Nonparametric predictive inference for future order statistics

Hana N. Alqifari

A Thesis presented for the degree of
Doctor of Philosophy

Statistics and Probability
Department of Mathematical Sciences
Durham University
United Kingdom

August 2017

Nonparametric predictive inference for future order statistics

Hana N. Alqifari

Submitted for the degree of Doctor of Philosophy

August 2017

Abstract

Nonparametric predictive inference (NPI) has been developed for a range of data types, and for a variety of applications and problems in statistics. In this thesis, further theory will be developed on NPI for multiple future observations, with attention to order statistics. The present thesis consists of three main, related contributions. First, new probabilistic theory is presented on NPI for future order statistics; additionally a range of novel statistical inferences using this new theory is discussed. Secondly, NPI for reproducibility is developed by considering two statistical tests based on order statistics. Thirdly, robustness of NPI is introduced which involves a first adaptation of some of the robustness theory concepts within the NPI setting, namely sensitivity curve and breakdown point.

In this thesis, we present NPI 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. We further present the use of such predictive probabilities for order statistics in statistical inference, in particular considering pairwise and multiple comparisons based on future order statistics of two or more independent groups of data.

This new theory enables us to develop NPI for the reproducibility of statistical hypothesis tests based on order statistics. Reproducibility of statistical hypothesis tests is an important issue in applied statistics: if the test were repeated, would the same conclusion be reached that is rejection or non-rejection of the null hypothesis? NPI provides a natural framework for such inferences, as its explicitly predictive nature fits well with the core problem formulation of a repeat of the test in the future. For inference on reproducibility of statistical tests, NPI provides lower and upper reproducibility probabilities (RP). The NPI-RP method is presented for two basic tests using order statistics, namely a test for a specific value for a population quantile and a precedence test for comparison of data from two populations, as typically used for experiments involving lifetime data if one wishes to conclude before all observations are available.

As every statistical inference has underlying assumptions about models and specific methods used, one important field in statistics is the study of robustness of inferences. The concept of robust inference is usually aimed at development of inference methods which are not too sensitive to data contamination or to deviations from model assumptions. In this thesis we use it in a slightly narrower sense. For our aims, robustness indicates insensitivity to small changes in the data, as our predictive probabilities for order statistics and statistical inferences involving future observations depend upon the given observations. We introduce some concepts for assessing the robustness of statistical procedures in the NPI framework, namely sensitivity curve and breakdown point. The classical breakdown point does not apply to our context as the predictive inferences are bounded, so we change the definition to suit our context. Most of our nonparametric inferences have a reasonably good robustness with regard to small changes in the data. Traditionally, in the robustness literature there has been quite a lot of emphasis and discussion on robustness properties of estimators for the location parameters. Thus, in our investigation of NPI robustness we also focus on differences in robustness of the mean and the median of the m future observations, and see how they relate to the classical concepts of robustness of the median and mean.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification.

Acknowledgements

Alhamdullilah, be grateful to Allah my God, for his blessing on me generally and in accomplishing this thesis especially.

I would like to express my sincere gratitude to my supervisor Prof. Frank Coolen for the continuous support of my Ph.D study and research, for his motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D study.

Many thanks to the Durham University. I learned a lot from the seminars organised and the training courses.

I would like to thank the sponsors and supporters of my study, Saudi Arabian Cultural Bureau in London, the Ministry of Higher Education in Saudi Arabia and Qassim University.

I am truly grateful to my husband Abdullah Alabdulatif for his belief in me and unlimited support. To my two lovely children, Abdulrahman and Mela, who are the pride and joy of my life.

Special thanks to my father who has been a great source of motivation and support and my mother for all her unlimited love and prayers and all my brothers and sisters for encouraging and believing.

Contents

Chapter 1

Introduction

Order statistics appear in many statistical problems and are widely applied in statistical modelling and inference [64]. Thus, a recognizable large body of literature has been developed to study order statistics [3, 36, 49]. In this thesis, we present nonparametric predictive inference (NPI) for order statistics of future observations.

NPI [7, 21] is a statistical framework based on the assumption $A_{(n)}$ by Hill [55], as introduced in Section 1.2, with inferences explicitly in terms of future observations. NPI has been developed for a range of data types, and for a variety of applications and problems in statistics, risk and reliability and operations research. NPI for real-valued random quantities has thus far been mostly restricted to a single future observation, although multiple future observations have been considered for NPI methods for statistical process control [4, 5] and system reliability [22]. In this thesis, further theory is developed on NPI for multiple future observations with attention to order statistics. The nature of this thesis is mostly theoretical with the implementation of the developed methods illustrated by example applications using data from the literature.

This thesis presents a new probabilistic theory of NPI for future order statistics that takes into account their interdependency. The results from these new predictive probabilities involving order statistics of m future observations enable the development of new NPI methods for statistical inference. We present pairwise and multiple comparisons based on future order statistics of two or more independent groups of real-valued data. This generalizes NPI results for a single future observation [30]. The new probabilistic theory contributes to development of NPI for test reproducibility, we consider two tests based on order statistics. The reproducibility of statistical hypothesis tests is an issue of major importance in applied statistics: if a test were repeated, would the same conclusion be reached about rejection of the null hypothesis? NPI provides a natural framework for such inferences, as its explicitly predictive nature fits well with the core problem formulation of a repeat of the test in the future. Robustness of the NPI is considered in this thesis. We have introduced some of the concepts of classical robust statistics to the NPI method.

The outline of this introductory chapter is as follows. Section 1.1 presents classical order statistics, and related prediction intervals. In Section 1.2, we present an overview of the NPI framework and imprecise probability. In Section 1.3 we present NPI for comparing groups. A detailed outline of this thesis is given in Section 1.4.

1.1 Classical order statistics

Order statistics are important tools in nonparametric statistical inference. If $X_1, \ldots, X_n$ are random quantities, their order statistics are arranged in non-decreasing order and denoted by $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. Classical results dealing with order statistics often assume that the X_i are independent and identically distributed (iid) with continuous cumulative distribution function (CDF) $F(x)$ and probability density function (PDF) $f(x)$ [3], although we will not make this assumption. Order statistics appear in many practical cases, including robust location estimates, detection of outliers, goodness-of-fit tests, analysis of censored data and reliability analysis [3]. Order statistics deal with the properties and applications of ordered random quantities $X_{(.)}$ and the functions involving them. These statistics have many applications, such as extremes that arise in the statistical study of floods and droughts, or in the problems of breaking strength and fatigue failure [36].

In this thesis we consider nonparametric predictive inference for order statistics. This topic has also been presented from a different perspective, in the classical literature, which we briefly introduce now.

In classical order statistics, it is assumed that $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ represent data observations, and $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(m)}$ represent future observations, where the underlying X and Y populations are assumed to consist of iid random quantities with the same continuous CDF $F(x)$. A prediction interval $(X_{(j_1)}, X_{(j_2)})$, for $j_1, j_2 \in \{1, \ldots, n\}$ with $j_1 < j_2$ is an interval in which future observations will fall, with a certain probability. Major interest in prediction intervals was first mentioned in the context of exceedances, such as to answer the question what is the probability that $X_{(j)}$, the j-th in magnitude of n annual maximum flows at a point in a river, is exceeded u times, $u = 1, \ldots, m$, in m future years [36, 49]? If $R(j)$ is the number of future observations greater than $X_{(j)}$, then

$$\begin{aligned} P(R(j) = u) &= n\binom{n-1}{j-1}\int_0^1 (F(x))^{j-1}(1-F(x))^{n-j}\binom{m}{u}(1-F(x))^u(F(x))^{m-u}dF(x) \\ &= n\binom{n-1}{j-1}\binom{m}{u}\int_0^1 (F(x))^{j-1+m-u}(1-F(x))^{n-j+u}dF(x) \end{aligned}$$

Thus, u is the number of exceedances and is a positive integer where $0 \leq u \leq m$. This distribution of exceedances depends on the variate u and the parameters n, j and m but not on the CDF $F(x)$. The exceedances probability $P(R(j) = u)$ is;

$$P(R(j) = u) = \frac{\binom{n-j+u}{u}\binom{m-u+j-1}{j-1}}{\binom{m+n}{n}} \tag{1.1}$$

This result also provides the probability that at least u Y's fall between $X_{(j_1)}$ and $X_{(j_2)}$ $(1 \leq j_1 < j_2 \leq n)$. If $C(j_1, j_2)$ is the number of future observations in $X_{(j_1)}, X_{(j_2)}$, then $P(C(j_1, j_2) = u)$ is derived by replacing j by $n + 1 + j_1 - j_2$ in Equation (1.1), [36], to give

$$P(C(j_1, j_2) = u) = \frac{\binom{j_2-j_1-1+u}{u}\binom{m-u+n+j_1-j_2}{m-u}}{\binom{m+n}{n}} \tag{1.2}$$

Note here that, in the classical approach, $X_{(j_1)}$ and $X_{(j_2)}$ are also considered to be random quantities.

Recent work by Volterman, et al. [83] considered nonparametric prediction of a subset of order statistics that was based on one or more independent type-II right censored samples. They were interested in predicting the unobserved failures from a type-II censored sample. They assumed two independent type-II right censored samples $X_{1:m}, \ldots, X_{r:m}$, and $Y_{1:n} < \ldots < Y_{d:n}$ where $1 \leq r < m$ and $1 \leq d \leq n$. They assumed that the X and Y values come from the same absolutely continuous distribution with CDF F and PDF f. They considered a nonparametric interval for a single $X_{(s)}$ $(r < s \leq m)$ and two order statistics $X_{(s_1)}$ and $X_{(s_2)}$ $(r < s_1 < s_2 \leq m)$ using the independent Y sample, where all prediction intervals must be conditional on $X_{(r)} < Y_{(k)}$ for any $1 \leq k \leq d$. For the missing s-th order statistic of group X, where $r < s \leq m$, the exceedance probability of two order statistics from independent samples, $X_{(s)} < Y_{(k)}$ is equal to

$$P(X_{(s)} < Y_{(k)}) = \sum_{i=0}^{k-1} \frac{\binom{s-1+i}{i}\binom{m-s+n-i}{n-i}}{\binom{n+m}{m}} \tag{1.3}$$

Volterman, et al. [83] have predicted the s-th order statistic of right-censored observations in a sample X using the information provided from the independent sample Y, where all the prediction intervals must be conditional on $X_{r:m} < Y_{k:n}$ for any $1 \leq k \leq n$. In the classical order statistics methods, they regard future order statistics on the sample X and the observations on group Y both as random quantities [36, 49].

For two sided intervals, the probability of $X_{(r)} < Y_{(a)} < X_{(s_1)} < X_{(s_2)} < Y_{(b)}$ [83] is equal to

$$P(X_{(r)} < Y_{(a)} < X_{(s_1)} < X_{(s_2)} < Y_{(b)}) = \sum_{i_1=r}^{s_1-1} \sum_{i_2=0}^{m-s_2} \frac{\binom{a-1+i_1}{i_1}\binom{m-i_1-i_2+b-a-1}{m-i_2-i_1}\binom{n-b+i_2}{i_2}}{\binom{m+n}{m}} \tag{1.4}$$

1.2 Nonparametric predictive inference

Hill [55, 56] proposed the assumption $A_{(n)}$ for prediction of a future observation. This assumption is suitable for situations in which there is extremely vague prior knowledge about the characteristics of the underlying distribution of the observations or, which may be more realistic, in situations in which one explicitly does not want to use such information. This may occur, for example, if one wants to study the (often hidden) effects of additional structural assumptions underlying statistical models or methods. Inferences based on such restricted knowledge have also been called 'low structure inferences' [45].

To introduce the assumption $A_{(n)}$, we first need to intrduce some notation. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are real-valued exchangeable random quantities. Let the ordered observed values of $X_1, \ldots, X_n$ be denoted by $x_1 < x_2 < \ldots < x_n$. For ease of notation, let $x_0 = -\infty$ and $x_{n+1} = \infty$ (or when dealing with non-negative random quantities, we set $x_0 = 0$) [58]. Note that x_{n+1} does not denote an observed value for X_{n+1}. These n observations divide the real-line into $n+1$ intervals $I_j = (x_{j-1}, x_j)$, where $j = 1, \ldots, n+1$. For a future observation X_{n+1}, the assumption $A_{(n)}$ is

$$P(X_{n+1} \in I_j) = \frac{1}{n+1} \quad \text{for each } j = 1, \ldots, n+1 \tag{1.5}$$

It is clear that $A_{(n)}$ is a post-data assumption related to exchangeability [39]. A natural interpretation of the assumption $A_{(n)}$ is that it is conditional on $X_1, \ldots, X_n$ and, thus, the next observation X_{n+1} is equally likely to fall in any of the open intervals between the sequential order statistics of the given sample [57]. If one wishes to allow ties [56], the probabilities $\frac{1}{n+1}$ can be assigned to the closed intervals $[x_{j-1}, x_j]$ instead of the open intervals I_j. Alternatively, ties can also be dealt with by assuming that such observations differ by a very small amount, a common method to break ties in statistics [56]. To keep presentation simple we will assume throughout this thesis that ties do not occur.

The assumption $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest. However, it does provide bounds for probabilities by what is essentially an application of De Finetti's fundamental theorem of probability [39] or Walley's concept of natural extension [84]. Weichselberger [87] developed a formal foundation of interval probability, which is also known as imprecise probability, in the spirit of Kolmogorov's axioms. $A_{(n)}$ can provide predictive probability bounds for a future observation without assuming any further assumptions about the probability distribution. These bounds are lower and upper probabilities in imprecise probability theory [7, 8]. The lower and upper probabilities for event A are denoted by $\underline{P}(A)$ and $\overline{P}(A)$, respectively [21], and can have several interpretations. From a subjective perspective [84], the lower and upper probabilities for an event can be interpreted as supremum buying price and infimum selling price, respectively, for a gamble that pays one if the event occurs and zero if the event does not occur. From a classical perspective, lower and upper probabilities can be interpreted as bounds on precise probabilities because of the constrained or limited knowledge available, or because one wishes not to add further assumptions. Throughout this thesis, and in the NPI approach in general, the classical interpretation is used [21]. The NPI lower and upper probabilities are the maximum lower bound and minimum upper bound for the event of interest, based on the $A_{(n)}$ assumption. For example, the NPI lower and upper probabilities for the event that $X_{n+1} \in B$, given the past data $x_1, \ldots, x_n$, where $B \subset \mathbb{R}$, are

$$\underline{P}(X_{n+1} \in B) = \sum_{j=1}^{n+1} 1\{I_j \subseteq B\} P(X_{n+1} \in I_j) \tag{1.6}$$

$$\overline{P}(X_{n+1} \in B) = \sum_{j=1}^{n+1} 1\{I_j \cap B \neq \emptyset\} P(X_{n+1} \in I_j) \tag{1.7}$$

where $\mathbf{1}\{E\}$ is an indicator function which is equal to 1 if event E occurs and 0 else. The lower probability (1.6) is derived by summing only the $A_{(n)}$-based probabilities that must be in interval B; similarly, the upper probability (1.7) is obtained by summing all the $A_{(n)}$-based probabilities that can be in B.

The theory of imprecise probability [7, 84, 85, 86, 87] makes it clear that bounds provide valuable information on uncertainty of events caused by restricted information. For an event A, the precise classical probability is just a special case of imprecise probability, when $\underline{P}(A) = \overline{P}(A)$ (i.e. the point probability case). However, the case in which $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$ represents a complete lack of information about the event A (the vacuous case). We briefly present some key aspects of the theory of imprecise probability as relevant to $A_{(n)}$-based inference [7]. Generally, in imprecise probability theory, $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, the lower and upper probabilities are conjugated, i.e. $\underline{P}(A) = 1 - \overline{P}(A^c)$, where A^c is the complementary event of A, and $\underline{P}(.)$ is super-additive and $\overline{P}(.)$ is sub-additive, i.e. for events A and B such that $A \cap B = \emptyset$:

$$\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$$
$$\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B)$$

These properties hold for the NPI lower and upper probabilities, as proven by Coolen and Augustin [7].

NPI has been developed for a range of data types and a variety of applications. For example, NPI has been presented for Bernoulli data [20, 26], real-valued data [28, 30, 70], data including right-censored observations [31, 32], circular data [21], multinomial data [9, 24] and bivariate data [33].

NPI for such inferences requires an exchangeability assumption for the observable random quantities. This assumption can also be applied in less straightforward situations [21]. For example, in case a data set contains right-censored observations, Coolen and Yan [31, 32] developed a generalization of $A_{(n)}$, called 'right-censoring-$A_{(n)}$', and they illustrated its use in inferential problems, mostly by the corresponding lower and upper survival functions. NPI for Bernoulli random quantities [20, 26] is based on a latent variable representation of Bernoulli data as real-valued outcomes of an experiment in which there is a completely unknown threshold value, such that outcomes to one side of the threshold are successes and to the other side failures. The use of $A_{(n)}$ provides lower and upper probabilities for the number of successes in m future trials, based on the number of successes in n observed trials. NPI for the next observation in case of multinomial data, with complete absence of knowledge on the number of possible categories apart from the information provided by $n > 0$ observations, is presented by Coolen and Augustin [24].

1.3 NPI for comparing groups

The NPI approach has been developed for a variety of applications, such as comparing random quantities that correspond to two or more groups. In classical statistics comparisons of two or multiple independent groups tend to be formulated as tests of hypothesis. However, NPI comparisons are explicitly predictive so one or more future observations for each group can be compared. Throughout this thesis, and in NPI in general [19, 30], we use the word 'quantity' instead of 'variable' following De Finetti [56], and the term 'group' instead of the more common term 'population' since the common view for observations from a single population is that these are independent and identically distributed (iid), which is an assumption that we do not make in the NPI framework. In Chapter 3, we show how pairwise and multiple comparisons can be based on future order statistics as presented in Chapter 2. Note that, in this thesis we use the term 'multiple comparison' for the comparisons of multiple groups, in line with the use of this term by Bechhofer [13].

The comparison of two groups of real-valued data within the NPI method was proposed by Coolen [19], considering one future observation for each group. Similar inferences for lifetime data, including right-censored observations, were presented by Coolen and Yan [32], and Maturi [69] introduced NPI for comparison of multiple groups of data including right-censored observations. Further work by Maturi, Coolen-Schrijner and Coolen [70] presented NPI for comparison of two groups of real-valued data where the tails of the data were possibly terminated, leading to small values being left-censored and large values being right-censored. As an alternative to classical nonparametric precedence tests considering a hypothesis of equal lifetime distributions [10, 43, 73], Coolen-Schrijner, Maturi and Coolen [35] proposed NPI precedence testing for two groups.

In many applications of statistics, the purpose is to compare multiple groups of data or to select the best group. Different methods that are designed particularly for the problem of selecting the best treatment or the optimal member of some group are known as selection procedures [47]. Selecting one or more best groups out of multiple groups under consideration, is a problem that often occurs in practice. For example, in terms of multinomial data, the researcher might be interested in selecting the category that has the largest probability of occurrence [12, 13]. Selection procedures have wide applications in many fields [47]. The first application of selection procedures was reported in the area of poultry science, particularly about poultry stocks offered for sale, to breeders, hatcherymen and potential buyers of chicks. The problem of selecting was first considered by Becker [14] to select the best chicken stock which has the largest hen-house egg production. Further, in drug development, a researcher might wish to select the drug that provides the best response. In advertising applications, specifically in relation to the problem of selecting a medium for advertising some product, the decision maker might want to select the most efficient media which reaches the highest proportion of potential purchasers of that product [47]. Most of the existing methods for selection are based on hypothesis testing and do not consider predictive inference [88]. Selection methods based on NPI use predictive inferences based on past observations and make use of Hill's assumption $A_{(n)}$. Such methods have been developed by Coolen and van der Laan [30] and Coolen and Coolen-Schrijner [26]. The NPI method for selection problems has been applied to many inferences, such as the selection between sources if observations are real-valued [30], selection of an optimal group of Bernoulli data [26, 27] and selection of a best category of multinomial data [9].

