^g$, $g = 1, \dots, k$. For this event, the following NPI lower and upper probabilities hold,

$$\underline{P}_l = \underline{P}(X_{(r)}^l = \max_{1 \leq g \leq k} X_{(r)}^g) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1 \\ g \neq l}}^{k} \sum_{j_g=1}^{n_g+1} \mathbf{1}\{x_{j_g}^g < x_{j_l-1}^l\} p(X_{(r)}^g \in I_{j_g}^g) p(X_{(r)}^l \in I_{j_l}^l) \quad (20)$$

$$\overline{P}_l = \overline{P}(X_{(r)}^l = \max_{1 \leq g \leq k} X_{(r)}^g) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1 \\ g \neq l}}^{k} \sum_{j_g=1}^{n_g+1} \mathbf{1}\{x_{j_g-1}^g < x_{j_l}^l\} p(X_{(r)}^g \in I_{j_g}^g) p(X_{(r)}^l \in I_{j_l}^l) \quad (21)$$

This NPI lower probability is obtained by putting the probability mass per interval at end points; for group l at the left end point and for all other groups at the right end point. Similarly, this NPI upper probability is obtained by putting the probability mass per interval for group l at the right end point and for all other groups at the left end point. We will refer to these as the lower and upper probabilities that group l is the best of all groups, where 'best group' is clearly to be interpreted in terms of the rth ordered future observation for each group.

In theory of statistical selection [12, 22] the interest is often in subsets of the groups, for example for use in screening processes where initially all groups are involved in tests, but later stages of testing can only involve a subset of the groups. One logical problem formulation involves selecting a subset of the groups which is most likely to contain all the best groups. We now derive the NPI method for such inferences, again with 'best group' in terms of the value of the r-th ordered value from m future observations. Suppose that a subset of the k independent groups contains w groups, with $2 \leq w \leq k-1$. Let $S = \{l_1, \dots, l_w\} \subset \{1, \dots, k\}$ be the subset of indices of these w groups, and let $NS = \{1, \dots, k\}/S$ be the subset of indices of the $k - w$

groups not in this subset. The NPI lower and upper probabilities for this event of interest are

$$\underline{P}_S = \underline{P}(\min_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} < \min_{l\in S} x^{l_s}_{j_{l_s}-1}\}\times$$
$$P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{22}$$
$$\overline{P}_S = \overline{P}(\min_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} < \min_{l\in S} x^{l_s}_{j_{l_s}}\}\times$$
$$P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{23}$$

The proofs of (22) and (23) are given in the appendix.

A second common group selection problem for which classical statistical methods have been presented is the variation with the aim that the selected subset should contain the single best group, so in our case the group which provides the maximum r-th ordered future observation. We can use the same notation as just introduced for selection of the subset containing all the best groups. The NPI lower and upper probabilities for the event that the rth future observation from (at least) one of the selected groups in S is greater than the rth future observation from all nonselected groups in NS, are derived similarly to the NPI lower and upper probabilities (22) and (23), as presented in the appendix, but with min everywhere replaced by max. These NPI lower and upper probabilities are

$$\underline{P}^*_S = \underline{P}(\max_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} < \max_{l\in S} x^{l_s}_{j_{l_s}-1}\}\times$$
$$P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{24}$$
$$\overline{P}^*_S = \overline{P}(\max_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} < \max_{l\in S} x^{l_s}_{j_{l_s}}\}\times$$
$$P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{25}$$

This second subset selection method is illustrated in the following example. There are, of course, a substantial number of further subset selection problem formulations that could be considered, including subsets containing the two best groups or criteria based on multiple future ordered observations. The NPI approach to such problems follows steps that are similar to those presented here, investigation of properties and performance will be of interest but is left as a topic for future research.