Coolen and van der Laan [30] proposed the NPI selection method for real-valued random quantities from different independent sources. Their aim was to select the source that provides the maximum next observation. This was done by comparing one future observation from each source. They considered two methods of selecting a subset of sources. The first method involved the selection of a subset of best sources, which is considered to be successful if the next observation from each source in the specific selected subset of sources is greater than the next observation from all sources that are not included in the subset. The second method involved the selection of a subset to include the best source, which is considered to be successful if the next observation from at least one of the selected sources is greater than the next observation from all sources that are not selected. In this thesis, we present similar NPI selection methods but based on the order statistics of multiple future observations.

1.4 Outline of the thesis

This thesis is organized as follows. In Chapter 2, we present probabilities for any collection of one or more future order statistics over the intervals I_j created by 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 Chapter 3 with main focus on pairwise and multiple comparisons. A paper presenting the results in Chapters 2 and 3 has been accepted for publication in Communications in Statistics - Theory and Methods [29]. These chapters have also been presented at several seminars and conferences, including the Northern Postgraduate Mini-Conference in Statistics in June 2015 in Durham, a lecture event on Nonparametric Predictive Inference in Durham in December 2015, and at the International Conference of the Royal Statistical Society at Manchester University in September 2016.

In Chapter 4, we develop NPI for reproducibility of two basic statistical tests based on order statistics, namely a population quantile test and a basic precedence test. A paper based on this chapter has been accepted for publication in REVSTAT [23]. This chapter has been presented at a research seminar in Durham in March 2017 and at the Research Students' Conference in Probability and Statistics in Durham in April 2017.

Robustness of NPI is considered in Chapter 5. This involves adapting some of the concepts of classical robust statistics within the NPI setting. A journal paper on this research topic is in preparation.

There are many interesting opportunities to develop and extend the research presented in this thesis. Some of these are mentioned in the final sections of Chapters 2 to 5 and in Chapter 6.

Chapter 2

NPI for future order statistics

2.1 Introduction

The development of NPI for real-valued random quantities has thus far mostly been restricted to a single future observation, although multiple future observations have been considered for NPI methods for statistical process control [4, 5] and system reliability [22]. In this thesis, we present several new contributions to the theory of NPI for future order statistics. We consider m future real-valued observations, given n data observations, and we focus on the order statistics of these m future observations.

This chapter derives the core probability results of NPI for future order statistics. These results will enable statistical inference involving order statistics for m future observations, presented in Chapters 3 and 4, as well as the development of NPI methods for a range of problems in probability, statistics and related topic areas, which will be explored in future research. We present the joint probability distribution for any collection of future order statistics over the intervals created by the partition of the real-line formed by 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 [8, 84] and interval probability [86, 87].

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 thesis that ties do not occur. We are interested in $m \geq 1$ future observations, X_{n+i} for $i = 1, \ldots, m$. It should be emphasised that the future observations X_{n+i} are assumed to come from the same data collecting process as the n data observations. Note that 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 $A_{(n)}$ [55], or more precisely, via consecutive application of $A_{(n)}, A_{(n+1)}, \ldots, A_{(n+m-1)}$. We refer to these generically as the $A_{(.)}$ assumptions, which can be considered as a post-data version of a finite exchangeability assumption for $n+m$ random quantities. The $A_{(.)}$ assumptions imply that all possible orderings of n data observations and 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 the random quantity S_j^i be defined as the number of m future observations in $I_j = (x_{j-1}, x_j)$ given a specific ordering, which is denoted by O_i, of the m future observations among n data observations, for $i = 1, \ldots, \binom{n+m}{n}$, so that $S_j^i = \#\{X_{n+l} \in I_j,\ l = 1, \ldots, m\}$. Then the $A_{(.)}$ assumptions lead to [28]

$$P\left(\bigcap_{j=1}^{n+1} \{S_j^i = s_j^i\}\right) = P(O_i) = \binom{n+m}{n}^{-1} \tag{2.1}$$

where s_j^i are non-negative integers with $\sum_{j=1}^{n+1} s_j^i = m$. Equation (2.1) implies that all $\binom{n+m}{n}$ orderings O_i of the m future observations among the n data observations are equally likely. Another convenient way to interpret the $A_{(.)}$ 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, which allow one to make inferences about the m unrevealed observations. The $A_{(.)}$ 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, Equation (2.1) implies that we can count the number of such orderings for which this event holds. Generally in NPI, as discussed in Section 1.2 the lower probability for the event of interest is derived by counting all the orderings for which this event has to hold, while the corresponding upper probability is derived by counting all the orderings for which this event can hold [7, 21].

NPI is close in nature to predictive inference for the low structure stochastic case as briefly outlined by Geisser [45], which is in line with many earlier non-parametric test methods where the interpretation of the inferences is in terms of confidence levels or intervals. In NPI the $A_{(.)}$ 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 [39], that corresponding optimal bounds can be derived for all events of interest [7]. NPI provides frequentist inferences which are exactly calibrated in the sence of Lawless and Fredette [66], and it has strong consistency properties in the theory of interval probability [7].

In NPI the n observations are explicitly used through the $A_{(.)}$ 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 $A_{(.)}$ 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 $A_{(.)}$ 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 [21].

This chapter is organised as follows. Section 2.2 provides an initial result which was presented in [28], which only includes the predictive probability for a single future order statistic. Section 2.2 provides NPI for the r-th future order statistic to be in any subset of the real values. In Section 2.3, we present the probability distributions for multiple future observations, while conditional probabilities given some future order statistics are derived in Section 2.4 along with some of the properties of these distributions. Section 2.5 presents NPI for the mean and median of the m future observations and some related inferences. Section 2.6 presents some concluding remarks.

2.2 NPI for the r-th ordered future observation

The probability distribution of a single order statistic of m future observations is useful for many inferences, e.g. pairwise and multiple comparisons which will be presented in Chapter 3. Let $X_{(r)}$, for $r = 1, 2, \ldots, m$, be the r-th ordered future observation, thus $X_{(r)} = X_{n+i}$ for one $i = 1, 2, \ldots, m$ and $X_{(1)} < X_{(2)} < \ldots\ldots < X_{(m)}$. In order to derive the probability distribution for the event that $X_{(r)}$ belongs to an interval $I_j = (x_{j-1}, x_j)$, we count the relevant orderings and use Equation (2.1). For $j = 1, \ldots, n+1$ and $r = 1, 2, \ldots, m$, this leads to

$$P(X_{(r)} \in I_j) = \frac{\binom{r+j-2}{j-1}\binom{n-j+1+m-r}{n-j+1}}{\binom{n+m}{n}} \tag{2.2}$$

The RHS of (2.2) can be derived as follows, and as illustrated in Figure 2.1. To the left of $X_{(r)}$, we count the number of ways of ordering $r-1$ future observations among $j-1$ past observations. This is multiplied by the number of ways of ordering the remaining $m-r$ future observations among $n-(j-1)$ past observations to the right of $X_{(r)}$. The denominator is the total number of ways of ordering m future observations among n past observations. For this event, NPI provides a precise probability, since 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.

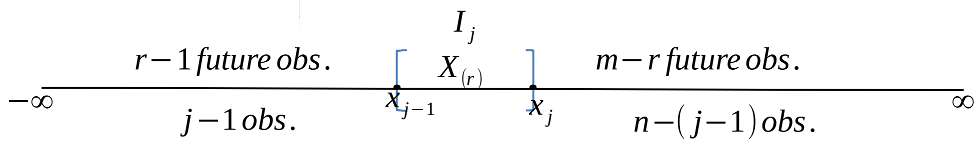

Figure 2.1: Locations of future and past observations corresponding to the event $X_{(r)} \in I_j$

As Equation (2.2) only specifies the probabilities for the events that $X_{(r)}$ belongs to interval 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 (2.2) 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 $\left(\frac{r-1}{m-1}\right)(n+1) \leq j^* \leq \left(\frac{r-1}{m-1}\right)(n+1)+1$. The proof of these properties is given in Appendix A. 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. 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. Although the special case $m = n$ is not explicitly considered further in this chapter, it plays an important role in the analysis of reproducibility of statistical hypothesis tests considered in Chapter 4.

It is worth commenting on extreme values, in particular, inference involving $X_{(1)}$ or $X_{(m)}$ for a large m 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 the lower bound for lifetime data.

This indicates that for these inferences, little can be concluded without making further assumptions on the probability masses within these end intervals, which are outside the range of observed data.

For an event $X_{(r)} \in I_j$, the $A_{(.)}$ assumptions provide precise probabilities. More generally, interest may be in an event $X_{(r)} \in S$, with S 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 [7, 8, 84, 86, 87]. This can be regarded as an application of De Finetti's 'Fundamental Theorem of Probability' [39]. Similar to Equation (1.6) and (1.7) in Section 1.2, for any subset S of the real values, we can derive the NPI lower probability

$$\underline{P}\left(X_{(r)} \in S\right) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \subseteq S\} P\left(X_{(r)} \in I_j\right) \tag{2.3}$$

and the corresponding NPI upper probability

$$\overline{P}\left(X_{(r)} \in S\right) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \cap S \neq \emptyset\} P\left(X_{(r)} \in I_j\right) \tag{2.4}$$

The probability (2.2) is important and we will use it in Chapter 3 for several statistical inferences, such as pairwise and multiple comparisons based on the r-th order statistic of m future observations for each group, and in Chapter 4 for reproducibility analysis of some basic nonparametric tests based on order statistics. The probability (2.2) and the NPI lower and upper probabilities (2.3) and (2.4) are briefly illustrated in the following example.

Example 2.1. Suppose that one has $n = 8$ observations and considers $m = 5$ future order statistics represented by the random quantities $X_{(1)} < \ldots < X_{(5)}$. The NPI probabilities for the events $X_{(r)} \in I_j$ for $r = 1, \ldots, 5$ and $j = 1, \ldots, 9$ are displayed in Table 2.1 and Figure 2.2.

Table 2.1 illustrates the obvious symmetry and unimodality. Each probability has only one local maximum which is indicated in Table 2.1 by $*$. Figure 2.2 shows

j	$P(X_{(1)} \in I_j)$	$P(X_{(2)} \in I_j)$	$P(X_{(3)} \in I_j)$	$P(X_{(4)} \in I_j)$	$P(X_{(5)} \in I_j)$
1	0.38462*	0.12821	0.03497	0.00699	0.00078
2	0.25641	0.18648	0.08392	0.02486	0.00389
3	0.16317	0.19580*	0.13054	0.05439	0.01166
4	0.09790	0.17405	0.16317	0.09324	0.02720
5	0.05439	0.13598	0.17483*	0.13598	0.05439
6	0.02720	0.09324	0.16317	0.17405	0.09790
7	0.01166	0.05439	0.13054	0.19580*	0.16317
8	0.00389	0.02486	0.08392	0.18648	0.25641
9	0.00078	0.00699	0.03497	0.12821	0.38462*

Table 2.1: $P(X_{(r)} \in I_j)$ for $m = 5$ and $n = 8$

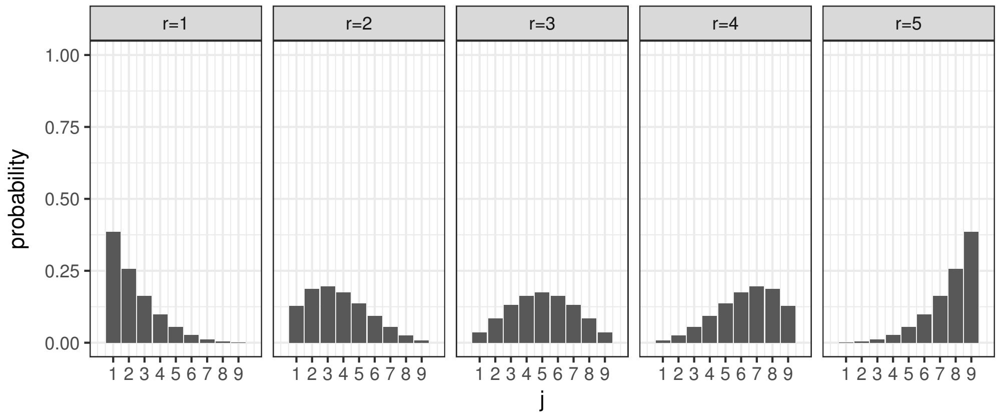

Figure 2.2: $P(X_{(r)} \in I_j)$ for $r = 1, \ldots, 5$ and $j = 1, \ldots, 9$

that the probabilities for the event $X_{(1)} \in I_j$ for $j = 1, \ldots, n+1$ are decreasing, while they are increasing for $X_{(5)} \in I_j$. When $r = 3$, the probability distribution is symmetric around the interval $j = 5$; this occurs for $r = \frac{m+1}{2}$ and $j = \frac{n+2}{2}$ if m is odd and n is even.

To illustrate the NPI lower and upper probabilities for the event $X_{(r)} \in S$, we consider an artificial data set consisting of the ordered numbers from 1 to 8, so $n = 8$. Note that these numbers should not be interpreted as counts, one could interpret them as ranks of actual real-valued data. The NPI lower and upper probabilities for event $X_{(r)} \in S$ for all $r = 1, \ldots, m$ and $S = (0, 2.5)$, $S = (1.5, 4.5)$, $S = (3.5, 5.5)$, $S = (4.2, 7.5)$ and $S = (6.5, 10)$ are presented in Table 2.2 for $m = 5$ and Figure 2.3 for $m = 10$. The plotted lines for each value of r represent the intervals bounded by the corresponding NPI lower and upper probabilities, hence the length of each line is the imprecision for that event. The results clearly illustrate the symmetry i.e. $[\underline{P},\overline{P}](X_{(1)} \in (0,2.5)) = [\underline{P},\overline{P}](X_{(5)} \in (6.5,10))$ and $[\underline{P},\overline{P}](X_{(r)} \in (3.5,5.5))$ symmetric around $r = 3$. The imprecision for $[\underline{P},\overline{P}](X_{(1)} \in (0,2.5))$, decreases as the value of r increases, which is because $P(X_{(r)} \in I_1)$ and $P(X_{(r)} \in I_3)$ are included in the upper probability but not in the lower probability, and the probability mass for $X_{(r)} \in I_1$ is large for $r = 1$.

$\Diamond$

	$S = (0,2.5)$		$S = (1.5,4.5)$		$S = (3.5,5.5)$		$S = (4.2,7.5)$		$S = (6.5,10)$	
r	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
1	0.25641	0.80420	0.26107	0.57187	0.05439	0.17949	0.03885	0.09713	0.00389	0.01632
2	0.18648	0.51049	0.36985	0.69231	0.13598	0.40326	0.14763	0.30847	0.02486	0.08625
3	0.08392	0.24942	0.29371	0.55245	0.17483	0.50117	0.29371	0.55245	0.08392	0.24942
4	0.02486	0.08625	0.14763	0.30847	0.13598	0.40326	0.36985	0.69231	0.18648	0.51049
5	0.00389	0.01632	0.03885	0.09713	0.05439	0.17949	0.26107	0.57187	0.25641	0.80420

Table 2.2: $P(X_{(r)} \in S)$ for $m = 5$ and $n = 8$

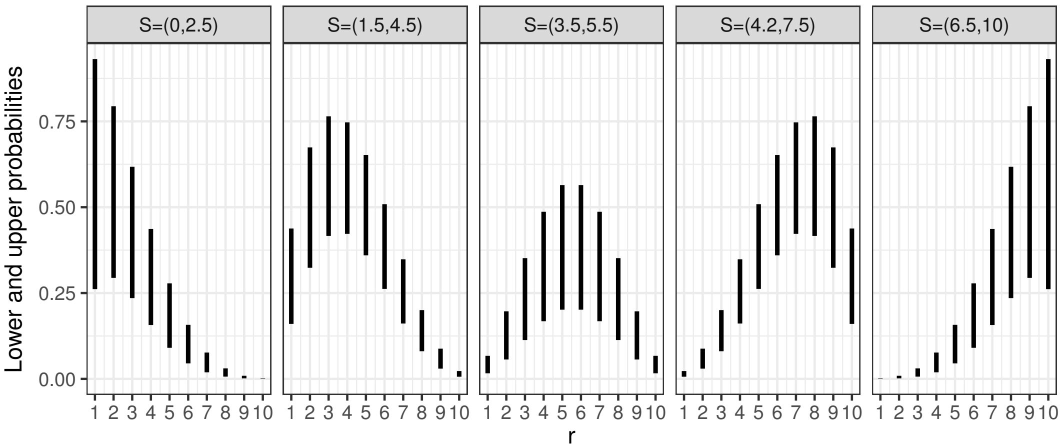

Figure 2.3: $P(X_{(r)} \in S)$ for $m = 10$ and $n = 8$

2.3 Multiple future order statistics

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 (2.1), 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 $A_{(.)}$ assumptions imply the probabilities

$$P\left(\bigcap_{r \in R} \left\{X_{(r)} \in I_{j_r}\right\}\right) =$$
$$\binom{n+m}{n}^{-1} \binom{r_1 + j_{r_1} - 2}{r_1 - 1} \times \prod_{i=2}^{t} \left\{\binom{r_i - r_{i-1} - 1 + j_{r_i} - j_{r_{i-1}}}{r_i - r_{i-1} - 1}\right\} \binom{m - r_t + n - j_{r_t} + 1}{m - r_t} \tag{2.5}$$

Note that, throughout this thesis, a product over an empty set is defined to be equal to 1; this applied here if $t = 1$. For the special case of two order statistics, $X_{(r)}$ and $X_{(s)}$ with $r < s$ and with $j \leq l$, we have

$$P(X_{(r)} \in I_j, X_{(s)} \in I_l) = \frac{\binom{r+j-2}{r-1}\binom{s-r-1+l-j}{s-r-1}\binom{m-s+n-l+1}{m-s}}{\binom{n+m}{n}} \tag{2.6}$$

The RHS of Equation (2.6) is derived as follows. We count the number of orderings of $r-1$ future observations among $j-1$ past observations to the left of $X_{(r)}$, multiply this by the number of orderings of $s-r-1$ future observations among $l-j$ past observations which are located between $X_{(r)}$ and $X_{(s)}$, and multiply further by the number of orderings of the remaining $m-s$ future observations among $n-(l-1)$ data observations to the right of $X_{(s)}$. The denominator is again the total number of orderings of m future observations among n past observations. The RHS of Equation (2.5) is derived similarly as Equation (2.6) but for two or more future observations. For the special case $j = l$, we have

$$P(X_{(r)} \in I_j, X_{(s)} \in I_j) = \frac{\binom{r+j-2}{r-1}\binom{m-s+n-j+1}{m-s}}{\binom{n+m}{n}} \tag{2.7}$$

Example 2.2. The joint probabilities for the event $X_{(2)} \in I_j, X_{(5)} \in I_l$ for $j, l = 1, \ldots, 9$ are given in Table 2.3. We illustrate these probabilities for the same values $n = 8$ and $m = 5$ as used in Example 2.1.

Table 2.3 shows that $\sum_{j=1}^{9} \sum_{l=1}^{9} P(X_{(2)} \in I_j, X_{(5)} \in I_l) = 1$. Comparing Table

2.3 with Table 2.1 illustrates that $\sum_{l=1}^{9} P(X_{(2)} \in I_j, X_{(5)} \in I_l) = P(X_{(2)} \in I_j)$ and $\sum_{j=1}^{9} P(X_{(2)} \in I_j, X_{(5)} \in I_l) = P(X_{(5)} \in I_l)$. The value of $P(X_{(2)} \in I_j, X_{(5)} \in I_l)$ increases as the value of l increases. The maximum probability for such an event occurs for $X_{(2)} \in I_3$ or $X_{(2)} \in I_4$ and $X_{(5)} \in I_9$.

					l					
j	1	2	3	4	5	6	7	8	9	Total
1	0.00078	0.00233	0.00466	0.00777	0.01166	0.01632	0.02176	0.02797	0.03497	0.12821
2	0	0.00155	0.00466	0.00932	0.01554	0.02331	0.03263	0.04351	0.05594	0.18648
3	0	0	0.00233	0.00699	0.01399	0.02331	0.03497	0.04895	0.06527	0.19580
4	0	0	0	0.00311	0.00932	0.01865	0.03108	0.04662	0.06527	0.17405
5	0	0	0	0	0.00389	0.01166	0.02331	0.03885	0.05828	0.13598
6	0	0	0	0	0	0.00466	0.01399	0.02797	0.04662	0.09324
7	0	0	0	0	0	0	0.00544	0.01632	0.03263	0.05439
8	0	0	0	0	0	0	0	0.00622	0.01865	0.02486
9	0	0	0	0	0	0	0	0	0.00699	0.00699
Total	0.00078	0.00389	0.01166	0.02720	0.05439	0.09790	0.16317	0.25641	0.38462	1

Table 2.3: $P(X_{(2)} \in I_j, X_{(5)} \in I_l)$ for $j, l = 1, \ldots, 9$

$\Diamond$

2.4 Conditional probabilities for future order statistics

In this section, we present conditional probabilities for events involving a subset of the future order statistics, given information about another subset of the future order statistics. Let $R = \{r_1, \ldots, r_t\}$ as introduced in Section 2.3, $D = \{d_1, \ldots, d_v\}$ with $d_1 < d_2 < \ldots < d_v$ and $1 \leq v \leq m - t$, such that $R \cap D = \emptyset$. 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 $P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right)$, 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 $A_{(.)}$ assumptions lead to the following conditional probabilities

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

$$= \frac{P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \cap \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right)}{P\left(\bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right)} = \frac{P\left(\bigcap_{c\in C}\left\{X_{(c)} \in I_{j_c}\right\}\right)}{P\left(\bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right)}$$

$$= \frac{\binom{c_1+j_{c_1}-2}{c_1-1} \times \prod_{i=2}^{w}\binom{c_i-c_{i-1}-1+j_{c_i}-j_{c_{i-1}}}{c_i-c_{i-1}-1} \times \binom{m-c_w+n-j_{c_w}+1}{m-c_w}}{\binom{d_1+j_{d_1}-2}{d_1-1} \times \prod_{i=2}^{v}\binom{d_i-d_{i-1}-1+j_{d_i}-j_{d_{i-1}}}{d_i-d_{i-1}-1} \times \binom{m-d_v+n-j_{d_v}+1}{m-d_v}} \tag{2.8}$$

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

$$P\left(X_{(r)} \in I_j \mid X_{(d)} \in I_l\right) = \frac{P\left(\left\{X_{(d)} \in I_l\right\} \cap \left\{X_{(r)} \in I_j\right\}\right)}{P\left(X_{(d)} \in I_l\right)} = \frac{\binom{r-d-1+j-l}{r-d-1}\binom{m-r+n-j+1}{m-r}}{\binom{n-l+1+m-d}{m-d}} \tag{2.9}$$

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

$$P\left(X_{(r)} \in I_j \mid X_{(d)} \in I_l\right) = \frac{P\left(\left\{X_{(r)} \in I_j\right\} \cap \left\{X_{(d)} \in I_l\right\}\right)}{P\left(X_{(d)} \in I_l\right)} = \frac{\binom{r+j-2}{r-1}\binom{d-r-1+l-j}{d-r-1}}{\binom{d+l-2}{d-1}} \tag{2.10}$$

For completeness, $P\left(X_{(r)} \in I_j \mid X_{(d)} \in I_l\right) = 0$ if either $r > d$ and $j < l$ or $r < d$ and $j > l$, and $P\left(X_{(d)} \in I_j \mid X_{(d)} \in I_l\right) = 1$ if $j = l$ and $P\left(X_{(d)} \in I_j \mid X_{(d)} \in I_l\right) = 0$ otherwise.

It is straightforward to show that for the general conditional probability (2.8), the following property holds

$$P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D_R}\left\{X_{(d)} \in I_{j_d}\right\}\right) \tag{2.11}$$

where $D_R \subset D$ consists of the elements of D which in the combined set $C = R \cup D$ have an element of R as a neighbour. Thus,

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

where the 'or' is of course not strict and events concerning the non-existent c_0 or c_{w+1} do not hold. Property (2.11) 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 Equation (2.8). 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 [79], 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

$$P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid X_{(d_v)} \in I_{j_{d_v}}\right)$$
$$= \frac{\binom{r_1-d_v-1+j_{r_1}-j_{d_v}}{r_1-d_v-1} \times \prod_{i=2}^{t}\binom{r_i-r_{i-1}-1+j_{r_i}-j_{r_{i-1}}}{r_i-r_{i-1}-1} \times \binom{m-r_t+n-j_{r_t}+1}{m-r_t}}{\binom{m-d_v+n-j_{d_v}+1}{m-d_v}} \tag{2.12}$$

The backward analogue of this result may also be of use, that is if $d_1 > r_t$ and $j_{d_1} \geq j_{r_t}$ then

$$P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid \bigcap_{d\in D}\left\{X_{(d)} \in I_{j_d}\right\}\right) = P\left(\bigcap_{r\in R}\left\{X_{(r)} \in I_{j_r}\right\} \mid X_{(d_1)} \in I_{j_{d_1}}\right)$$
$$= \frac{\binom{r_1+j_{r_1}-2}{r_1-1} \times \prod_{i=2}^{t}\binom{r_i-r_{i-1}-1+j_{r_i}-j_{r_{i-1}}}{r_i-r_{i-1}-1} \times \binom{d_1-r_t-1+j_{d_1}-j_{r_t}}{d_1-r_t-1}}{\binom{d_1+j_{d_1}-2}{d_1-1}} \tag{2.13}$$

An interesting special case of the probability (2.9) 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 $j \geq l$,

$$\begin{aligned} P(X_{(r)} \in I_j | X_{(r-1)} \in I_l) &= \frac{\binom{r-r+1-1+j-l}{r-r+1-1}\binom{m-r+n-j+1}{m-r}}{\binom{n-l+1+m-r+1}{m-r+1}} = \frac{\binom{m-r+n-j+1}{m-r}}{\binom{n-l+m-r+2}{m-r+1}} \\ &= \frac{(m-r+1)}{(n-l+m-r+2)} \prod_{k=2}^{m-r+1}\left[\frac{n-j+k}{n-l+k}\right] \end{aligned} \tag{2.14}$$

This is equal to the probability for the event that $X_{(1)} \in I_{j-l+1}$, as given by Equation (2.2), for the case with $n-l+1$ data observations and $m-r+1$ future observations based on the corresponding $A_{(.)}$ assumption. A more general form of this result is presented in the following proposition, followed by a special case given as a corollary. Since the proofs of these properties are straightforward, they are not included.

Proposition 2.1. For $r > d$ and $j \geq l$, the NPI probability for the event that the

r-th ordered future observation belongs to interval I_j, given that the d-th ordered future observation belongs to I_l, as given by Equation (2.9), is equal to the NPI probability for the event that $X_{(r-d)} \in I_{j-l+1}$, as given by Equation (2.2), 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 r-th future observation belongs to I_j, given that the d-th future observation belongs to I_l, as given by Equation (2.10), is equal to the probability for the event that $X_{(d-r)} \in I_{l-j+1}$, as given by Equation (2.2), for $l-1$ data observations and $d-1$ future observations.

Corollary 2.1. The conditional probability (2.9) for $j = l$ and $r > d$ is,

$$\begin{aligned} P(X_{(r)} \in I_j | X_{(d)} \in I_j) &= \prod_{k=0}^{r-d-1} \frac{m-r+1+k}{m-r+n-j+2+k} \\ &= \prod_{k=0}^{r-d-1} P(X_{(r-k)} \in I_j | X_{(r-k-1)} \in I_j) \end{aligned} \tag{2.15}$$

That is, the probability for the event that the r-th future observation belongs to I_j given that the d-th 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 $X_{(r-k-1)} \in I_j$, for $k = 0, \ldots, r-d-1$. From Proposition 2.1, this probability is equal to the probability for the event that $X_{(r-d)} \in I_1$, as given by Equation (2.2), for $n-j+1$ data observations and $m-d$ future observations.

Corollary 2.2. The conditional probability (2.10) for $j = l$ and $r < d$, is

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

That is, the probability for the event that the r-th future observation belongs to I_j given that the d-th 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 $X_{(r+1+k)} \in I_j$, for $k = 0, \ldots, d-r-1$. From Proposition 2.1, this probability is equal to the probability for the event that $X_{(d-r)} \in I_1$, as given by Equation (2.2), for $j-1$ data observations and $d-1$ future observations.

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

Proposition 2.2. 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 (2.2), for $j_{d_2} - j_{d_1}$ data observations and $d_2 - d_1 - 1$ future observations.

The proof of proposition 2.2 is straightforward, using Equation (2.8) 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 (2.2) 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.

The information used in the conditional probability (2.8) provides for each $X_{(d)}$, with $d \in D$, the interval in the partition created by the n observations to which this future order statistic belongs. 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 $A_{(.)}$-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.

Example 2.3. Suppose that one has $n = 2$ observations, $x_1 = 4$ and $x_2 = 10$, so $I_1 = (-\infty, 4)$, $I_2 = (4, 10)$ and $I_3 = (10, \infty)$, and consider $m = 3$ future observations. Suppose that we are interested in $X_{(3)}$ given either $X_{(2)} \in (4, 10)$ or $X_{(2)} = 8$. The corresponding conditional probabilities are $P(X_{(3)} \in (4, 10) | X_{(2)} \in (4, 10)) = 0.5$, $P(X_{(3)} \in (8, 10) | X_{(2)} = 8) = 0.5$. The important difference between these two cases is just based on the more detailed information in the latter case about $X_{(2)}$ which reduces interval $(4, 10)$ for $X_{(3)}$ to $(8, 10)$, but it does not affect the corresponding probability. The absence of this detailed information in the first case avoids reduction of the interval $(4, 10)$ for $X_{(3)}$, but of course the order $X_{(2)} < X_{(3)}$ has to hold.

$\Diamond$

Analysis of the conditional probability (2.10) 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 for $j = 1, \ldots, l$ and $l = 1, \ldots, n+1$, is defined as

$$F_{r|d}(j|l) = P(X_{(r)} \in \bigcup_{k=1}^{j} I_k | X_{(d)} \in I_l) = \sum_{k=1}^{j} P(X_{(r)} \in I_k | X_{(d)} \in I_l) \qquad (2.16)$$

where $F_{r|d}(n+1|l) = 1$ and $F_{r|d}(1|1) = 1$. Let $F_r(j)$ be the CDF for $X_{(r)} \in I_j$, i.e. $F_r(j) = \sum_{k=1}^{j} P(X_{(r)} \in I_k)$.

Consider two future order statistics $X_{(r)}$ 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 $F_{r|d}(j|l-1) \geq F_{r|d}(j|l)$, for all j, then $X_{(r)} \in I_j | X_{(d)} \in I_{l-1}$ is said to be stochastically less than $X_{(r)} \in I_j | X_{(d)} \in I_l$, denoted by $X_{(r)} \in I_j | X_{(d)} \in I_{l-1} \leq_{st} X_{(r)} \in I_j | X_{(d)} \in I_l$.

Theorem 2.1. For $l = 1, \ldots, n+1$, we have

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

Therefore, $X_{(r)} \in I_j | X_{(d)} \in I_1 \leq_{st} X_{(r)} \in I_j | X_{(d)} \in I_2 \leq_{st} \ldots \leq_{st} X_{(r)} \in I_j | X_{(d)} \in I_{n+1}$.

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

$$\sum_{k=1}^{j} P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \geq \sum_{k=1}^{j} P(X_{(r)} \in I_k | X_{(d)} \in I_l) \qquad (2.17)$$

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 (2.17) to hold is if there exists one value w_r such that

$$P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \geq P(X_{(r)} \in I_k | X_{(d)} \in I_l) \quad \text{for all} \quad k \leq w_r \qquad (2.18)$$

and

$$P(X_{(r)} \in I_k | X_{(d)} \in I_{l-1}) \leq P(X_{(r)} \in I_k | X_{(d)} \in I_l) \quad \text{for all} \quad k > w_r \tag{2.19}$$

Using Equation (2.10), it is straightforward to show that Equation (2.18) holds if and only if

$$k \leq \frac{r(l-1)}{(d-1)} + 1 \tag{2.20}$$

Similarly, Equation (2.19) holds if and only if

$$k \geq \frac{r(l-1)}{(d-1)} + 1 \tag{2.21}$$

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

□

Example 2.4. Consider $n = 8$ observations and $m = 5$ future observations, as also used in Example 2.1. Table 2.4 presents the conditional probabilities for the event $X_{(2)} \in I_j | X_{(5)} \in I_l$ for all $j, l = 1, \ldots, 9$. Table 2.4 illustrates that Equations (2.18) and (2.19) hold, namely when $w_2 = \frac{2(l-1)}{4}$, as given in the proof of Theorem 2.1, we have

$$P(X_{(2)} \in I_j | X_{(5)} \in I_{l-1}) \geq P(X_{(2)} \in I_j | X_{(5)} \in I_l) \qquad \text{for} \quad j \leq \frac{l-1}{2} + 1$$

and

$$P(X_{(2)} \in I_j | X_{(5)} \in I_{l-1}) \leq P(X_{(2)} \in I_j | X_{(5)} \in I_l) \qquad \text{for} \quad j \geq \frac{l-1}{2} + 1$$

Thus, the conditional CDF of $X_{(2)} \in I_j | X_{(5)} \in I_l$ is less than the conditional CDF of $X_{(2)} \in I_j | X_{(5)} \in I_{l-1}$, then $X_{(2)} \in I_j | X_{(5)} \in I_l \geq^{st} X_{(2)} \in I_j | X_{(5)} \in I_{l-1}$, as illustrated in Table 2.5.

◊

	l								
j	1	2	3	4	5	6	7	8	9
1	1	0.6	0.4	0.28571	0.21428	0.16666	0.13333	0.10909	0.09090
2	0	0.4	0.4	0.34285	0.28571	0.23809	0.20000	0.16969	0.14545
3	0	0	0.2	0.25714	0.25714	0.23809	0.21428	0.19090	0.16969
4	0	0	0	0.11428	0.17142	0.19047	0.19047	0.18181	0.16969
5	0	0	0	0	0.07142	0.11904	0.14285	0.15151	0.15151
6	0	0	0	0	0	0.04761	0.08571	0.10909	0.12121
7	0	0	0	0	0	0	0.03333	0.06363	0.08484
8	0	0	0	0	0	0	0	0.02424	0.04848
9	0	0	0	0	0	0	0	0	0.01818
Total	1	1	1	1	1	1	1	1	1

Table 2.4: $P(X_{(2)} \in I_j | X_{(5)} \in I_l)$

	l								
j	1	2	3	4	5	6	7	8	9
1	1	0.6	0.4	0.2857	0.2143	0.1667	0.1333	0.1091	0.0909
2	1	1	0.8	0.6286	0.5000	0.4048	0.3333	0.2788	0.2364
3	1	1	1	0.8857	0.7571	0.6429	0.5476	0.4697	0.4006
4	1	1	1	1	0.9286	0.8333	0.7381	0.6515	0.5758
5	1	1	1	1	1	0.9524	0.8810	0.8030	0.7273
6	1	1	1	1	1	1	0.9667	0.9121	0.8485
7	1	1	1	1	1	1	1	0.9758	0.9333
8	1	1	1	1	1	1	1	1	0.9818
9	1	1	1	1	1	1	1	1	1

Table 2.5: $F_{2|5}(j|l) = \sum_{k=1}^{j} P(X_{(2)} \in I_k | X_{(5)} \in I_l)$

2.5 Median and mean of future observations

In this section, we derive the NPI probability distributions for the median and the mean of m future observations. These have not yet been considered in NPI, and will be used in Chapter 5.

2.5.1 The median of m future observations

We denote the median of m future observations by M_m. To keep it simple, we first consider the case in which m is odd, so $M_m = X_{(\frac{m+1}{2})}$, as it can be derived straightforwardly by replacing r by $\frac{m+1}{2}$ in Equation (2.2). Using Equation (2.2),

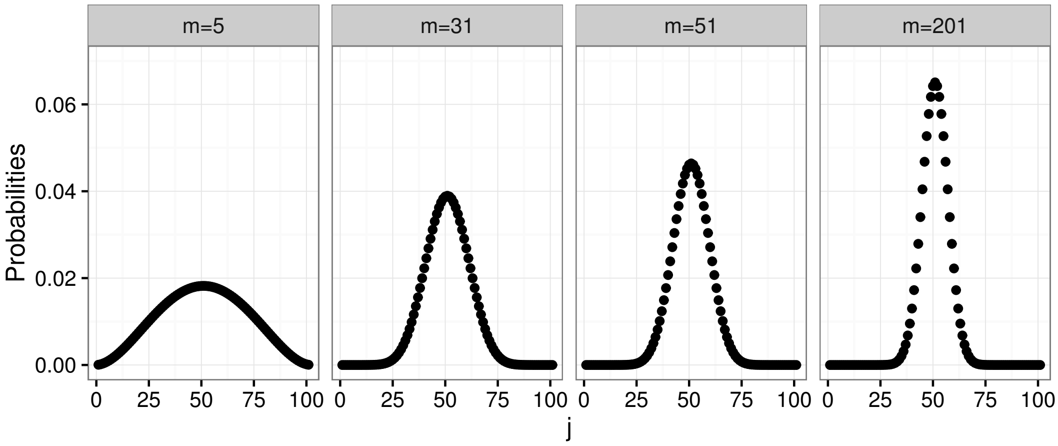

Figure 2.4: $P(M_m \in I_j)$ for $n = 100$

the NPI probability for the event $M_m \in I_j = (x_{j-1}, x_j)$ is

$$P(M_m \in I_j) = P(X_{(\frac{m+1}{2})} \in I_j) = \frac{\binom{\frac{m-1}{2}+j-1}{j-1}\binom{n-j+1+\frac{m-1}{2}}{n-j+1}}{\binom{n+m}{n}} \tag{2.22}$$

Figure 2.4 illustrates probability (2.22) for $n = 100$ and for different values of m. Note that $P(M_m \in I_j)$ is symmetric around the value I_{51} for $m = 5, 31, 51, 201$ and $j = 1, \ldots, 101$. The lower and upper probabilities for the event $M_m > z$, if $z \neq x_j$ for $j = 1 \ldots, n$, are

$$\underline{P}(M_m > z) = \sum_{j=1}^{n+1} 1\{x_{j-1} > z\} P(M_m \in I_j) \tag{2.23}$$

$$\overline{P}(M_m > z) = \sum_{j=1}^{n+1} 1\{x_j > z\} P(M_m \in I_j) \tag{2.24}$$

The probabilities (2.23) and (2.24) are step functions of z, as illustrated in Figure 2.5 for $m = 5$ and different values of n. Figure 2.5 shows that as n increases, the imprecision of the lower and upper probabilities decreases.