Example 2. We illustrate some of the above presented NPI methods for multiple comparisons based on order statistics of future observations, using data from Coolen and van der Laan [12] with sample sizes $n_1 = 20$, $n_2 = 18$, $n_3 = 15$, $n_4 = 3$ as presented in Table 2. A wider range of such methods are considered and illustrated by Alqifari [2].

l	Data
1	5.01 5.04 5.60 5.78 6.43 6.53 6.96 7.00 7.21 7.58 8.12 8.26 8.27 8.34 8.62 8.66 8.91 8.94 9.05 9.16
2	4.50 4.86 5.10 5.15 5.17 5.34 5.99 6.18 6.72 7.39 7.44 7.46 7.47 7.76 8.38 8.42 8.52 8.81
3	6.84 6.91 7.22 7.24 7.25 7.35 7.55 7.62 7.69 7.98 7.99 8.04 8.08 8.18 8.97
4	4.71 8.20 9.03

Table 2: Ordered data, Example 2

The NPI lower and upper probabilities for group l to be best, in terms of providing the largest value of the r-th ordered out of $m = 5$ future observations, for each group, are presented in Table 3, so these are $\underline{P}_l$ and $\overline{P}_l$ as given in Equations (20) and (21). These NPI lower and upper probabilities are also presented, for the case with $m = 10$ and all $r = 1, \ldots, 10$, in Figure 4. The imprecision in these lower and upper probabilities tends to be largest for small and large values of r, reflecting the earlier discussed feature of increased imprecision due to probabilities assigned to the first or last intervals. Group 3 is most likely to provide the largest future value for $r = 1$, but is quite unlikely to provide the largest future value for $r > m/2$, which appears most likely to come from Group 4. However, imprecision in these lower and upper probabilities is largest for Group 4, which reflects the fact that there are only 3 data observations from this group.

r	$\underline{P}_1$	$\overline{P}_1$	$\underline{P}_2$	$\overline{P}_2$	$\underline{P}_3$	$\overline{P}_3$	$\underline{P}_4$	$\overline{P}_4$
1	0.0682	0.2732	0.0199	0.1296	0.3798	0.7752	0.1034	0.3804
2	0.1342	0.2946	0.0373	0.1207	0.1818	0.4700	0.2389	0.5883
3	0.1893	0.4216	0.0422	0.1427	0.0543	0.2218	0.3138	0.6753
4	0.1716	0.4922	0.0364	0.1736	0.0185	0.1424	0.3166	0.7268
5	0.1003	0.6010	0.0079	0.2582	0.0105	0.2981	0.1840	0.7965

Table 3: $[\underline{P}_l, \overline{P}_l](X^l_{(r)} = \max_{1 \leq g \leq k} X^g_{(r)})$ for $m = 5$

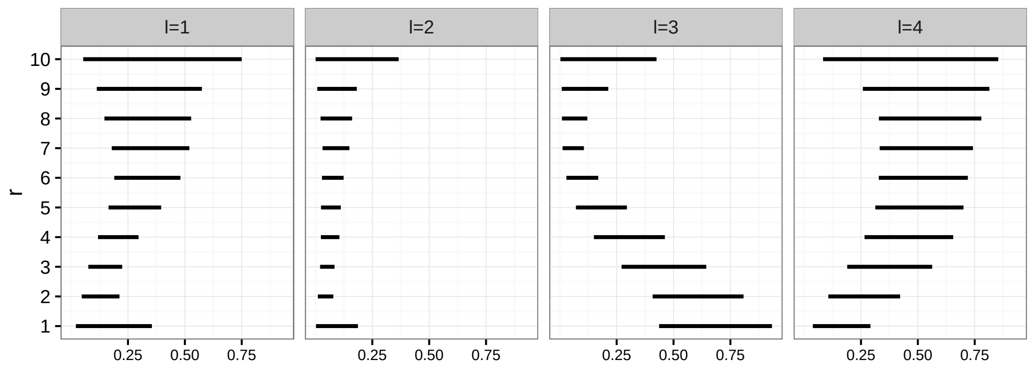

Figure 4: $[\underline{P}_l, \overline{P}_l](X^l_{(r)} = \max X^g_{(r)})$ for $m = 10$

As Group 4 only has 3 data observations, it is of interest to consider the effect on these inferences when this group is deleted. We denote the NPI lower and upper probabilities in this case by $\underline{P}_l^{(-4)}$ and $\overline{P}_l^{(-4)}$, they are presented in Table 4 for $m = 5$ and in Figure 5 for $m = 10$. Of course, as Group 4 was quite likely to lead to the largest r-th ordered future observation for the larger values of r, with this group deleted the corresponding lower and upper probabilities for the 3 remaining groups have increased, where particularly Group 1 benefits from the absence of Group 4. The overall pattern of these lower and upper probabilities for different values of r, as best seen from Figure 5, remains quite similar for these 3 groups in both cases with and without Group 4, but imprecision has decreased. This shows that the presence of a group with only few observations results in more imprecision for the other groups, so inclusion of a group with only few observations may reduce the overall quality of statistical inferences for such selection problems in the following sense. NPI provides exactly calibrated frequentist inferences [20], as discussed in Section 1, but it only provides inferences in terms of lower and upper probabilities. Hence, one can consider the level of imprecision a reflection of quality of the statistical inferences, which remain exactly calibrated both with and without inclusion of Group 4 in this example, but less imprecision provides more insight.