More generally, interest may be in the event $M_m \in Z = (z_1, z_2)$, in which both of z_1, z_2 are not equal to data observations. The NPI lower and upper probabilities for this event are

$$\underline{P}(M_m \in (z_1, z_2)) = \sum_{j=1}^{n+1} 1\{I_j \subseteq (z_1, z_2)\} P(M_m \in I_j) \tag{2.25}$$

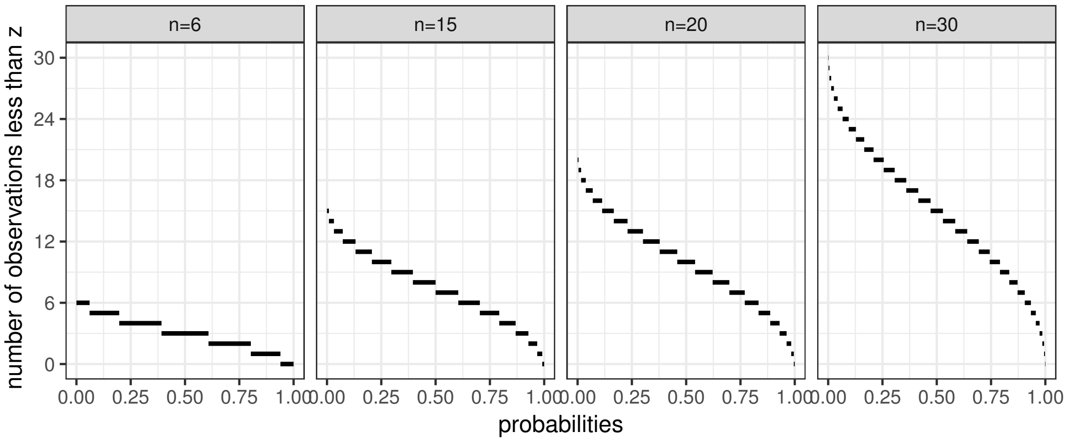

Figure 2.5: $[\underline{P}, \overline{P}](M_m > z)$ for $m = 5$ and $n = 6, 15, 20, 30$.

$$\overline{P}(M_m \in (z_1, z_2)) = \sum_{j=1}^{n+1} 1\{I_j \cap (z_1, z_2) \neq \emptyset\} P(M_m \in I_j) \tag{2.26}$$

NPI for the median of m future observations is relatively more complex if m is even, in which case $M_m = \frac{X_{(\frac{m}{2})}+X_{(\frac{m}{2}+1)}}{2}$. In this case NPI does not provide precise probabilities for the event $M_m \in I_j$, but instead the NPI lower and upper probabilities for the event $M_m \in I_j$ are

$$\underline{P}(M_m \in I_j) = \\ \sum_{i=1}^{n+1} \sum_{l=i}^{n+1} 1\{x_{j-1} \leq \frac{x_{i-1} + x_{l-1}}{2} \wedge \frac{x_i + x_l}{2} \leq x_j\} \times P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_l) \tag{2.27}$$

$$\overline{P}(M_m \in I_j) = \\ \sum_{i=1}^{n+1} \sum_{l=i}^{n+1} 1\{(\frac{x_{i-1} + x_{l-1}}{2}, \frac{x_i + x_l}{2}) \cap I_j \neq \emptyset\} \times P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_l) \tag{2.28}$$

Example 2.5. Consider an artificial data set consisting of the ordered numbers from 3 to 50, so $n = 48$. The lower and upper probabilities for the event that the median for $m = 6, 32, 52$ and 202 is in interval $I_j = (x_{j-1}, x_j)$ are shown in Figure 2.6. The results clearly illustrate that by increasing the value of m the NPI lower and upper probabilities for event that the $M_m \in I_j$ increase for nearly all j, apart

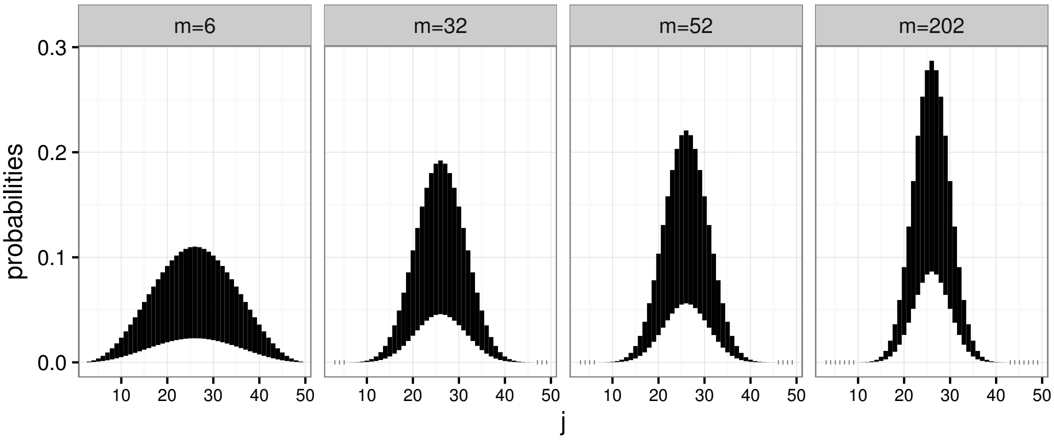

Figure 2.6: $[\underline{P}\ ,\overline{P}](M_m \in I_j)$ for $m = 6, 32, 52, 202$, $n = 48$ and $j = 1, \ldots, 49$

from small and large values of j. The maximum lower and upper probabilities are achieved for event $M_m \in I_{25}$, for which the imprecision is also maximal.

$\Diamond$

Similarly, for m even, the lower and upper probabilities for the event $M_m > z$, with $z \neq x_j$, for $j = 1, \ldots, n$, are

$$\underline{P}(\frac{X_{(\frac{m}{2})} + X_{(\frac{m}{2}+1)}}{2} \geq z)$$
$$= \sum_{i=1}^{n+1} \sum_{l=i}^{n+1} 1\{(\frac{x_{i-1} + x_{l-1}}{2}, \frac{x_i + x_l}{2}) \geq z\} \times P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_l) \tag{2.29}$$

$$\overline{P}(\frac{X_{(\frac{m}{2})} + X_{(\frac{m}{2}+1)}}{2} \geq z)$$
$$= \sum_{i=1}^{n+1} \sum_{l=i}^{n+1} 1\{(\frac{x_{i-1} + x_{l-1}}{2}, \frac{x_i + x_l}{2}) \cap (z, \infty) \neq \emptyset\} \times P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_l) \tag{2.30}$$

2.5.2 The mean of m future observations

We denote the mean of m future observations by μ_m and the mean corresponding to a specific ordering O_i of the future observations among n observations by μ_m^i. When we consider μ_m^i, we must avoid possible probability mass in $-\infty$ or ∞, because it affects the mean of the m future observations. We assume finite bounds $L < R$ for the data observations and future observations, such that $L < x_1 < \ldots < x_n < R$, and define $x_0 = L$ and $x_{n+1} = R$ for the $A_{(.)}$ assumptions.

The randomness of μ_m can be considered as a result from not knowing the ordering O_i and the place of the s^i_j future observations within each interval I_j. Given the orderings O_i, the mean μ^i_m is imprecise, because we do not assume where the s^i_j future observations precisely are within interval I_j, we remind the reader that s^i_j is the number of future observations which, according to ordering O_i, are in interval I_j. Thus, based on the $A_{(.)}$ assumptions, we can only derive bounds for μ_m and for probabilities for events involving μ_m. The maximum lower bound and the minimum upper bound for the mean μ^i_m of the m future observations, for given ordering O_i, are

$$\underline{\mu^i_m} = \frac{1}{m}\sum_{j=1}^{n+1} s^i_j x_{j-1} \tag{2.31}$$

$$\overline{\mu^i_m} = \frac{1}{m}\sum_{j=1}^{n+1} s^i_j x_j \tag{2.32}$$

The expected value for the mean μ_m of the m future observations is

$$E(\mu_m) = \sum_{i=1}^{\binom{n+m}{n}} E(\mu_m|O_i)P(O_i) = \frac{1}{\binom{n+m}{n}}\sum_{i=1}^{\binom{n+m}{n}} \mu^i_m$$

The maximum lower bound for $E(\mu_m)$, which is also called the lower expectation, is

$$\begin{aligned}
\underline{E}(\mu_m) &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\underline{\mu^i_m} = \frac{1}{\binom{n+m}{n}}\sum_{i=1}^{\binom{n+m}{n}} \frac{1}{m}\sum_{j=1}^{n+1} s^i_j x_{j-1} \\
&= \frac{1}{\binom{n+m}{n}}\frac{1}{m}\sum_{j=1}^{n+1} x_{j-1}[\sum_{i=1}^{\binom{n+m}{n}} s^i_j] = \frac{1}{\binom{n+m}{n}}\frac{1}{m}\sum_{j=1}^{n+1} x_{j-1}[\frac{m\binom{n+m}{n}}{n+1}] \\
&= \frac{1}{n+1}\sum_{j=1}^{n+1} x_{j-1} = \frac{1}{n+1}\sum_{j=0}^{n} x_j
\end{aligned} \tag{2.33}$$

In this derivation, the third equality is based on the equality $\sum_{i=1}^{\binom{n+m}{n}} s^i_j = \frac{m\binom{n+m}{n}}{n+1} = \binom{n+1+m-1}{n+1}$, which holds for all j, since all orderings O_i are equally likely by Equation (2.1). Similarly, the minimum upper bound for $E(\mu_m)$, also called the

A:	$L=-2$	1	10							$R=12$
B:	$L=-2$	1	2	11	12					$R=14$
C:	$L=-2$	1	2	3	4	13	14	15	16	$R=18$

Table 2.6: Data sets

upper expectation, is

$$\overline{E}[\mu_m] = \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\overline{\mu_m^i} = \frac{1}{n+1}\sum_{j=1}^{n+1} x_j \tag{2.34}$$

The NPI lower and upper probabilities for the event $\mu_m > z$, are

$$\underline{P}(\mu_m \geq z) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{\underline{\mu_m^i} \geq z\} \tag{2.35}$$

$$\overline{P}(\mu_m \geq z) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{\overline{\mu_m^i} \geq z\} \tag{2.36}$$

For any interval $Z = (z_1, z_2)$ the NPI lower and upper probabilities for the event $\mu_m \in Z$ are

$$\underline{P}(\mu_m \in (z_1, z_2)) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{z_1 \leq \underline{\mu_m^i} \leq \overline{\mu_m^i} \leq z_2\} \tag{2.37}$$

$$\overline{P}(\mu_m \in (z_1, z_2)) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{(\underline{\mu_m^i}, \overline{\mu_m^i}) \cap (z_1, z_2) \neq \emptyset\} \tag{2.38}$$

Example 2.6. To illustrate the inference involving the mean of the m future observations given in Equations (2.37) and (2.38), we consider the data sets given in Table 2.6. It is often of interest to investigate whether the mean of the future observations falls in an interval (z_1, z_2), e.g. we want to know if the μ_m is somewhere around the centre of the given data.

Figure 2.7 presents the lower and upper probabilities for the events $\mu_m \in (z_1, z_2)$, for $m = 2, \ldots, 30$, corresponding to the datasets in cases A, B and C, for the events $\mu_m \in (2.5, 9.5)$, $\mu_m \in (3.5, 10.5)$ and $\mu_m \in (5.5, 12.5)$ respectively. The behaviour of each probability for the different datasets is quite similar as the

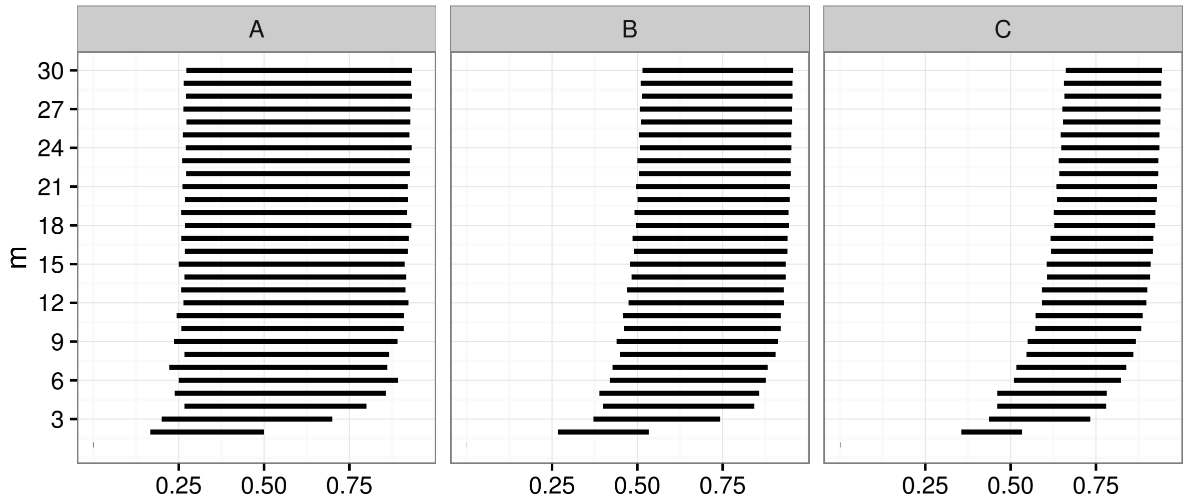

Figure 2.7: Lower and upper probabilities for the event $\mu^i_m \in (z_1, z_2)$ for data set in case A, B and C for $m = 2, \dots, 30$

intervals (z_1, z_2) for all data sets have the same length. However, it is clear from the results that the effect of increasing the sample size leads to decreasing the imprecision, because more information is available.

$\Diamond$

2.5.3 The expected value of a future order statistic

It may also be of interest to consider the NPI maximum lower and minimum upper bounds for the expectation of $X_{(r)}$. As in Section 2.5.2, we assume bounds L and R for the observations $x_1, \dots, x_n$, and for the future observations.

The lower and upper bounds for the expected value of $X_{(r)}$ are easily derived by putting all probability masses per interval to the left for the lower bound, and to the right for the upper bound, which leads to the lower and upper expected values

$$\underline{E}(X_{(r)}) = \sum_{j=1}^{n+1} x_{j-1} P(X_{(r)} \in I_j) \quad (2.39)$$

$$\overline{E}(X_{(r)}) = \sum_{j=1}^{n+1} x_j P(X_{(r)} \in I_j) \quad (2.40)$$

2.6 Concluding remarks

The results presented in this chapter provide new tools for predictive inference on order statistics of future observations. While some inferences coincide with classical results on order statistics [3], the explicit use of the $A_{(.)}$ assumptions and restriction to m future observations make the 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 developed 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 involving future order statistics will be illustrated in the next chapter, in particular on pairwise and multiple comparisons.

Chapter 3

Statistical inferences involving future order statistics

3.1 Introduction

In this chapter, we develop statistical methods based on the probability results presented in Chapter 2, so we consider inference involving future order statistics. Applications of statistical inference often involve comparison of two or more independent groups of data, e.g. resulting from two different treatments. In classical statistics, a popular method is to test equality of parameters in assumed parametric models, or to use the ranks of the observations in nonparametric methods, for example Wilcoxon's or Kruskal–Wallis tests can be used for two or more independent populations, respectively [46]. These methods assume that the random quantities for each population are independent and identically distributed. Comparisons of real-valued observations in NPI have been formulated in terms of a single future observation from each of two or more groups of data [19, 30]. We provide a generalisation by considering m future observations instead of a single future observation.

The key difference between comparison in NPI and classical statistics is in formulating the question of interest. In classical tests, usually the starting point is the hypothesis that both groups come from the same distribution, which is often impractical. The NPI approach does not require a particular hypothesis to be formulated; it uses a direct approach involving only future observations, which enables a natural manner of comparison that is particularly well suited if a decision must be made about, for example, the best treatment for the next units or individuals [35].

This chapter is organized as follows. Section 3.2 provides a generalisation of pairwise comparisons as presented by Coolen and Maturi [28], and different events of interest are considered. Section 3.3 presents the difference between future order statistics of two independent groups. Section 3.4 extends multiple comparisons as presented by Coolen and van der Laan [30], by considering the comparison in terms of the r-th order statistic out of m future observations from each group using the results derived in Chapter 2. Section 3.5 briefly discusses some further inferences that could be of interest, namely prediction intervals, the number of future order statistics in an interval, and the spacing between future order statistics. The chapter ends with some concluding remarks in Section 3.6.

3.2 Pairwise comparisons

Suppose that we have two independent groups of real-valued observations, X and Y; so the values in one group contain no information about the values in the other group, and 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, on X_{n_x+i} and Y_{n_y+i} for $i = 1, \ldots, m$. The theory presented in this chapter does not require limitation to the case $m_x = m_y$, but it seems to be quite logical when comparing future order statistics to consider the same number of future observations for each group; the generalisation 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

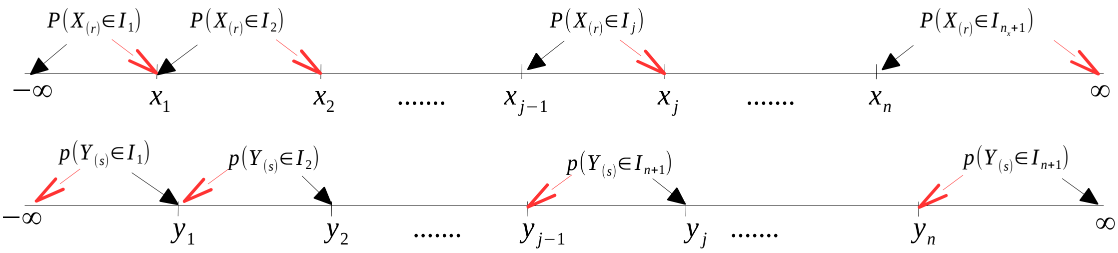

Figure 3.1: Locations of probability masses corresponds to lower and upper probabilities of the event $X_{(r)} < Y_{(s)}$, represented by red dots and black arrows respectively.

the event $X_{(r)} < Y_{(s)}$. The inference depends on $A_{(.)}$ assumptions for each group, denoted by $A^x_{(.)}$ and $A^y_{(.)}$ for group X and Y respectively. The NPI lower and upper probabilities for this event, based on the $A^x_{(.)}$ and $A^y_{(.)}$ assumptions per group, are

$$\underline{P}(X_{(r)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x} < y_{j_y-1}\} P(X_{(r)} \in I^x_{j_x}) P(Y_{(s)} \in I^y_{j_y}) \quad (3.1)$$

$$\overline{P}(X_{(r)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{j_x-1} < y_{j_y}\} P(X_{(r)} \in I^x_{j_x}) P(Y_{(s)} \in I^y_{j_y}) \quad (3.2)$$

The NPI lower (upper) probability is obtained by putting the probability mass per interval at an end point, for group X at right (left) end point and for group Y at the left (right) end point. This is illustrated in Figure 3.1. The special case with $r = s$ was earlier presented by Coolen and Maturi [28].

One may wish to compare two groups by taking multiple future order statistics into account. This can be done using the joint probability (2.5) presented in Chapter 2. For example, suppose that we are interested in comparing two independent groups X and Y by simultaneously considering the r-th and the s-th future order statistics from each group. We can use the joint probability given by Equation (2.6) for any event involving the r-th and s-th 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 3.1. The NPI lower and upper probabilities for the event $X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}$ are

$$\underline{P}(X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x+1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \sum_{l_y=j_y}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}, x_{l_x-1} > y_{l_y}\}$$
$$\times P(X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x}) \times P(Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{l_y}) \tag{3.3}$$

$$\overline{P}(X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} \sum_{j_y=1}^{n_y+1} \sum_{l_y=j_y+1}^{n_y+1} \Big[1\{x_{j_x-1} < y_{j_y}, x_{l_x} > y_{l_y-1}\}.$$
$$\times P(X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x}) \times P(Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{l_y}) \Big]$$
$$+ \sum_{j_y=1}^{n_y+1} \max_{y^*_{j_y} \in I_{j_y}} \left[P_{(j_y)} \right] \tag{3.4}$$

where

$$P_{(j_y)} = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} 1\{x_{j_x-1} < y^*_{j_y}, x_{l_x} > y^*_{j_y}\} P(X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x})$$
$$\times P(Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{j_y}) \tag{3.5}$$

The proof of Theorem 3.1 is given in Appendix B; the remaining maximisation for the derivation of the upper probability is also discussed in that appendix.

Another event that may be of interest is $X_{(r)} < Y_{(r)}, X_{(s)} < Y_{(s)}$. The NPI lower and upper probabilities for this event are

$$\underline{P}(X_{(r)} < Y_{(r)}, X_{(s)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} \sum_{j_y=1}^{n_y+1} \sum_{l_y=j_y}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}, x_{l_x} < y_{l_y-1}\}$$
$$\times P(X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x}) \times P(Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{l_y}) \tag{3.6}$$

$$\overline{P}(X_{(r)} < Y_{(r)}, X_{(s)} < Y_{(s)}) = \sum_{j_x=1}^{n_x+1} \sum_{l_x=j_x}^{n_x+1} \sum_{j_y=1}^{n_y+1} \sum_{l_y=j_y}^{n_y+1} 1\{x_{j_x-1} < y_{j_y}, x_{l_x-1} < y_{l_y}\}$$
$$\times P(X_{(r)} \in I^x_{j_x}, X_{(s)} \in I^x_{l_x}) \times P(Y_{(r)} \in I^y_{j_y}, Y_{(s)} \in I^y_{l_y}) \tag{3.7}$$

These lower and upper probabilities are derived similarly to those in Theorem 3.1, but note that, for this event, the upper probability does not require remaining optimization. We illustrate the NPI pairwise comparisons based on future order statistics in the following example.

Example 3.1. We consider the data set of a study of the effect of ozone environment on rats growth [42, 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 3.1.

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 3.1: Rats weight gain data

The NPI lower and upper probabilities (3.1) and (3.2) for the events $X_{(r)} < Y_{(r)}$, $r = 1, \ldots, m$, are displayed in Figure 3.2. The first row corresponds to the full data for the cases with $m = 5, 25, 200$, while the second row presents the corresponding figures with the observation -16.9 removed from group Y. This value could e.g. be removed because it could be considered as an outlier, so 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. Since the 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.

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

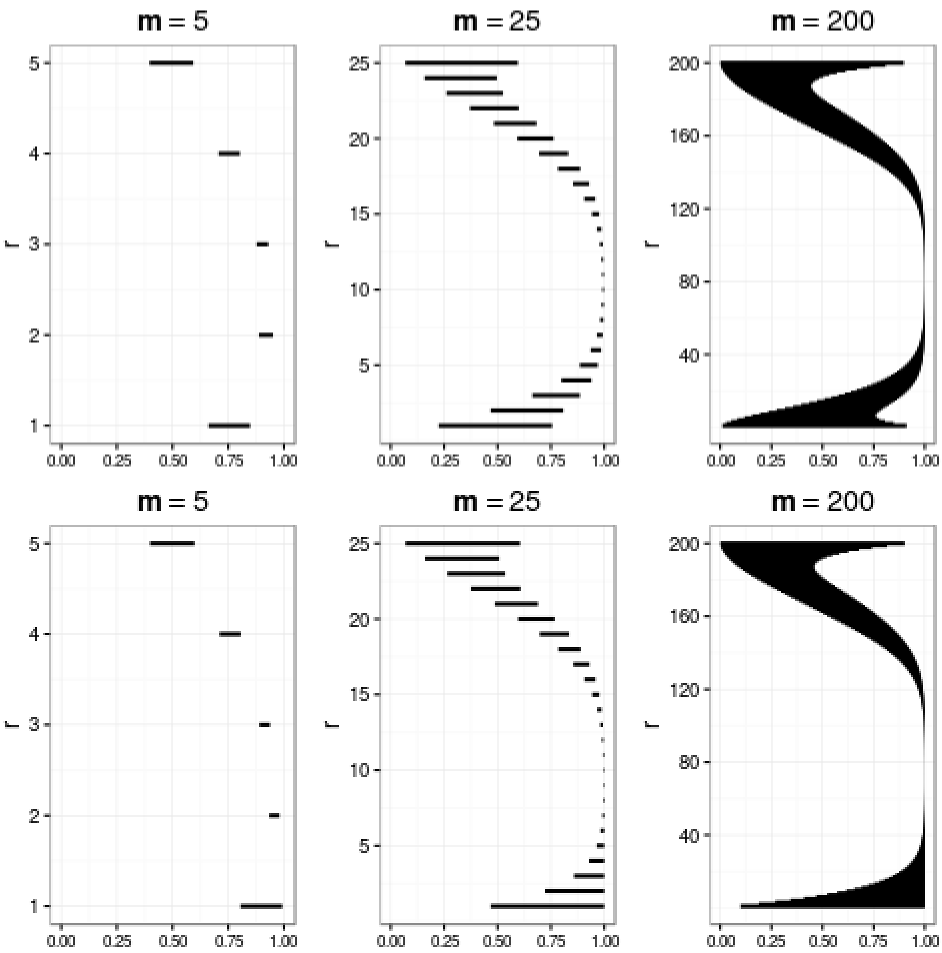

Figure 3.2: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)})$

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 both to 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 3.2 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 3.3 presents the NPI lower and upper probabilities 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 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, it is very unlikely that $Y_{(s)} > X_{(200)}$ for nearly all s, apart from large values of s for which we get imprecision close to 1, due to the X group data containing the two overall largest observations. In Figure 3.3, for $r = 1$ and $r = 50$ and $s > 80$ the NPI lower and upper probabilities for event $Y_{(s)} > X_{(r)}$ are close to one. The plotted line, which represents the interval bounded by lower and upper probabilities for each of these events is not shown in the figures, similarly for $r = 200$, and $s < 100$, in which cases $\underline{P}(Y_{(s)} > X_{(200)}) \simeq \overline{P}(Y_{(s)} > X_{(200)}) \simeq 0$. 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.

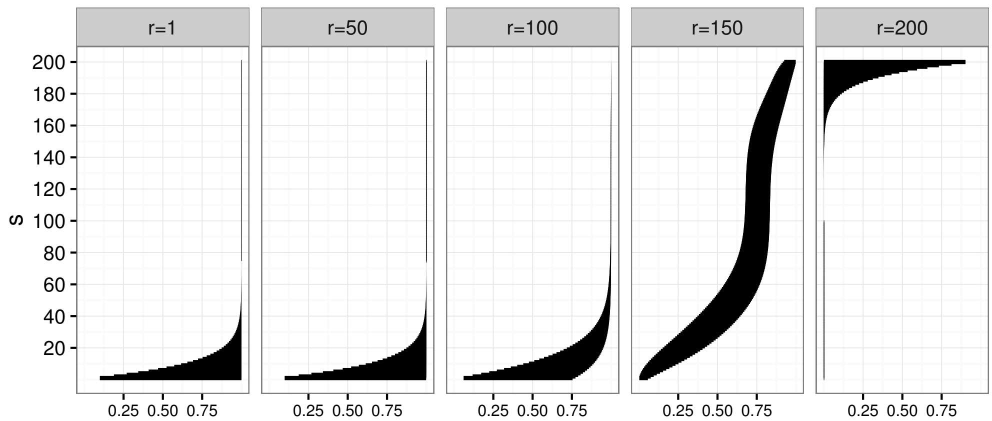

Figure 3.3: $[\underline{P}\ , \overline{P}](X_{(r)} < Y_{(s)})$ for $m = 200$

The NPI lower and upper probabilities for the events $X_{(r)} < Y_{(r)}, X_{(s)} > Y_{(s)}$, with $r < s$, are presented for these data in Figure 3.4, for $m = 100$ future observations for both groups X and Y. The presented cases are for $r = 5, 10, 25, 50, 75$, and for all $s = r + 1, \ldots, m$. The plotted lines in Figure 3.4 start at $s = r + 1$ as

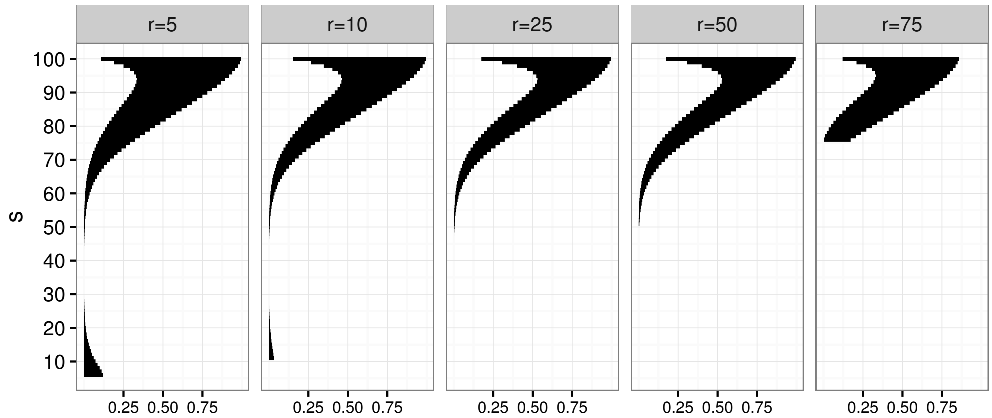

Figure 3.4: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)}, Y_{(s)} < X_{(s)})$ for $r < s$ and $m = 100$

we consider $r < s$. Note that again the smallest Y observation, -16.9, has been deleted. 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.

$\Diamond$

Example 3.2. We consider different values of n and m to illustrate their effect on the NPI inference of pairwise comparisons of the event $X_{(r)} < Y_{(r)}$. The NPI comparisons depend on the combined ordering of the two groups. Let us consider data from two independent groups X and Y, consisting of an even number n of observations from each group, with the following ranks; $x_1 < \ldots < x_{\frac{n}{2}} < y_1 < \ldots < y_n < x_{\frac{n}{2}+1} < \ldots < x_n$. The NPI lower and upper probabilities for the event $X_{(r)} < Y_{(r)}$ are plotted as a function of n in Figure 3.5 for $m = 5$ and $r = 1, \ldots, 5$, and in Figure 3.6 for $m = 100$ and $r = 1, 25, 50, 75, 200$. The plotted line for each value of n represents the interval bounded by the lower and upper probabilities, so the length of each line is the imprecision for the event $X_{(r)} < Y_{(r)}$.

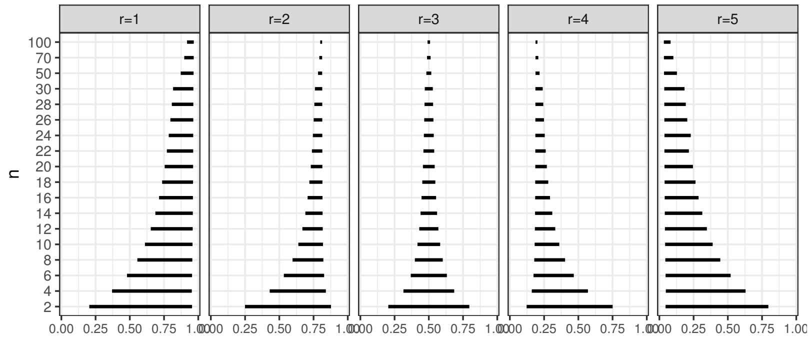

Figure 3.5: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)})$ for $m = 5$

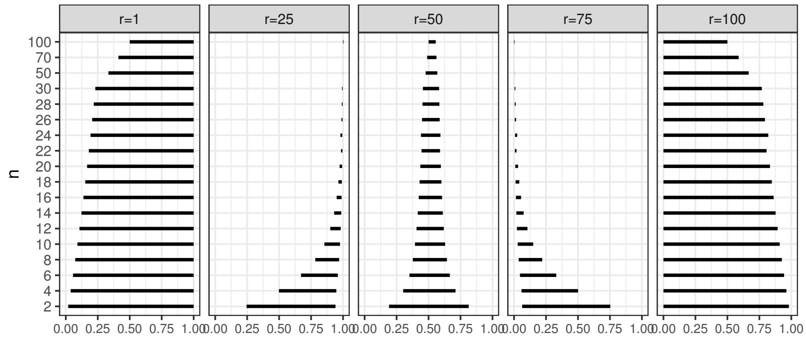

Figure 3.6: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)})$ for $m = 100$

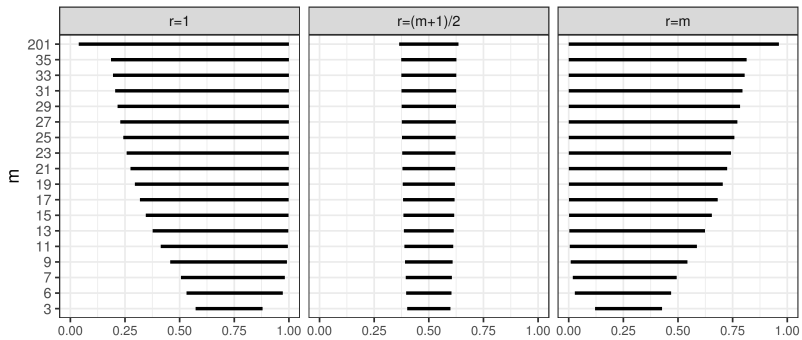

Figure 3.7: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)})$ for $n = 8$

These results clearly show that increasing the number n of data observations decreases the imprecision for such an event. However, for large m, as illustrated in Figure 3.6, the imprecision remains high for the extreme future order statistics, as no assumptions are made about the spread of the probability mass in the end intervals in both groups.

Figure 3.7 presents the lower and upper probabilities for the event that $X_{(r)} < Y_{(r)}$ for $n = 8$ and $r = 1, \frac{m+1}{2}, m$, considering different odd values of m. The results show that imprecision tends to increase as a function of m, particularly for the event involving extreme future observations and where m is much greater than n.

$\Diamond$

3.3 Differences between future order statistics of two groups

The pairwise comparison methods presented in Section 3.2 do not provide insight into the actual differences between these future order statistics. To achieve this, we generalise this method by considering the event that $X_{(r)} < Y_{(s)} + \delta$ for $\delta \in \mathbb{R}$. This is similar to the use of the so-called effect size in hypothesis testing, which is used to quantify a difference between two groups [17]. The generalisations of Equations (3.1) and (3.2) are as follows,

$$\underline{P}(X_{(r)} < Y_{(s)} + \delta) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} 1\{x_{j_x} < y_{j_y-1} + \delta\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(s)} \in I_{j_y}^y) \quad (3.8)$$

$$\overline{P}(X_{(r)} < Y_{(s)} + \delta) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} 1\{x_{j_x-1} < y_{j_y} + \delta\} P(X_{(r)} \in I_{j_x}^x) P(Y_{(s)} \in I_{j_y}^y) \quad (3.9)$$

The NPI lower and upper probabilities only change if δ is large enough to change the ordering of the $y_1, \ldots, y_{n_x}$ among the values $x_1, \ldots, x_{n_x}$. This change in ordering can occur for at most $n_x \times n_y$ different values of δ. Thus the lower and upper probabilities of the event that $X_{(r)} < Y_{(s)} + \delta$ for fixed r and s can have at most $n_x \times n_y + 1$ different values, including the case $\delta = 0$. As a function of δ, $\underline{P}(X_{(r)} < Y_{(s)} + \delta)$ and $\overline{P}(X_{(r)} < Y_{(s)} + \delta)$ are step functions that change value at $n_x \times n_y$ different values of δ.

Example 3.3. To illustrate the NPI lower and upper probabilities (3.8) and (3.9), we consider the data set given in Example 3.1, about the effect of ozone environment on rats growth. We consider $m = 10$ future observations for both groups. The smallest Y observation, -16.9, has again been deleted from the

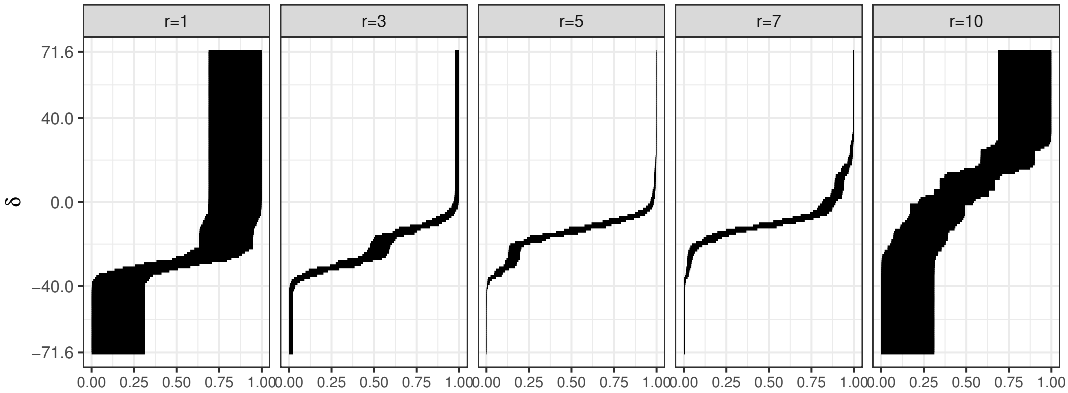

Figure 3.8: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(r)} + \delta)$, $m = 10$

data. Figure 3.8 presents the NPI lower and upper probabilities for the event $X_{(r)} < Y_{(r)} + \delta$, as a function of δ, for $r = 1, 3, 5, 7, 10$. Figures 3.9 and 3.10 present the NPI lower and upper probabilities for the event $X_{(r)} < Y_{(s)} + \delta$, again as function of δ, for $(r = 1, s = 3)$, $(r = 1, s = 5)$, $(r = 1, s = 7)$, $(r = 1, s = 10)$ and $(r = 3, s = 5)$, in Figure 3.9, and for $(r = 3, s = 7)$, $(r = 3, s = 10)$, $(r = 5, s = 7)$, $(r = 5, s = 10)$ and $(r = 7, s = 10)$ in Figure 3.10.

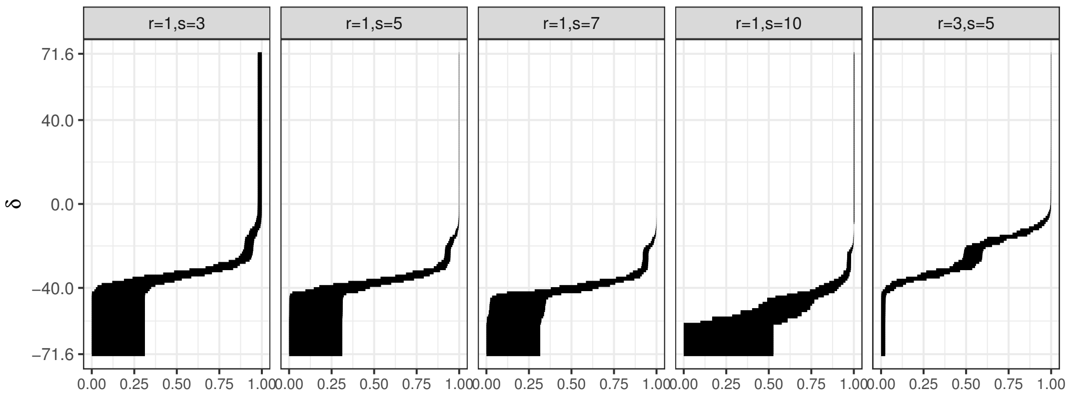

Figure 3.9: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(s)} + \delta)$, $m = 10$

These NPI lower and upper probabilities are monotonically increasing as the value of δ increases. For these data, the lower and upper probabilities remain constant for values of δ less than -57 or greater than 42; in these cases, the two data sets are completely non-overlapping as all observations for group Y become less than all observations for group X or vice versa. Thus, the lower probabilities for this event are equal to 0 for $\delta < -57$, while the upper probabilities for this

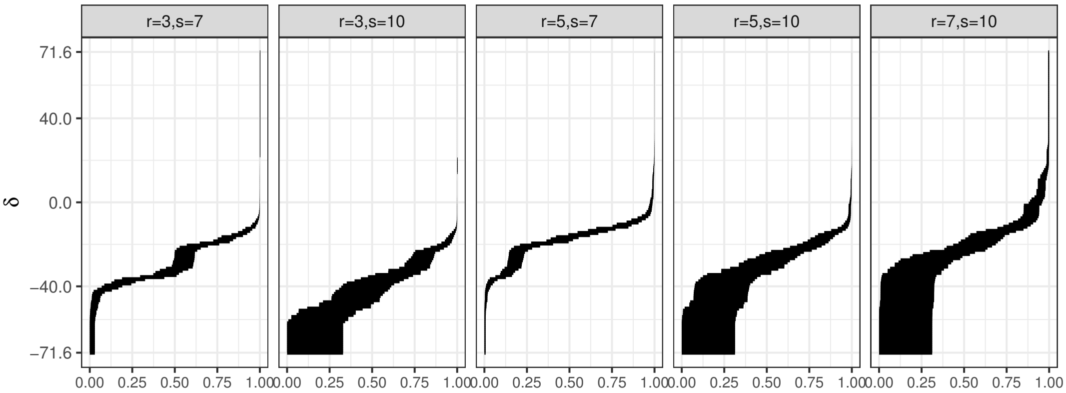

Figure 3.10: $[\underline{P}, \overline{P}](X_{(r)} < Y_{(s)} + \delta)$, $m = 10$

event are equal to 1 for $\delta > 42$. Figure 3.8 illustrates that events involving the extreme future order statistics have high imprecision. $[\underline{P}, \overline{P}](X_{(1)} < Y_{(1)} + \delta)$ increase as functions of δ, staring when the largest observations in group Y begin to exceed the smallest observations in group X. Figures 3.9 and 3.10 illustrate that events involving the extreme future order statistics, i.e. $X_{(1)}$ or $X_{(10)}$ or both, have less imprecision for nearly all δ, apart from small values of δ, i.e. $\delta < -40$. For $\delta > -8.6$, the plotted line, which represents the interval bounded by lower and upper probabilities for the event $X_{(1)} < Y_{(10)} + \delta$ is not shown in Figure 3.9, where the NPI lower and upper probabilities for $X_{(1)} < Y_{(10)} + \delta$ are close to one. Similarly for $\delta > 30$, the plotted line for the lower and upper probabilities for the event $X_{(3)} < Y_{(7)} + \delta$ is not shown in Figure 3.10.