r	$\underline{P}_1^{(-4)}$	$\overline{P}_1^{(-4)}$	$\underline{P}_2^{(-4)}$	$\overline{P}_2^{(-4)}$	$\underline{P}_3^{(-4)}$	$\overline{P}_3^{(-4)}$
1	0.0987	0.3022	0.0311	0.1481	0.6076	0.8682
2	0.2425	0.3569	0.0883	0.1595	0.5079	0.6564
3	0.4284	0.5504	0.1166	0.2064	0.2891	0.4200
4	0.5239	0.6971	0.1353	0.2587	0.1323	0.2673
5	0.4003	0.7648	0.0744	0.3266	0.1045	0.3865

Table 4: $[\underline{P}_l^{(-4)}, \overline{P}_l^{(-4)}](X_{(r)}^l = \max_{1\leq g\leq k} X_{(r)}^g)$ for $m = 5$

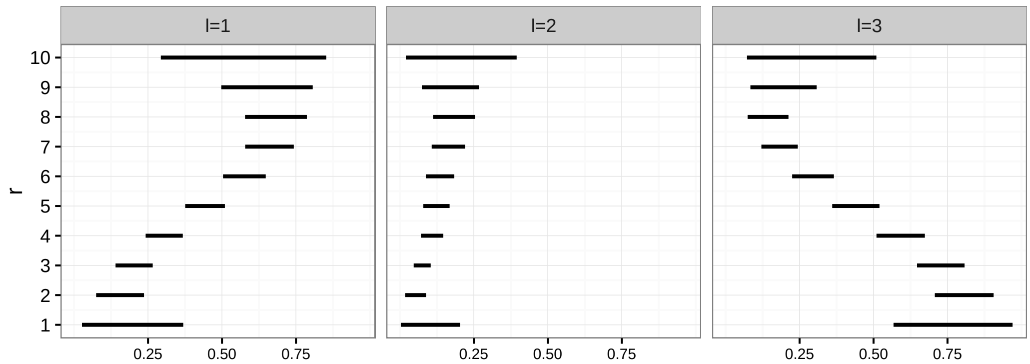

Figure 5: $[\underline{P}_l^{(-4)}, \overline{P}_l^{(-4)}](X_{(r)}^l = \max X_{(r)}^g)$ for $m = 10$

Figure 6 presents the NPI lower and upper probabilities for pairwise comparisons between these groups based on the r-th ordered future observation, for $m = 10$ and each $r = 1, \ldots, 10$. So the events considered are $X_{(r)}^l > X_{(r)}^g$ for $l, g = 1, \ldots, 4$ and $l \neq g$. It should be noted that NPI lower and upper probabilities for events not included in this figure can be deduced using the conjugacy property, that is $\underline{P}(A) = 1 - \overline{P}(A^c)$, for any event A and its complementary event A^c, which holds for NPI-based inferences and is a common property in theory of imprecise probability [6, 7]. These pairwise comparisons also show that Group 3 is most likely to provide the largest r-th ordered future observation for small values of r, while it is also clear that the lower and upper probabilities for comparisons involving Group 4 are more imprecise than for comparisons not involving Group 4, which again results from the small data set for Group 4.