$\Diamond$

3.4 Multiple comparisons

Coolen and van der Laan [30] presented NPI methods for comparisons of multiple groups with different events of interest formulated in terms of the next future observation from each group. This included selecting the best group, the subset of the best groups, and the subset that includes the best group. In this section, we present similar NPI multiple comparisons methods but based on order statistics of multiple future observations. We are interested in the question, which group will give the maximum r-th ordered future observation or the minimum r-th ordered future observation? An answer to this question might be useful if one wants to choose between the groups.

3.4.1 Selecting the best group

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

$$\begin{aligned}
P(X_{(r)}^l = \max_{1 \leq g \leq k} X_{(r)}^g) &= P(\bigcap_{\substack{g=1 \\ g \neq l}}^{k} \{X_{(r)}^l > X_{(r)}^g\} \\
&= \sum_{j_l=1}^{n_l+1} P(\bigcap_{\substack{g=1 \\ g \neq l}}^{k} \{X_{(r)}^g < X_{(r)}^l\} | X_{(r)}^l \in I_{j_l}^l) P(X_{(r)}^l \in I_{j_l}^l) \\
&\geq \sum_{j_l=1}^{n_l+1} P(\bigcap_{\substack{g=1 \\ g \neq l}}^{k} \{X_{(r)}^g < x_{j_l-1}^l\}) P(X_{(r)}^l \in I_{j_l}^l) \\
&= \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1 \\ g \neq l}}^{k} P(X_{(r)}^g < x_{j_l-1}^l) P(X_{(r)}^l \in I_{j_l}^l) \\
&\geq \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1 \\ g \neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x_{j_g}^g < x_{j_l-1}^l\} P(X_{(r)}^g \in I_{j_g}^g) P(X_{(r)}^l \in I_{j_l}^l)
\end{aligned} \tag{3.10}$$

The lower probability is derived by putting the probability mass per interval at the left end point for the group l, and at the right end point for all other groups. By placing the probability mass for each interval this way, the RHS of Equation (3.10), can actually be attained so it is the maximum lower bound for the probability for this event, hence it is the NPI lower probability. We use the notation

$$\underline{P}_l = \underline{P}(X^l_{(r)} = \max_{1\leq g\leq k} X^g_{(r)}) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1\\g\neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} < x^l_{j_l-1}\} P(X^g_{(r)} \in I^g_{j_g}) P(X^l_{(r)} \in I^l_{j_l}) \tag{3.11}$$

We can similarly derive the NPI upper probability for the event $X^l_{(r)} = \max_{1\leq g\leq k} X^g_{(r)}$ by

$$\begin{aligned}
P(X^l_{(r)} = \max_{1\leq g\leq k} X^g_{(r)}) &= P(\bigcap_{\substack{g=1\\g\neq l}}^{k} \{X^l_{(r)} > X^g_{(r)}\} \\
&= \sum_{j_l=1}^{n_l+1} P(\bigcap_{\substack{g=1\\g\neq l}}^{k} \{X^g_{(r)} < X^l_{(r)}\} | X^l_{(r)} \in I^l_{j_l}) P(X^l_{(r)} \in I^l_{j_l}) \\
&\leq \sum_{j_l=1}^{n_l+1} P(\bigcap_{\substack{g=1\\g\neq l}}^{k} \{X^g_{(r)} < x^l_{j_l}\}) P(X^l_{(r)} \in I^l_{j_l}) \\
&\leq \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1\\g\neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} < x^l_{j_l}\} P(X^g_{(r)} \in I^g_{j_g}) P(X^l_{(r)} \in I^l_{j_l})
\end{aligned} \tag{3.12}$$

The upper bound is obtained by putting the probability mass per interval at the right end point for group l, and at left end point for all other groups. This upper bound can also be attained and hence is the NPI upper probability, and we use the notation

$$\overline{P}_l \;=\; \overline{P}(X^l_{(r)} = \max_{1\leq g\leq k} X^g_{(r)}) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1\\g\neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} < x^l_{j_l}\} P(X^g_{(r)} \in I^g_{j_g}) P(X^l_{(r)} \in I^l_{j_l}) \tag{3.13}$$

The lower and upper probabilities for the event that the r-th future observation from group X^l is the minimum of all r-th future observations from the other groups

$X^g_{(r)}$, $1 \leq g \leq k$, so $X^l_{(r)} = \min_{1\leq g\leq k} X^g_{(r)}$, are derived similarly and are equal to

$$\underline{P}'_l = \underline{P}(X^l_{(r)} = \min_{1\leq g\leq k} X^g_{(r)}) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1\\g\neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} > x^l_{j_l}\} P(X^l_{(r)} \in I^l_{j_l}) P(X^g_{(r)} \in I^g_{j_g}) \tag{3.14}$$

$$\overline{P}'_l = \overline{P}(X^l_{(r)} = \min_{1\leq g\leq k} X^g_{(r)}) = \sum_{j_l=1}^{n_l+1} \prod_{\substack{g=1\\g\neq l}}^{k} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} > x^l_{j_l-1}\} P(X^l_{(r)} \in I^l_{j_l}) P(X^g_{(r)} \in I^g_{j_g}) \tag{3.15}$$

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 r-th ordered future observation for each group. We present examples to illustrate these lower and upper probabilities in Subsection 3.4.4.

3.4.2 Selecting a subset containing only the best groups

In the theory of statistical selection [30, 82] 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 [13]. We consider a subset of the k independent groups with a subset containing w groups, with $1 \leq w \leq k-1$, where $w = 1$ is the case of a single group as presented in Section 3.4.1. One logical problem formulation involves selecting a subset of the groups that is most likely to contain all the w best groups. We now derive the NPI method for such inferences with the best group in terms of the value of the r-th ordered value from the m future observations. Suppose that a subset of the k independent groups contains w groups with $1 \leq w \leq k-1$. Let $S = \{l_1, \ldots, l_w\} \subset \{1, \ldots, k\}$ be the subset of indices of these w groups, and let $NS = \{1, \ldots, k\}/S$ be the subset of indices of the $k-w$ groups not in this subset. The NPI lower and upper probabilities for the event $\min_{l\in S} X^l_{(r)} > \max_{g\in NS} X^g_{(r)}$ are derived as follows. First, we find a lower bound for this probability

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

This lower bound is obtained by putting the probability mass per interval at the left end point for all w groups in the subset and at the right end point for the other $k-w$ groups. This construction makes clear that this is the maximum lower bound, and hence the NPI lower probability which we denote by

$$
\begin{aligned}
\underline{P}_S &= \underline{P}(\min_{l\in S} X^{l}_{(r)} > \max_{g\in NS} X^{g}_{(r)})\\
&= \sum_{j_{l_1}=1}^{n_{l_1}+1}\cdots\sum_{j_{l_w}=1}^{n_{l_w}+1}\prod_{g\in NS}\sum_{j_g=1}^{n_g+1} 1\{x^{g}_{j_g} < \min_{l\in S} x^{l_s}_{j_{l_s}-1}\}P(X^{g}_{(r)} \in I^{g}_{j_g})\\
&\qquad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}},\ldots,X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \qquad (3.16)
\end{aligned}
$$

Because the groups are assumed to be independent, the joint probability $P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}},\ldots,X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}})$ is equal to the product of the factors $P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}})$ to $P(X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}})$. The corresponding upper bound for the event $\min_{l\in S} X^{l}_{(r)} > \max_{g\in NS} X^{g}_{(r)}$ is derived similarly, leading to the upper probability

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

(3.17)

The lower and upper probabilities for the event that the r-th ordered future observation from each group in the subset is less than the r-th ordered future observation from all groups not in the subset, such that $\max_{l\in S} X^l_{(r)} < \min_{g\in NS} X^g_{(r)}$, are similarly derived and are equal to

$$\underline{P}'_S = \underline{P}(\max_{l\in S} X^l_{(r)} < \min_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} > \max_{l\in S} x^{l_s}_{j_{l_s}}\} \times P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{3.18}$$

$$\overline{P}'_S = \overline{P}(\max_{l\in S} X^l_{(r)} < \min_{g\in NS} X^g_{(r)}) = \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g\in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} > \max_{l\in S} x^{l_s}_{j_{l_s}-1}\} \times P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \tag{3.19}$$

We will refer to these as the NPI lower and upper probabilities for a subset S containing only the best groups. These inferences will be illustrated in Subsection 3.4.4.

3.4.3 Selecting a subset including the best group

A second common group selection problem for which classical statistical methods have been presented aims to have the selected subset containing the single best group. In our case, this is the group which provides the maximum r-th ordered future observation. We can use the same notation as introduced for the selection of a subset containing all the best groups. The NPI lower and upper probabilities for the event that the r-th future observation from at least one of the selected groups in S is greater than the r-th future observation from all unselected groups in NS, are derived similarly to the NPI lower and upper probabilities (3.16) and (3.17), but with the minimum over S everywhere replaced by the maximum over S in the event of interest. These NPI lower and upper probabilities are

$$
\begin{aligned}
\underline{P}^*_S &= \underline{P}(\max_{l \in S} X^l_{(r)} > \max_{g \in NS} X^g_{(r)}) \\
&= \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g \in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} < \max_{l \in S} x^{l_s}_{j_{l_s}-1}\} P(X^g_{(r)} \in I^g_{j_g}) \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \qquad (3.20) \\
\overline{P}^*_S &= \overline{P}(\max_{l \in S} X^l_{(r)} > \max_{g \in NS} X^g_{(r)}) \\
&= \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g \in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} < \max_{l \in S} x^{l_s}_{j_{l_s}}\} P(X^g_{(r)} \in I^g_{j_g}) \\
&\quad \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \qquad (3.21)
\end{aligned}
$$

Similarly, the NPI lower and upper probabilities for the event that the r-th future observation from at least one of the selected groups in the subset S is less than the r-th future observation from all nonselected groups in NS, are

$$
\begin{aligned}
\underline{P}^{*\prime}_S = \underline{P}(\min_{l \in S} X^l_{(r)} < \min_{g \in NS} X^g_{(r)}) &= \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g \in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g-1} > \min_{l \in S} x^{l_s}_{j_{l_s}}\} \times \\
&\quad P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \qquad (3.22) \\
\overline{P}^{*\prime}_S = \overline{P}(\min_{l \in S} X^l_{(r)} < \min_{g \in NS} X^g_{(r)}) &= \sum_{j_{l_1}=1}^{n_{l_1}+1} \cdots \sum_{j_{l_w}=1}^{n_{l_w}+1} \prod_{g \in NS} \sum_{j_g=1}^{n_g+1} 1\{x^g_{j_g} > \min_{l \in S} x^{l_s}_{j_{l_s}-1}\} \times \\
&\quad P(X^g_{(r)} \in I^g_{j_g}) \times P(X^{l_1}_{(r)} \in I^{l_1}_{j_{l_1}}, \ldots, X^{l_w}_{(r)} \in I^{l_w}_{j_{l_w}}) \qquad (3.23)
\end{aligned}
$$

These results are also illustrated in Subsection 3.4.4.

The results presented in Subsections 3.4.2 and 3.4.3 can be used in a variety of ways. For example, one may want to select a subset of minimum size for which $\underline{P}_S > 0.5$, or from all subsets of the same size, one may want to select the subset with the largest lower probability for the event of interest. 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.

3.4.4 Examples

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

Example 3.4. 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 [30] with sample sizes $n_1 = 20$, $n_2 = 18$, $n_3 = 15$ and $n_4 = 3$, as presented in Table 3.2.

Group	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 3.2: Ordered data for Example 3.4

The NPI lower and upper probabilities for group $l = 1, \ldots, 4$ to be best, in terms of providing the largest value of the r-th ordered observation from a future sample of size $m = 5$, for each group, are presented in Table 3.3, so these are $\underline{P}_l$ and $\overline{P}_l$ as given in Equations (3.11) and (3.13). These NPI lower and upper probabilities are also presented, for the case with $m = 10$ and all $r = 1, \ldots, 10$, in Figure 3.11. 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.

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

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 3.3: $\underline{P}_l$ and $\overline{P}_l$ for $m = 5$

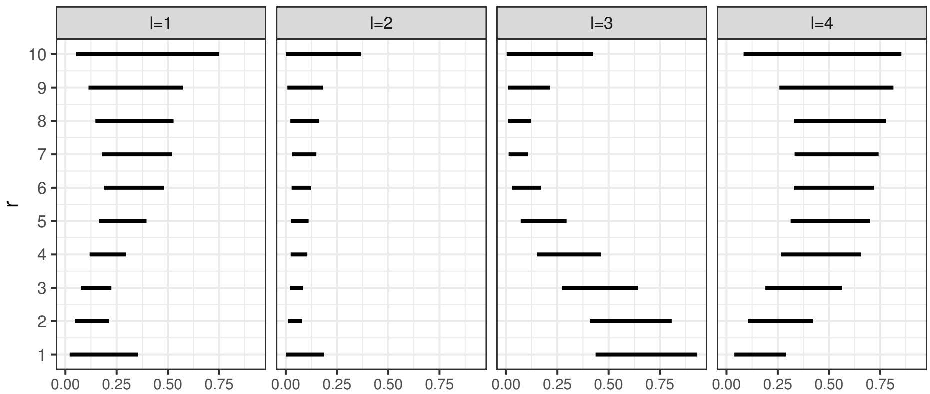

Figure 3.11: $\underline{P}_l$ and $\overline{P}_l$ for $m = 10$

upper probabilities in this case by $\underline{P}_l^{(-4)}$ and $\overline{P}_l^{(-4)}$, they are presented in Table 3.4 for $m = 5$ and in Figure 3.12 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 3.12, 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 may result 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 in the sense of Lawless and Fredette [66], but it only provides inferences in terms of lower and upper probabilities. Hence, one can consider the level of imprecision as a reflection of quality of the statistical inferences, which remain exactly calibrated both with and without inclusion of Group 4 in this

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 3.4: $\underline{P}_l^{(-4)}$ and $\overline{P}_l^{(-4)}$ for $m = 5$

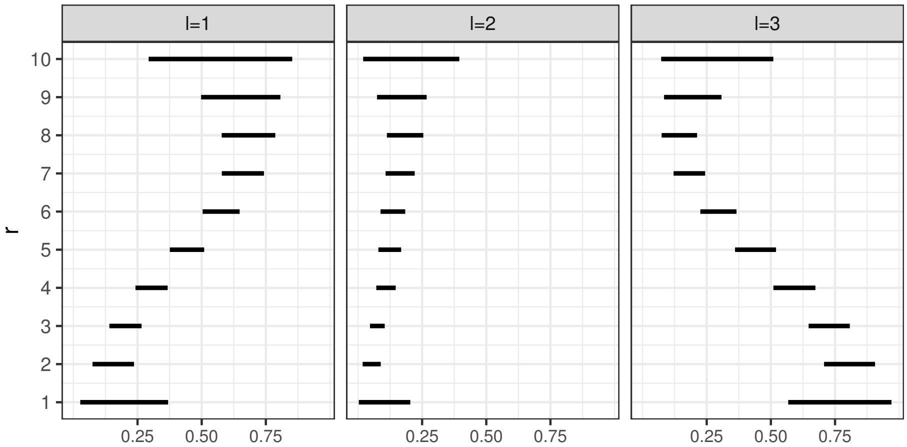

Figure 3.12: $\underline{P}_l^{(-4)}$ and $\overline{P}_l^{(-4)}$ for $m = 10$

example, but less imprecision provides more insight.

Figure 3.13 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 $r = 1, \dots, 10$. So the events considered are $X^l_{(r)} > X^g_{(r)}$ for $l, g = 1, \dots, 4$ 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 as discussed in Chapter 1. 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.

To illustrate subset selection, we first consider the event that the subset S

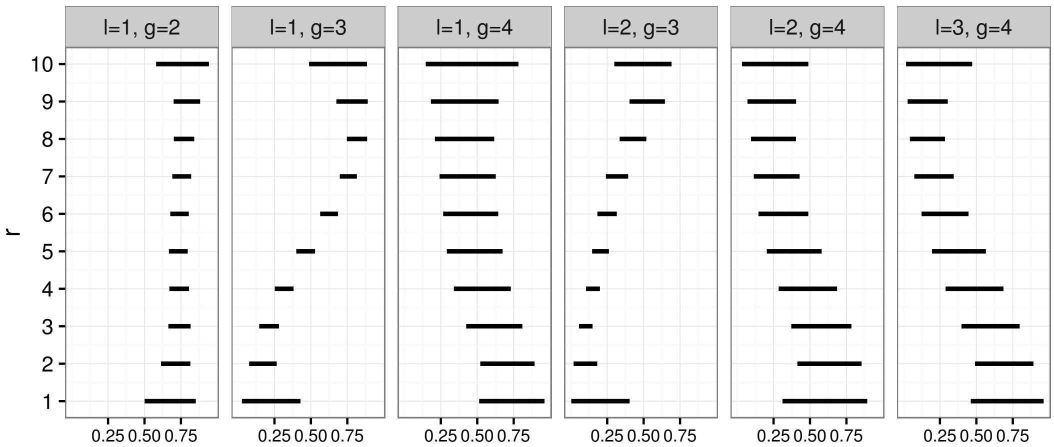

Figure 3.13: $\underline{P}$ and $\overline{P}$ for $m = 10$

contains all the best groups, as presented in Subsection 3.4.2. Figure 3.14 and Table 3.5 present the NPI lower and upper probabilities, which are given in Equations (3.16) and (3.17), for any subset S containing two of the four groups, for $m = 5$ in Table 3.5 and $m = 10$ in Figure 3.14. These results show that the subset that contains groups 1 and 3 is most likely to provide the largest future value for $r = 1$; however, it quite unlikely to provide the largest future value for $r > m/2$, which is most likely to come from the subset that contains groups 1 and 4.

We also illustrate the selection of subsets containing the best group, as presented in Subsection 3.4.3 and for which the NPI lower and upper probabilities are given by Equations (3.20) and (3.21). The NPI lower and upper probabilities for any subset S consisting of two of the four groups are shown for $m = 5$ in Table 3.6, for $m = 10$ in the first row in Figure 3.15, and for $m = 100$ in the second row in this figure. 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 three of the four 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 3.3 for $m = 5$ and in Figure 3.11 for $m = 10$.

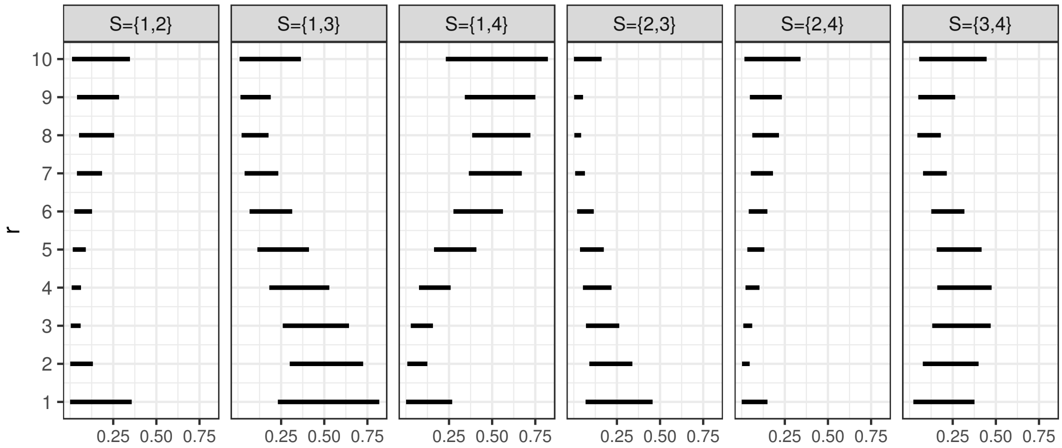

Figure 3.14: $\underline{P}_S$ and $\overline{P}_S$ for $m = 10$

$S:$	$\{1,2\}$		$\{1,3\}$		$\{1,4\}$		$\{2,3\}$		$\{2,4\}$		$\{3,4\}$	
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	0.0023	0.2142	0.2613	0.6971	0.0106	0.1751	0.1003	0.3782	0.0033	0.0868	0.0651	0.3750
2	0.0134	0.0858	0.2014	0.5210	0.0690	0.2311	0.0757	0.2511	0.0248	0.1020	0.1451	0.4230
3	0.0304	0.1235	0.0948	0.3362	0.2074	0.4458	0.0361	0.1531	0.0518	0.1591	0.1432	0.3407
4	0.0541	0.2118	0.0388	0.2096	0.3282	0.6159	0.0107	0.0719	0.0739	0.2125	0.0822	0.2248
5	0.0359	0.2789	0.0196	0.2402	0.2855	0.7050	0.0041	0.0926	0.0521	0.2744	0.0683	0.3227

Table 3.5: $\underline{P}_S$ and $\overline{P}_S$ for $m = 5$

$S:$	$\{1,2\}$		$\{1,3\}$		$\{1,4\}$		$\{2,3\}$		$\{2,4\}$		$\{3,4\}$	
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 3.6: $\underline{P}^*_S$ and $\overline{P}^*_S$ for $m = 5$

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.

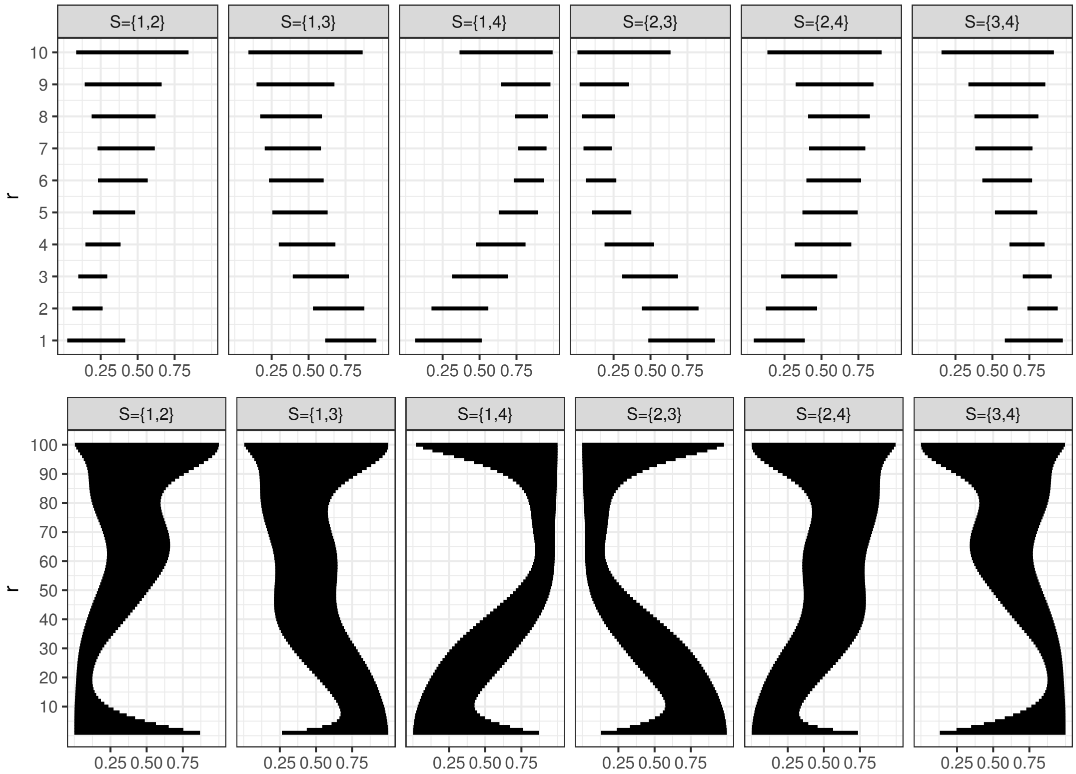

Figure 3.15: $\underline{P}_S^*$ and $\overline{P}_S^*$ for $m = 10$ and $m = 100$

Cloth l	x_1^l	x_2^l	x_3^l	x_4^l
1	337	344	369	396
2	520	537	602	627
3	233	240	251	278
4	196	211	248	273
5	160	185	195	199
6	442	563	595	606
7	226	252	297	300

Table 3.7: Ordered weight losses (in 0.1 mg).

Example 3.5. The data for this example are presented in Table 3.7; the data are from Bechhofer et al. [13, p. 84] and were also used by Box et al. [18, p. 277] and by Coolen and van der Laan [30]. The data are from an experiment conducted to determine the weight loss for seven types of cloth during a mechanical test. The measured weight loss is in tenths of a milligram for each cloth following 1000 revolutions on a wear tester.

Based on the given data, suppose we want to choose the cloth with the smallest weight loss at the r-th ordered future observation. The NPI lower and

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	0.55556	0	0.55556	0	0.55580	0.00008	0.56975
2	0	0.27778	0	0.27778	0	0.28485	0.00217	0.34172
3	0	0.11922	0	0.11906	0	0.16437	0.00752	0.25740
4	0	0.04540	0	0.04070	0	0.18911	0.00900	0.29025
5	0	0.10244	0	0.04485	0	0.38214	0.00424	0.49790

Table 3.8: $\underline{P}'_l$ and $\overline{P}'_l$ for $m = 5$

r	$\underline{P}'_5$	$\overline{P}'_5$	$\underline{P}'_6$	$\overline{P}'_6$	$\underline{P}'_7$	$\overline{P}'_7$
1	0.00746	0.99559	0	0.55556	0	0.55593
2	0.12836	0.98898	0	0.27778	0	0.28315
3	0.38170	0.98583	0	0.11906	0	0.14221
4	0.53352	0.98898	0	0.04093	0	0.10527
5	0.41310	0.99559	0	0.05454	0	0.21775

Table 3.9: $\underline{P}'_l$ and $\overline{P}'_l$ for $m = 5$

upper probabilities $\underline{P}'_l$ and $\overline{P}'_l$ for the event that cloth l is the best, for $l = 1, \ldots, 7$, as given in Equations (3.14) and (3.15), are displayed in Tables 3.8 and 3.9 for $m = 5$ and $r = 1, \ldots, 5$, and in Figure 3.16 for $m = 5$, $m = 10$ and $m = 100$, and for all $r = 1, \ldots, m$. The results clearly indicate that cloths 2 and 6 are unlikely to lead to the best r-th ordered future observation, with the upper probabilities for cloth 6 slightly greater than for cloth 2 for large values of r. Obviously, in this example, there is a strong suggestion that cloth 5 is the best, as it provides the highest $\underline{P}'_l$ and $\overline{P}'_l$ that the r-th future ordered statistic is the minimum for all r-th future order statistics. The rather large difference between the lower and upper probabilities reflects the fact that there are only few observations available. All the lower probabilities that are equal to 0 in Tables 3.8 and 3.9 are caused by all the related observations being larger than all the observations for cloth 5.

Table 3.10 presents the NPI lower and upper probabilities for some subsets, $\underline{P}'_S$ and $\overline{P}'_S$, for the event that the subset S contains all the best cloths, as given by Equations (3.18) and (3.19), for $m = 5$. These NPI lower and upper probabilities are also presented for the case with $m = 10$ in Figure 3.17. These results illustrate that $\overline{P}'_{\{3,4,5,7\}} = 1$, because all observations for these four cloths are greater than all the observations for cloths 1,2 and 6. If we select cloth 1 instead of cloth 7, we

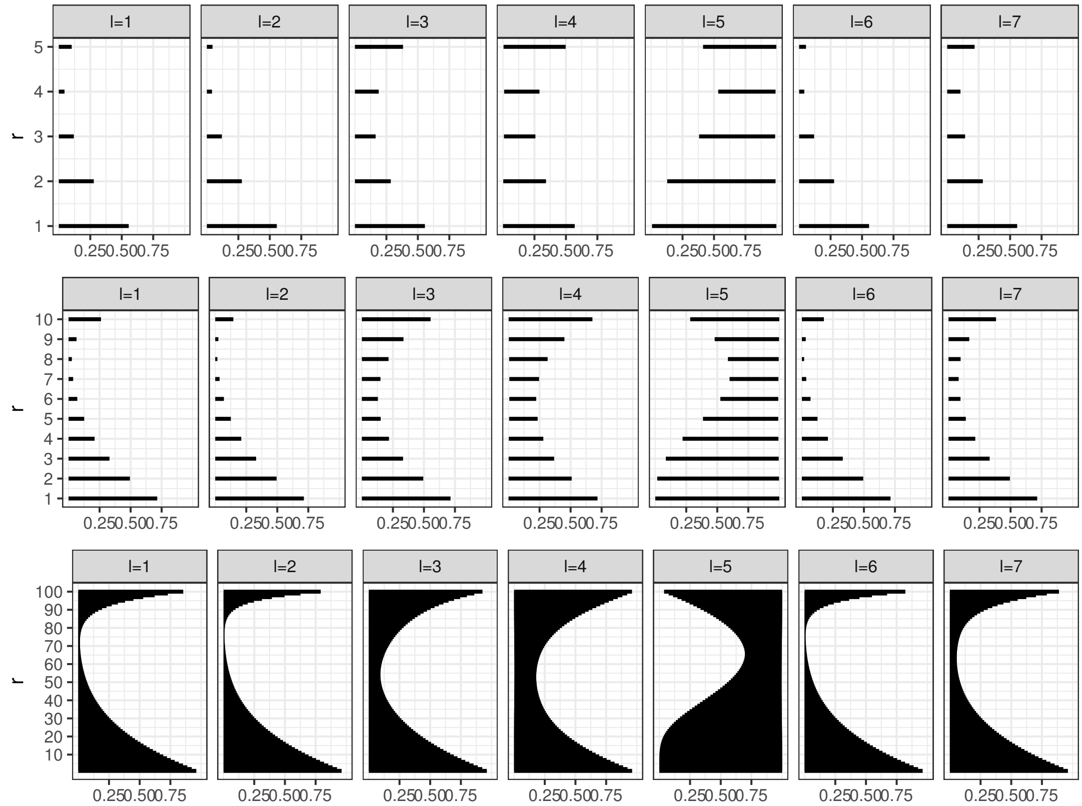

Figure 3.16: $\underline{P}'_l$ and $\overline{P}'_l$ for $l = 1, \ldots, 7$, $m = 5, 10, 100$

get $\underline{P}'_{\{1,3,4,5\}} = 0$ as all the observations in cloth 7 are less than those for cloth 1.

Tables 3.11 and Figure 3.18 present some of the NPI lower and upper probabilities for the event that the subset S contains the cloth that provides the smallest r-th ordered future observation, as given in Equations (3.22) and (3.23), for $m = 5$ in Table 3.11 and for $m = 10$ in Figure 3.18. The upper probabilities $\overline{P}^{*\prime}_S$ are all equal to one as group 5 is included in all these subsets. The results from Table 3.11 and Figure 3.18 can be used to select the best subset. Suppose we want to select a subset with minimum size for which $\underline{P}^{*\prime}_S > 0.5$. Table 3.11 shows that a subset contains cloths 4 and 5 is the best for $r = 4$ and $r = 5$ with $\underline{P}^{*\prime}_S > 0.5$. From all subsets of the size three, suppose we want to select the subset with maximum lower probability for the event $\min_S X^l_{(r)} < \min_{NS} X^g_{(r)}$. Table 3.11 gives that $S = \{3, 4, 5\}$ is the best for $r = 4$ with $\underline{P}^{*\prime}_S = 0.7924$.

$S:$	$\{4,5\}$		$\{3,4,5\}$		$\{4,5,7\}$		$\{3,4,5,7\}$		$\{1,3,4,5\}$	
r	$\underline{P}'_S$	$\overline{P}'_S$	$\underline{P}'_S$	$\overline{P}'_S$	$\underline{P}'_S$	$\overline{P}'_S$	$\underline{P}'_S$	$\overline{P}'_S$	$\underline{P}'_S$	$\overline{P}'_S$
1	0.0144	0.9554	0.0135	0.7340	0.0184	0.8055	0.0850	1	0	0.4113
2	0.1184	0.8731	0.1154	0.7618	0.0521	0.5769	0.3204	1	0	0.2403
3	0.2125	0.8050	0.2381	0.8367	0.0469	0.4440	0.4118	1	0	0.2045
4	0.2074	0.7966	0.2235	0.9002	0.0262	0.4462	0.2410	1	0	0.3025
5	0.1171	0.8700	0.0729	0.9602	0.0053	0.6047	0.0381	1	0	0.5588

Table 3.10: $\underline{P}'_S$ and $\overline{P}'_S$ for $m = 5$

$S:$	$\{4,5\}$		$\{3,4,5\}$		$\{4,5,7\}$		$\{3,4,5,7\}$		$\{1,3,4,5\}$	
r	$\underline{P}^*_S$	$\overline{P}^{*\prime}_S$	$\underline{P}^{*\prime}_S$	$\overline{P}^{*\prime}_S$	$\underline{P}^{*\prime}_S$	$\overline{P}^{*\prime}_S$	$\underline{P}^{*\prime}_S$	$\overline{P}^{*\prime}_S$	$\underline{P}^{*\prime}_S$	$\overline{P}^{*\prime}_S$
1	0.0173	1	0.0390	1	0.0390	1	0.0878	1	0.0877	1
2	0.1936	1	0.2700	1	0.2694	1	0.3767	1	0.3739	1
3	0.4962	1	0.5865	1	0.5713	1	0.6835	1	0.6657	1
4	0.6696	1	0.7924	1	0.7181	1	0.8803	1	0.8251	1
5	0.5735	1	0.7638	1	0.6033	1	0.8834	1	0.7699	1

Table 3.11: $\underline{P}^{*\prime}_S$ and $\overline{P}^{*\prime}_S$ for $m = 5$

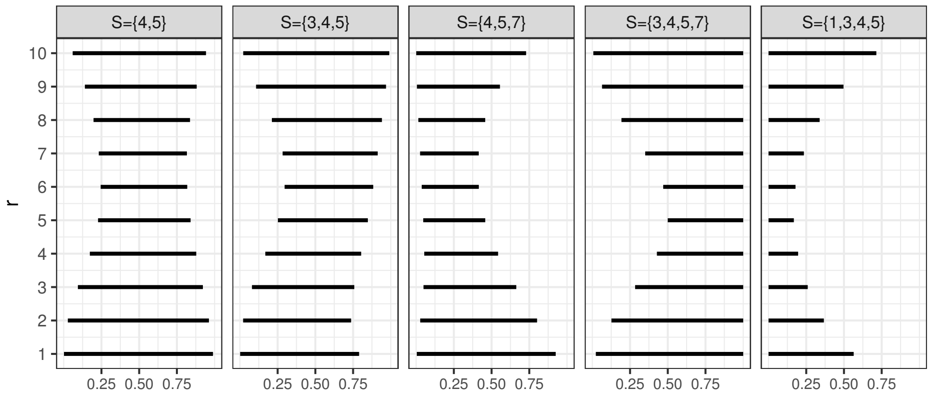

Figure 3.17: $\underline{P}'_S$ and $\overline{P}'_S$ for $m = 10$

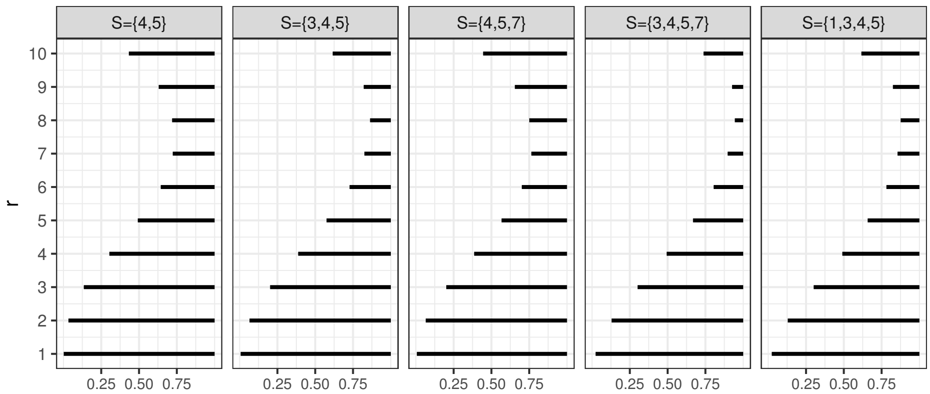

Figure 3.18: $\underline{P}^{*\prime}_S$ and $\overline{P}^{*\prime}_S$ for $m = 10$

◊

3.5 Further inferences

The NPI methods for future order statistics presented in this thesis 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 can consider NPI for prediction intervals [44], for the number of future order statistics in an interval [3, 49], and for spacings between order statistics [3]. These are briefly discussed next.

3.5.1 Prediction intervals

In many practical situations, it may be of interest to use the observations from the original sample to construct an interval which contains some order statistics of future observations with a certain probability [2]. Such an interval is called a prediction interval 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 an interval of the future order statistics $[X_{(r)}, X_{(s)}]$ for $r < s$. The corresponding predictive probability is easily computed using Equation (2.6), and is

$$P(x_a < X_{(r)} < X_{(s)} < x_b) = \sum_{j=a+1}^{b} \sum_{l=j}^{b} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \tag{3.24}$$

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

$$P(X_{(r)} < x_c < x_d < X_{(s)}) = \sum_{j=1}^{c} \sum_{l=d+1}^{n+1} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \tag{3.25}$$

One may typically be interested in the shortest outer interval, or the longest inner interval, for which the corresponding probability (3.24) or (3.25) 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. The core problem of the prediction intervals fits well within the NPI

					b				
a	1	2	3	4	5	6	7	8	9
0	0.00078	0.00466	0.01632	0.04351	0.09790	0.19580	0.35897	0.61538	1
1	0	0.00155	0.00855	0.02797	0.07071	0.15229	0.29371	0.52214	0.87179
2	0	0	0.00233	0.01243	0.03963	0.09790	0.20668	0.39161	0.68531
3	0	0	0	0.00311	0.01632	0.05128	0.12510	0.26107	0.48951
4	0	0	0	0	0.00389	0.02020	0.06294	0.15229	0.31546
5	0	0	0	0	0	0.00466	0.02409	0.07459	0.17949
6	0	0	0	0	0	0	0.00544	0.02797	0.08625
7	0	0	0	0	0	0	0	0.00622	0.03186
8	0	0	0	0	0	0	0	0	0.00699

Table 3.12: $P(x_a < X_{(2)} < X_{(5)} < x_b)$ for $n = 8$ and $m = 5$

framework for the future order statistics, as presented in this thesis, as it explicitly focuses on the future order statistics based on the past data observations.

Example 3.6. Suppose we have $n = 8$ observations and $m = 5$ future observations. Table 3.12 presents the NPI probabilities for the event that the outer prediction intervals (x_a, x_b) contains an interval of the future order statistics $[X_{(2)}, X_{(5)}]$, with $a < b$, $a = 0, \dots, 8$ and $b = 1, \dots, 9$. Comparing Table 3.12 with Table 2.3 shows that $P(x_{j-1} < X_{(2)} < X_{(5)} < x_j) = P(X_{(2)} \in I_j, X_{(5)} \in I_j)$ for $j = 1, \dots, 9$. Suppose we want to find the shortest outer interval for which $P(x_a < X_{(2)} < X_{(5)} < x_b) > 0.5$. Table 3.12 shows that (1,8) and (2,9) are the shortest intervals with $P(x_a < X_{(2)} < X_{(5)} < x_b) > 0.5$, no other interval of length 7 has probability greater than 0.5 for this event. Of course, it may be logical to choose $(2, 9)$ in this case, as the probability for it to contain $[X_{(2)}, X_{(5)}]$ is larger than for $(1, 8)$.

Table 3.13 presents the NPI probabilities for the event that the inner prediction interval (x_c, x_d) is contained in $[X_{(2)}, X_{(5)}]$, for $n = 8$, $m = 5$, $c = 1, \dots, 8$ and $d = 1, 2, \dots, 8$. Suppose one may be interested in the longest inner interval for which $P(X_{(2)} < x_c < x_d < X_{(5)}) > 0.5$, then from Table 3.13, the longest inner interval is $(4, 6)$ for which $P(X_{(2)} < x_4 < x_6 < X_{(5)}) = 0.50894$.