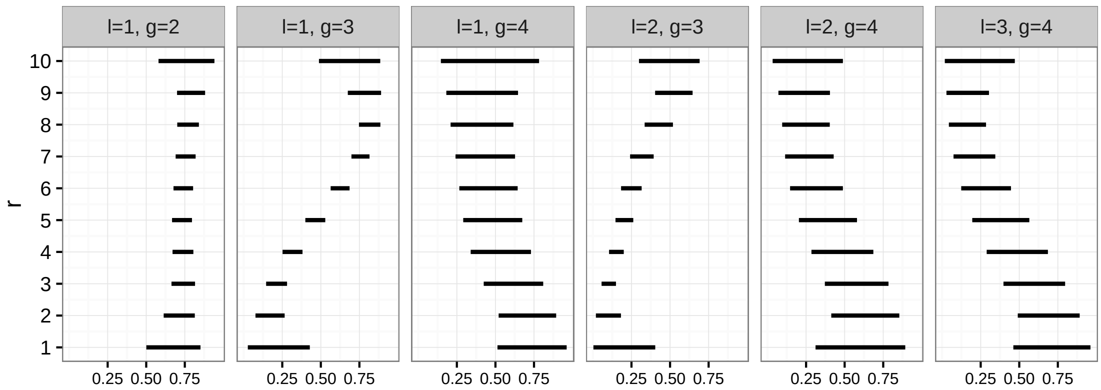

Figure 6: $[\underline{P}, \overline{P}](X^l_{(r)} > X^g_{(r)})$ for $m = 10$

To illustrate subset selection, we focus attention on subsets containing the best group, for which the NPI lower and upper probabilities are given by Equations (24) and (25). Similar illustrations for subsets containing all best groups are provided by Alqifari [2], who also investigates the effect of different values of m on such inferences. Table 5 and Figures 7 and 8 present the NPI lower and upper probabilities for any subset S, consisting of 2 of the 4 groups, to contain the group which provides the largest r-th ordered future observation out of m observations for each group, with $m = 5$ in Table 5, $m = 10$ in Figure 7 and $m = 100$ in Figure 8. Imprecision is again largest for extreme values of r and the values in this table and these figures illustrate the conjugacy relation $\underline{P}(A) = 1 - \overline{P}(A^c)$. Note that the NPI lower and upper probabilities for these events with subset S consisting of 3 of the 4 groups can be derived, again by the conjugacy relation, from the corresponding lower and upper probabilities for such events with S consisting of a single group, as presented in Table 3 for $m = 5$ and in Figure 4 for $m = 10$.

S:	$\{1,2\}$		$\{1,3\}$		$\{1,4\}$		$\{2,3\}$		$\{2,4\}$		$\{3,4\}$	
r	$\underline{P}^*_S$	$\overline{P}^*_S$	$\underline{P}^*_S$	$\overline{P}^*_S$	$\underline{P}^*_S$	$\overline{P}^*_S$	$\underline{P}^*_S$	$\overline{P}^*_S$	$\underline{P}^*_S$	$\overline{P}^*_S$	$\underline{P}^*_S$	$\overline{P}^*_S$
1	0.0893	0.3540	0.5389	0.8685	0.1949	0.5634	0.4366	0.8051	0.1315	0.4611	0.6460	0.9107
2	0.1768	0.3974	0.3483	0.6919	0.4510	0.7642	0.2358	0.5490	0.3081	0.6517	0.6026	0.8232
3	0.2437	0.5340	0.2599	0.5965	0.6639	0.8951	0.1049	0.3361	0.4035	0.7401	0.4660	0.7563
4	0.2282	0.6108	0.2085	0.5814	0.7074	0.9395	0.0605	0.2926	0.4186	0.7915	0.3892	0.7718
5	0.1390	0.7175	0.1477	0.7370	0.5193	0.9758	0.0242	0.4807	0.2630	0.8523	0.2825	0.8610

Table 5: $[\underline{P}^*_S, \overline{P}^*_S](\max_{l \in S} X^l_{(r)} > \max_{g \in NS} X^g_{(r)})$ for $m = 5$

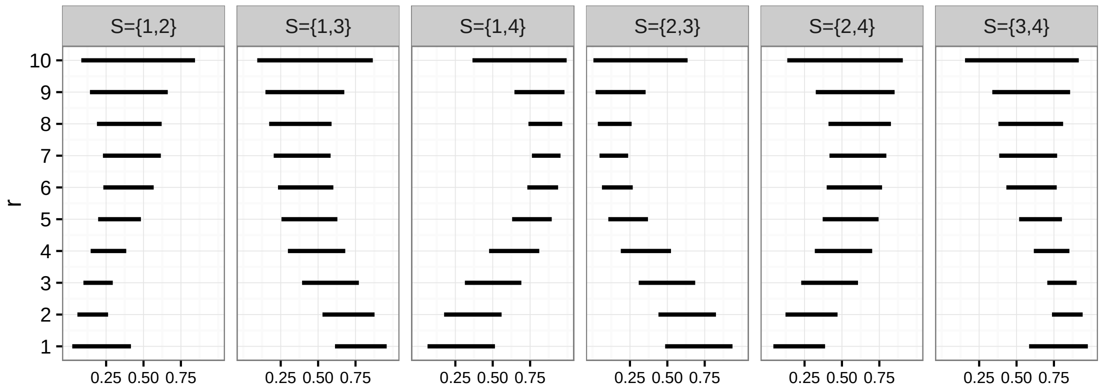

Figure 7: $[\underline{P}^*_S, \overline{P}^*_S](\max_S X^l_{(r)} > \max_{NS} X^g_{(r)})$ for $m = 10$

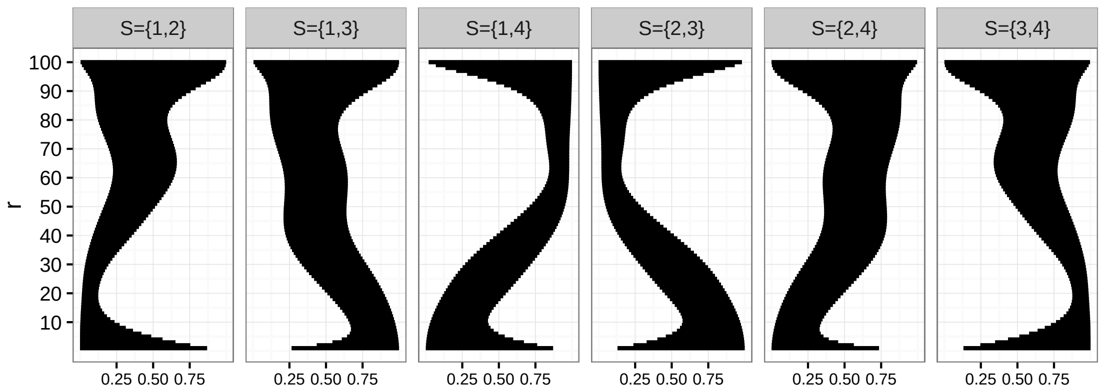

Figure 8: $[\underline{P}^*_S, \overline{P}^*_S](\max_S X^l_{(r)} > \max_{NS} X^g_{(r)})$ for $m = 100$

These NPI lower and upper probabilities can be used in a variety of ways. For example, one may be interested in a subset of smallest size such that the lower probability of it containing the best subset in terms of a specific r-th ordered future observation exceeds a specific value. The flexibility of the NPI approach provides methods for a wide range of problem formulations, for which derivations of the lower and upper probabilities always follow the same basic steps.

3.3 Further inferences

The NPI methods for future order statistics presented in this paper enable a wide range of further statistical inferences, as long as problems of interest are formulated in terms of such future order statistics. For example, one may be interested in prediction intervals [17], e.g. outer prediction intervals can be derived as the interval between two of the first n observations (or possibly with $-\infty$ or ∞ as end points), say (x_a, x_b) with $a < b$, such that this interval contains the predictive interval $[X_{(r)}, X_{(s)}]$ for $r < s$. The corresponding predictive probability of interest, which is easily computed using Equation (4), is

$$P(x_a < X_{(r)} < X_{(s)} < x_b) = \sum_{j=a+1}^{b} \sum_{l=j}^{b} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \tag{26}$$

One may also be interested in a corresponding inner prediction interval (x_c, x_d) which is contained in $[X_{(r)}, X_{(s)}]$, the corresponding predictive probability is

$$P(X_{(r)} < x_c < x_d < X_{(s)}) = \sum_{j=1}^{c} \sum_{l=d+1}^{n+1} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \tag{27}$$

One may typically be interested in the shortest outer interval, or the longest inner interval, for which the corresponding probability (26) or (27) exceeds a chosen threshold value, for given r and s. Of course, one may also just want to use these probabilities directly for inferences on $X_{(r)}$ and $X_{(s)}$. The idea of such outer and inner prediction intervals is used by Ahmadi et al [1] for intervals between future records.

One may also be interested in the number of future observations in an interval between two data observations. Let $C^m_{a,b} = u$ denote the event that exactly u out of m future observations are in the interval (x_a, x_b), with $1 \leq a < b \leq n$ and $1 \leq u \leq m$. The NPI probability for this event is equal to

$$P(C^m_{a,b} = u) = \sum_{m_a=0}^{m-u} \frac{\binom{a-1+m_a}{m_a}\binom{b-a-1+u}{u}\binom{n-b+m-u-m_a}{m-u-m_a}}{\binom{n+m}{n}} \tag{28}$$

This probability only depends on the number of intervals in the partition of the real line created by the data between x_a and x_b, hence only on the value $b-a$. An alternative expression for this NPI probability is

$$P(C^m_{a,b} = u) = \frac{\binom{n+a-b+m-u}{m-u}\binom{b-a-1+u}{u}}{\binom{n+m}{n}}$$

Both these expressions are easily derived by combinatorics using the basic probability results presented in this paper. For the special case with $b = n+1$, so considering the interval (x_a, ∞), we have

$$P(C^m_{a,n+1} = u) = \frac{\binom{a-1+m-u}{m-u}\binom{n-a+u}{u}}{\binom{n+m}{n}} \tag{29}$$

This result is equal to the distribution of the number of exceedances in the classical theory of statistics, although the derivation method differs due to the different starting points of NPI and the classical theory [3].

Spacings between order statistics have also attracted interest [3, p.32], the NPI approach enables consideration of spacings between future order statistics. Let $W_{r,s} = X_{(s)} - X_{(r)}$ for $1 \leq r < s \leq m$. We can use the joint probabilities given in Equation (4), for the event that

$X_{(r)} \in I_j = (x_{j-1}, x_j)$ and $X_{(s)} \in I_l = (x_{l-1}, x_l)$, for $j \leq l$, for inferences on $W_{r,s}$, which will mostly be in the form of lower and upper probabilities. For example, for the event $W_{r,s} < T$ for some $T > 0$, the NPI lower and upper probabilities are derived by

$$\underline{P}(W_{r,s} < T) = \sum_{j=1}^{n+1} \sum_{l=j}^{n+1} \mathbf{1}\{x_l - x_{j-1} < T\} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \quad (30)$$

$$\overline{P}(W_{r,s} < T) = \sum_{j=1}^{n+1} \sum_{l=j}^{n+1} \mathbf{1}\{x_{l-1} - x_j < T\} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \quad (31)$$

The NPI lower probability (30) is derived by summing up the probabilities for events $X_{(r)} \in I_j,\ X_{(s)} \in I_l)$ for which $X_{(s)} - X_{(r)} < T$ necessarily holds while the corresponding NPI upper probability (31) is derived by summing up the probabilities for all such events values for which $X_{(s)} - X_{(r)} < T$ can possibly be true, given that the predictive probabilities based on the $A_{(n)}$ assumptions are only specified on the intervals between consecutive data observations without any further assumptions on the distributions of such probabilities within these intervals. A range of further inferences, together with illustrative examples, will be presented by Alqifari [2].

4 Concluding remarks

The results presented in this paper provide new tools for predictive inference on order statistics of future observations. While for some inferences these coincide with classical results on order statistics [3], the explicit use of the $A_{(n)}$ assumptions and restriction to m future observations, make derivation of some results more straightforward than in the classical framework, where typically both the data observations and future observations are considered to be random quantities, sampled from an unknown population probability distribution, with predictive inference arrived at through conditioning on the data observations. The use of lower and upper probabilities widens the range of possible inferences compared to the classical approach. Several inferences are illustrated in this paper, in particular on multiple comparisons; the main ideas are similar for other inferences as long as these are explicitly expressed in terms of one or more future order statistics. Alqifari [2] presents NPI methods for a wider range of such inferences, and also investigates the influence of the particular choice of the number m of future observations.

A major research challenge is the generalization of NPI for future order statistics in case of lifetime data containing right-censored observations [23], which will enable such methods to be created for many applications in medical and engineering applications, where e.g. multiple comparisons methods are often applied. The NPI approach has been presented for right-censored data, leading to predictive lower and upper survival functions that bound the well-known Kaplan-Meier estimate [13], and related results for multiple comparisons have also been presented [14], however these were explicitly in terms of only a single future observation. The development of NPI for multiple future observations and for future order statistics, based on right-censored data, is a challenging topic for future research.

Appendix

Proof. **Proof of Theorem 1**
We must show that, for all $j \leq l-1$

$$\sum_{k=1}^{j} P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \geq \sum_{k=1}^{j} p(X_{(r)} \in I_k | X_{(d)} \in I_l) \tag{32}$$

Note that the conditional CDF for the event $X_{(r)} \in I_{l-1} | X_{(d)} \in I_{l-1}$ is

$$F_{r|d}(l-1|l-1) = \sum_{k=1}^{l-1} p(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) = 1$$

and the conditional CDF for the event $X_{(r)} \in I_{l-1} | X_{(d)} \in I_l$ is

$$F_{r|d}(l-1|l) = \sum_{k=1}^{l-1} p(X_{(r)} \in I_k | X_{(d)} \in I_l) < 1$$

A sufficient condition for property (32) to hold is if there exists one value w_r such that

$$P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \geq P(X_{(r)} \in I_k | X_{(d)} \in I_l) \quad \text{for all} \quad k \leq w_r \tag{33}$$

and

$$P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \leq P(X_{(r)} \in I_k | X_{(d)} \in I_l) \quad \text{for all} \quad k > w_r \tag{34}$$

Using Equation (7), it is straightforward to show that Equation (33) holds if and only if

$$k \leq \frac{r(l-1)}{(d-1)} + 1 \tag{35}$$

Similarly, Equation (34) holds if and only if

$$k \geq \frac{r(l-1)}{(d-1)} + 1 \tag{36}$$

Hence, by defining $w_r = \frac{r(l-1)}{(d-1)} + 1$ the sufficient condition holds and the proof is complete. □

Proof. **Proof of Theorem 2**
These NPI lower and upper probabilities are, as always, obtained by putting the probability

masses per interval at end points in order to minimize or maximize the probability for the event of interest, given the joint probabilities for the order statistics based on the $A_{(n)}$ assumptions. The lower probability given by Equation (17) is derived by summing up the joint probabilities for the events $X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x}$ and $Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{l_y}$ for which $x_{j_x} < y_{j_y-1}, x_{l_x-1} > y_{l_y}$ and $x_{l_x-1} \geq x_{j_x}$ hold. This follows from putting the probability masses for $X_{(r)}$ and $Y_{(s)}$ to the right end points of their respective intervals, and for $X_{(s)}$ and $Y_{(r)}$ to the left end points of their respective intervals. For the case where $X_{(r)}$ and $X_{(s)}$ belong to the same interval, we can achieve a lower probability of zero for the event that both $Y_{(r)}$ and $Y_{(s)}$ are between these two ordered future X observations, due to the fact that the $A_{(n)}$ assumptions do not imply any assumptions on the distribution of such probability masses within an interval between two consecutive data observations.

The NPI upper probability given by Equation (18) is derived similarly, by putting the probability masses for all 4 ordered future observations at the opposite end points of the intervals compared to the derivation of the lower probability, as explained above. However, for the upper probability the case where $Y_{(r)}$ and $Y_{(s)}$ belong to the same interval must be taken into account, this leads to the additional term in Equation (18), which actually involves maximisation of the probability given in Equation (19). In this case, $Y_{(r)}$ and $Y_{(s)}$ can be assumed to be extremely close to each other, effectively both equal to a value $y^*_{j_y} \in I^y_{j_y}$. This is possible due to the flexibility of placing the respective probability masses at any convenient point within the data intervals, note that now we do not just put these probability masses at end points of the interval. The remaining task is to maximize the term in Equation (19) with regard to $y^*_{j_y} \in I^y_{j_y}$, with this term dependent on whether or not any X group data observations are within the interval $I^y_{j_y}$. If there are no such X observations, then one can just put the Y probability mass in this interval at either of its end-points. However, if there are X observations in the interval $I^y_{j_y}$, then these partition this interval and we must calculate the term in Equation (19) for $y^*_{j_y}$ in each of the sub-intervals of this partition, and finally take the maximum over these values. Clearly, while this is slightly awkward since there is no closed-form expression for this upper probability, it is a straightforward algorithm which takes little computational effort due to the limited number of X values in each Y interval.

□

Proof. **Proof of the lower and upper probabilities** (22) **and** (23)
We present the derivation of the NPI lower probability (22) for the event $\min_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}$, the corresponding NPI upper probability (23) is derived by similar arguments, just replacing lower bounds by upper bounds (it is presented in detail by Alqifari [2]). To derive the NPI lower probability, we derive a lower bound for the probability for this event, based on the $A_{(n)}$ assumptions for all groups. This lower bound can actually be achieved by putting the probability masses for the future order statistics per group at end points of the respective intervals between consecutive data observations, with probability masses for groups in the selected subset S put at left end points and for groups not in S at right end points. Hence, we derive the maximum

lower bound for this probability, which therefore is the NPI lower probability. The steps to derive this lower bounds are as follows

$$
\begin{aligned}
P\left(\min_{l\in S} X^{l}_{(r)} > \max_{g\in NS} X^{g}_{(r)}\right) &= P\left(\bigcap_{g\in NS}\{X^{g}_{(r)} < \min_{l\in S} X^{l}_{(r)}\}\right) \\
&= \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} P\Bigg(\bigcap_{g\in NS}\{X^{g}_{(r)} < \min_{l\in S} X^{l}_{(r)}\} | X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}\Bigg) \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \\
&\geq \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \left[P\left(\bigcap_{g\in NS}\{X^{g}_{(r)} < \min_{l\in S} x^{l_s}_{j_{l_s}-1}\}\right)\right] \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \\
&\geq \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \left[\prod_{g\in NS} P\left(X^{g}_{(r)} < \min_{l\in S} x^{l_s}_{j_{l_s}-1}\right)\right] \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \\
&\geq \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \left[\prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^{g}_{j_g} < \min_{l\in S} x^{l_s}_{j_{l_s}-1}\} P(X^{g}_{(r)} \in I^{g}_{j_g})\right] \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}})
\end{aligned}
$$

$\square$

Acknowledgements

The authors are grateful to two anonymous reviewers whose supportive comments led to improved presentation of the paper.

References

[1] Ahmadi, J., MirMostafaee, S.M.T.K. and Balakrishnan, N. (2010). Nonparametric prediction intervals for future record intervals based on order statistics. *Statistics and Probability Letters*, **80**, 1663-1672.

[2] Alqifari, H.N. *Nonparametric predictive inference for future order statistics*. PhD Thesis, Durham University, in preparation.

[3] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics*. Wiley, New York.

[4] Arts, G.R.J. and Coolen, F.P.A. (2008). Two nonparametric predictive control charts. *Journal of Statistical Theory and Practice*, **2**, 499-512.

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

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

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

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

[9] Coolen, F.P.A. and Al-nefaiee, A.H. (2012). Nonparametric predictive inference for failure times of systems with exchangeable components. *Journal of Risk and Reliability*, **226**, 262-273.

[10] Coolen, F.P.A. and Bin Himd, S. (2014). Nonparametric predictive inference for reproducibility of basic nonparametric tests. *Journal of Statistical Theory and Practice*, **8**, 591-618.

[11] Coolen, F.P.A. and Maturi, T.A. (2010). Nonparametric predictive inference for order statistics of future observations. In: *Combining Soft Computing and Statistical Methods in Data Analysis.* C. Borgelt, *et al.* (Eds). Springer, Berlin (Advances in Intelligent and Soft Computing 77), pp. 97-104.

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

[13] Coolen, F.P.A. and Yan, K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, **126**, 25-54.

[14] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2012). Nonparametric predictive multiple comparisons of lifetime data. *Communications in Statistics - Theory and Methods*, **41**, 4164-4181.

[15] De Finetti, B. (1974). *Theory of Probability* (2 volumes). Wiley, London.

[16] Desu, M.M. and Raghavarao, D. (2004). *Nonparametric Statistical Methods for Complete and Censored Data.* Chapman and Hall, Boca Raton

[17] Frey, J. (2014). Shorter nonparametric prediction intervals for an order statistic from a future sample. *Statistics & Probability Letters*, **91**, 69-75.

[18] Geisser, S. (1993). *Predictive Inference: an Introduction.* Chapman and Hall, London.

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

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

[21] Sultan, K.S. and Alshami, S.A. (2010). Prediction of future order statistics from the uniform distribution. In: *Combining Soft Computing and Statistical Methods in Data Analysis*, C. Borgelt, G. Gonzalez-Rodriguez, W. Trutschnig, M. Asuncion Lubiano, M. Angeles Gil, P. Grzegorzewski, O. Hryniewicz (Eds). Springer, Berlin, pp. 593-602.

[22] Verheijen, J.H.M., Coolen, F.P.A. and van der Laan, P. (1997). Combining two classical approaches for statistical selection. *Communications in Statistics: Theory and Methods*, **26**, 1291-1312.

[23] Volterman, W., Davies, K.F., Balakrishnan, N. and Ahmadi, J. (2014). Nonparametric prediction of future order statistics. *Journal of Statistical Computation and Simulation*, **84**, 683-695.

[24] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

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

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