$\Diamond$

	d							
c	1	2	3	4	5	6	7	8
1	0.12743	0.12510	0.12044	0.11267	0.10101	0.08469	0.06294	0.03497
2	0	0.31002	0.30070	0.28361	0.25641	0.21678	0.16239	0.09091
3	0	0	0.49417	0.47009	0.42890	0.36597	0.27661	0.15618
4	0	0	0	0.64103	0.59052	0.50894	0.38850	0.22145
5	0	0	0	0	0.72261	0.62937	0.48563	0.27972
6	0	0	0	0	0	0.71795	0.56022	0.32634
7	0	0	0	0	0	0	0.60917	0.35897
8	0	0	0	0	0	0	0	0.37762

Table 3.13: $P(X_{(2)} < x_c < x_d < X_{(5)})$ for $n = 8$ and $m = 5$

3.5.2 Number of future observations in an interval

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

$$P(C_{a,b}^m = u) = \sum_{m_a=0}^{m-u} \frac{\binom{a-1+m_a}{m_a}\binom{b-a-1+u}{u}\binom{n-b+m-u-m_a}{m-u-m_a}}{\binom{n+m}{n}} \tag{3.26}$$

This probability depends only 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

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

Both these expressions are easily derived by combinatorics using the basic probability results presented in Chapter 2. For the special case with $b = n+1$, so considering the interval (x_a, ∞), we have

$$P(C_{a,n+1}^m = u) = \frac{\binom{a-1+m-u}{m-u}\binom{n-a+u}{u}}{\binom{n+m}{n}} \tag{3.27}$$

This result is equal to the distribution of the number of exceedances in the classical theory of statistics [3], although the derivation method differs due to the different starting points of NPI compared with the classical theory. Analysis of this probability (3.27) leads to the logical symmetry

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

for $u = 0, \ldots, m$ and $a = 1, \ldots, n$. This will be illustrated in Example 3.7.

We have the following simpler expressions for some special cases. Let us consider the event $C^m_{a,n+1} = 0$, so no future observations in interval (x_a, ∞), then for $a = 1$

$$P(C^m_{1,n+1} = 0) = \frac{m!n!}{(n+m)!} = P(C^m_{n,n+1} = m)$$

and for $a = n$,

$$P(C^m_{n,n+1} = 0) = \frac{n}{n+m} = P(C^m_{1,n+1} = m)$$

In particular, if $m = n$,

$$P(C^m_{n,n+1} = 0) = \frac{n}{2n} = 0.5 = P(C^m_{1,n+1} = m)$$

A further interesting event is that at least u future observations are in the interval (x_a, x_b), denoted by $C^m_{a,b} \geq u$. The NPI probability for this event is equal to:

$$\begin{aligned} P(C^m_{a,b} \geq u) &= \sum_{i=u}^{m} \sum_{m^i_a=0}^{m-i} P(C^m_{-\infty,a} = m^i_a, C^m_{a,b} = i, C^m_{b,\infty} = m - i - m^i_a) \\ &= \sum_{i=u}^{m} \sum_{m^i_a=0}^{m-i} \frac{\binom{a-1+m^i_a}{m^i_a}\binom{b-a-1+i}{i}\binom{n-b+m-i-m^i_a}{m-i-m^i_a}}{\binom{n+m}{n}} \end{aligned} \qquad (3.28)$$

This probability (3.28) also only depends on the value $b - a$, we can also write it as

$$P(C^m_{a,b} \geq u) = \sum_{i=u}^{m} \frac{\binom{b-a-1+i}{i}\binom{m-i+n-b+a}{m-i}}{\binom{n+m}{n}} \qquad (3.29)$$

An interesting special case of probability (3.28) for the event that at least u future observations belong to the interval $(-\infty, x_b)$

$$P(C^m_{0,b} \geq u) = \sum_{i=u}^{m} \frac{\binom{b-1+i}{i}\binom{n-b+m-i}{m-i}}{\binom{n+m}{n}} = \sum_{j=1}^{b} P(X_{(u)} \in I_j) \qquad (3.30)$$

This gives the same result as the CDF of the probability in Equation (2.2).

The joint probability for the event $X_{(r)} \in (-\infty, x_a), X_{(s)} \in (x_a, x_b)$ is

$$P(X_{(r)} \in (-\infty, x_a), X_{(s)} \in (x_a, x_b)) = \sum_{i_1=r}^{s-1} \sum_{i_2=s}^{m} \frac{\binom{i_1+a-1}{i_1}\binom{i_2-i_1+b-a-1}{i_2-i_1}\binom{m-i_2+n-b}{m-i_2}}{\binom{n+m}{n}} \tag{3.31}$$

This result is equal to the probability in Equation (1.4), given in classical theory of statistics [83], although the derivation method differs due to the different starting points of NPI compared with the classical theory. Where as in classical theory both the data and future observations are considered to be random quantities, predictive inference involves conditioning on the data observations.

We also derive the expected value for the number of future observations in interval (x_a, x_b)

$$E(C_{a,b}^m = u) = \sum_{u=1}^{m} uP(C_{a,b}^m = u) = \frac{(b-a)m}{n+1} \tag{3.32}$$

As a special case, the expected number of future observations in interval $(-\infty, x_b)$, is $\frac{bm}{n+1}$, which is the same result as the classic mean of exceedances derived by Gumbel [49], although they use a different setting than NPI.

Example 3.7. Suppose we have $n = 8$ observations and $m = 5$ future observations. The NPI probabilities for the events that there are exactly u future observations in interval (x_a, x_9) are given in Table 3.14. This table illustrates the logical symmetry $P(C_{a,9}^5 = u) = P(C_{9-a,9}^5 = 5 - u)$ for $a = 1, \ldots, 8$ and $u = 0, \ldots, 5$, so the last row is equal to the second one in reversed order, the last column is equal to the first one in reversed order, and so on.

To illustrate the inference in Equation (3.28), Table 3.15 presents the NPI probabilities for the event that at least one future observation falls in interval (x_a, x_b), i.e. $C_{a,b}^5 \geq 1$, with $a = 0, \ldots, 8$ and $b = 1, \ldots, 9$. Table 3.15 illustrates that $P(C_{a,b}^m \geq u)$ depends only on the value of $b - a$.

Table 3.16 illustrates the probabilities $P(C_{a,b}^5 \geq u)$, as given in Equation (3.29), for $n = 8$ and $m = 5$, where rows represent the number of the intervals $b - a$, and columns represent the number of the future observations u . Comparing Table

			u			
a	0	1	2	3	4	5
0	0	0	0	0	0	1
1	0.00078	0.00622	0.02797	0.09324	0.25641	0.61538
2	0.00466	0.02720	0.08702	0.19580	0.32634	0.35897
3	0.01632	0.06993	0.16317	0.26107	0.29371	0.19580
4	0.04351	0.13598	0.23310	0.27195	0.21756	0.09790
5	0.09790	0.21756	0.27195	0.23310	0.13598	0.04351
6	0.19580	0.29371	0.26107	0.16317	0.06993	0.01632
7	0.35897	0.32634	0.19580	0.08702	0.02720	0.00466
8	0.61538	0.25641	0.09324	0.02797	0.00622	0.00078

Table 3.14: $P(C^5_{a,9} = u)$

					b				
a	1	2	3	4	5	6	7	8	9
0	0.38461	0.64102	0.80419	0.90209	0.95648	0.98368	0.99534	0.99922	1
1	0	0.38461	0.64102	0.80419	0.90209	0.95648	0.98368	0.99533	0.99922
2	0	0	0.38461	0.64102	0.80419	0.90209	0.95649	0.98368	0.99534
3	0	0	0	0.38461	0.64102	0.80419	0.90209	0.95649	0.98368
4	0	0	0	0	0.38461	0.64102	0.80419	0.90209	0.95649
5	0	0	0	0	0	0.38461	0.64102	0.80419	0.90209
6	0	0	0	0	0	0	0.38461	0.64102	0.80419
7	0	0	0	0	0	0	0	0.38461	0.64102
8	0	0	0	0	0	0	0	0	0.38461

Table 3.15: $P(C^5_{a,b} \geq 1)$

3.16 with Table 3.14 shows that $P(C^5_{a,b} \geq 5) = P(C^5_{9-a,9} = 0) = P(C^5_{a,9} = 5)$. Table 3.17 presents $F_r(j) = \sum_{k=1}^{j} P(X_{(r)} \in I_k)$, as given in Equation (2.2), for $n = 8$, $m = 5$, and $r = 1, \ldots, 5$. Comparing Table 3.17 with Tables 2.1 and Table 3.16 illustrates that

$$P(C^5_{j,j+1} \geq u) = P(X_{(u)} \in I_1), \quad u = 1, \ldots, 5$$

$$P(C^5_{0,j} \geq u) = F_u(j), \quad u = 1, \ldots, 5 \quad j = 1, \ldots, 9$$

$\Diamond$

			u		
$b-a$	1	2	3	4	5
1	0.38462	0.12821	0.03497	0.00699	0.00078
2	0.64103	0.31469	0.11888	0.03186	0.00466
3	0.80420	0.51049	0.24942	0.08625	0.01632
4	0.90210	0.68454	0.41259	0.17949	0.04351
5	0.95649	0.82051	0.58741	0.31546	0.09790
6	0.98368	0.91375	0.75058	0.48951	0.19580
7	0.99534	0.96814	0.88112	0.68531	0.35897
8	0.99922	0.99301	0.96503	0.87179	0.61538
9	1	1	1	1	1

Table 3.16: $P(C^5_{a,b} \geq u)$

j	$F_1(j)$	$F_2(j)$	$F_3(j)$	$F_4(j)$	$F_5(j)$
1	0.38462	0.12821	0.03497	0.00699	0.00078
2	0.64103	0.31469	0.11888	0.03186	0.00466
3	0.80420	0.51049	0.24942	0.08625	0.01632
4	0.90210	0.68454	0.41259	0.17949	0.04351
5	0.95649	0.82051	0.58741	0.31546	0.09790
6	0.98368	0.91375	0.75058	0.48951	0.19580
7	0.99534	0.96814	0.88112	0.68531	0.35897
8	0.99922	0.99301	0.96503	0.87179	0.61538
9	1	1	1	1	1

Table 3.17: $F_u(j) = \sum_{k=1}^{j} P(X_{(u)} \in I_k)$

3.5.3 Spacings between order statistics

Spacings between order statistics have also attracted interest [3, p.32], they play a role in many research fields of statistics, for example, goodness of fit tests, reliability analysis and survival analysis [3, 36, 65]. 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 (2.6), 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 $W_{r,s} < w$ for some $w > 0$, the NPI lower and upper probabilities are

$$\underline{P}(W_{r,s} < w) \quad = \quad \sum_{j=1}^{n+1} \sum_{l=j}^{n+1} \mathbf{1}\{x_l - x_{j-1} < w\} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \quad (3.33)$$

$$\overline{P}(W_{r,s} < w) \quad = \quad \sum_{j=1}^{n+1}\sum_{l=j}^{n+1} \mathbf{1}\{x_{l-1} - x_j < w\} P(X_{(r)} \in I_j, X_{(s)} \in I_l) \quad (3.34)$$

Example 3.8. The data in Table 3.18 are birth weights in grams of 44 babies born in one 24-hour period in hospital Brisbane, Australia [78].

1745	2121	2184	2208	2383	2576	2635	2846	2902	3034	3116	3150	3166	3208
3278	3294	3300	3334	3345	3370	3380	3402	3406	3428	3428.1	3430	3480	3500
3520	3521	3523	3542	3554	3625	3630	3690	3736	3746	3783	3837	3838	3866
3920	4162												

Table 3.18: Data set of birth weights in grams

The NPI lower and upper probabilities for the event $W_{r,s} < w$, as a function of w, for $w = 0, \ldots, 2000$ and for $m = 25$ are displayed in Figures 3.19 for some events of r and s, corresponding to the full data. Figures 3.20 shows the same but with the first five observations removed from the data, as these values could be considered to be outliers, thus it might be interesting to see their influence on these inferences. The results illustrate that events involving extreme future order statistics have relatively higher imprecision then events that do not involve extreme future order statistics. Deleting the first five observations from the data, as shown in Figure 3.20, has some effect on inferences especially inferences that involves a small value of r, i.e. $r = 1$ and $r = 7$.

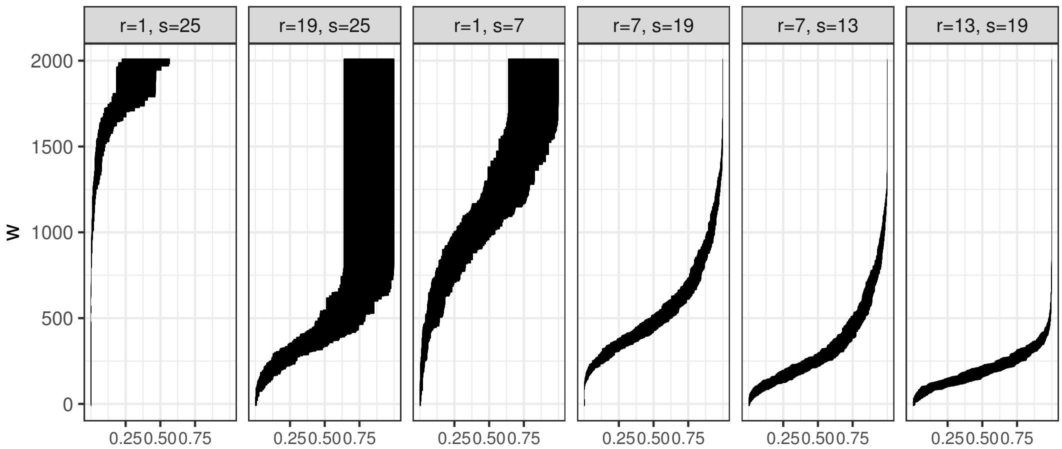

Figure 3.19: $\underline{P}, \overline{P}(W_{r,s} < w)$ for $m = 25, n = 44$

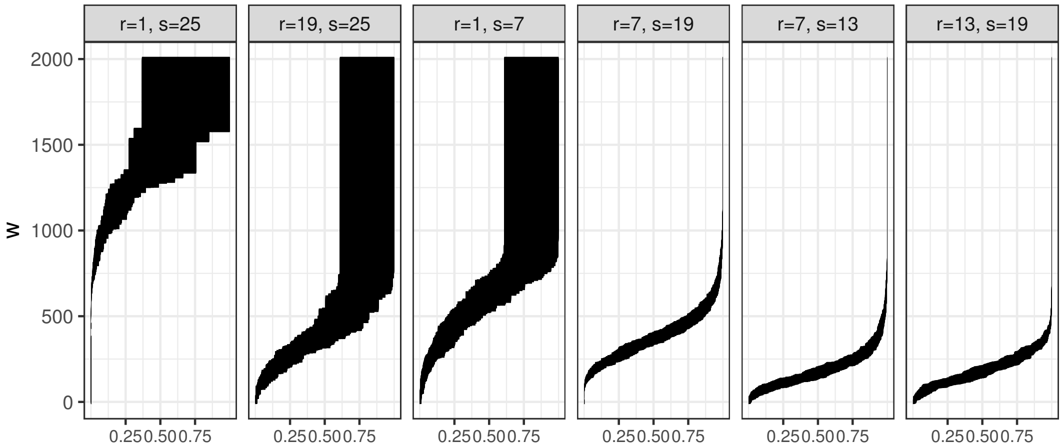

Figure 3.20: $\underline{P}, \overline{P}(W_{r,s} < w)$ for $m = 25, n = 39$

$\Diamond$

3.6 Concluding remarks

This chapter has presented NPI for several events of interest including order statistics, in particular multiple comparisons. We have presented NPI methods for a range of such inferences, and also illustrated the influence of the particular choice of the number m of future observations. The main ideas are similar for other inferences as long as these are explicitly expressed in terms of one or more future order statistics.

A major research challenge is the generalization of NPI for future order statistics in case of lifetime data containing right-censored observations [83], which will enable such methods to be created for many medical and engineering applications. 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 [31], and related results for multiple comparisons have also been presented [34], 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.

Chapter 4

NPI for test reproducibility

4.1 Introduction

Reproducibility of statistical hypothesis tests is an issue of major importance in applied statistics: if the test were repeated, would the same conclusion be reached that is rejection or non-rejection of the null hypothesis? NPI provides a natural framework for such inferences, as its explicitly predictive nature fits well with the core problem formulation of a repeat of the test in the future. For inference on reproducibility of statistical tests, NPI provides lower and upper reproducibility probabilities (RP).

In this chapter, the NPI method for reproducibility of statistical tests is presented for two basic tests using order statistics, namely a test for a specific value for a population quantile and a precedence test for comparison of data from two populations, as typically used for experiments involving lifetime data if one wishes to conclude before all observations are available. Now we use the term 'population' within those classical tests, but we do not use it as a concept in NPI as mentioned in Section 1.3.

Testing of hypotheses is one of the main tools in statistics and crucial in many applications. While many different tests have been developed for a wide range of scenarios, the aspect of reproducibility of tests has long been neglected: the question addressed is whether or not a test, if it were repeated under the same circumstances, would lead to the same overall conclusion with regard to rejection or non-rejection of the null hypothesis. Recently, this topic has started to gain attention, in particular through the publication of a ‘handbook on reproducibility’ [6] which provides a collection of papers on the issue. Nevertheless, whilst hypothesis testing is mainly seen as a frequentist statistics procedure, the classic frequentist framework is not suited for inference on reproducibility as this is neither an estimation nor a testing problem. The very nature of reproducibility is predictive, namely given the result of one test one wishes to predict the outcome of a possible future test. Coolen and Bin Himd [25] presented NPI for reproducibility of some basic tests, with more attention to this topic in the PhD thesis of Bin Himd [16]. These publications also provide a critical discussion of earlier methods for reproducibility presented in the literature, which is briefly considered in Section 4.2. This chapter contributes to development of NPI for reproducibility by considering two tests based on order statistics, namely a one sample quantile test and a two sample precedence test. Central to these inferences are the NPI results for future order statistics presented in Chapter 2.

This chapter is organised as follows. Section 4.2 discusses aspects of reproducibility of statistical tests and explains the NPI perspective on such inferences. Section 4.3 presents the NPI approach to reproducibility of a basic quantile test. Section 4.4 considers a precedence test used for comparison of two populations. Some concluding remarks are given in Section 4.5.

4.2 An overview of test reproducibility

Statistical hypothesis testing is used in many application areas and normally results in either non-rejection of the stated null hypothesis or its rejection in favour of a stated alternative, at a predetermined level of significance. Whilst this procedure is embedded in the successful long-standing tradition of statistics, a related aspect that had received relatively little attention in the literature until recently is the reproducibility of such tests: if the test were repeated, would it lead to the same overall conclusion?

The reproducibility of a statistical test is an issue that was raised by Goodman [48] to address some common misunderstandings regarding the meaning of a statistical p-value. Goodman claimed that the replication probability can be used to show that the p-value may overstate the evidence against the null hypothesis. The reproducibility probability for a test is the probability for the event that, if the test is repeated under similar circumstances as the original experiment, the test result, which is the rejection or not of the null hypothesis, will be the same in the repeated test. The focus is usually on the reproducibility of tests where the null hypothesis is rejected, because significant effects tend to introduce, for instance, new treatments in clinical applications. However, for a complete view we believe that also reproducibility of tests that did not reveal significant effects is important. Reproducibility is an important aspect of the practical significance of test outcomes and there has been growing interest in the reproducibility probability in recent years. In a discussion of Goodman's study, Senn [76] highlighted the distinctive natures of reproducibility probability and the p-value. Senn [76] agreed with Goodman [48] regarding the importance of reproducibility of test outcomes and the reproducibility probability. However, Senn disagreed with Goodman's argument that p-values exaggerate the evidence against the null-hypothesis, and emphasised differences between the p-value and the reproducibility probability.

Senn [76] also discussed issues with the reproducibility of tests in real-world situations where a repeated test may be under different circumstances and may involve a different team of analysts carrying out the test. Shao and Chow [77] introduced a concept of reproducibility probability (RP) for a given clinical trial. Shao and Chow estimated RP of a statistically significant result for the t-test, and argued that a single clinical trial is sufficient if the result from the first clinical trial is strongly significant. Shao and Chow [77], considered three approaches to study reproducibility probability; a frequentist approach that depends on estimating the power of a future test in relation to available test data, a corresponding approach where RP is associated with a lower confidence bound for the power estimate of the second test, and a Bayesian approach. They stressed the use of RP in situations where evidence in a clinical trial strongly supports a new treatment. De Martini [40] estimated the reproducibility probability of statistically significant results (RP) for one-sided and two-sided alternative hypotheses, and defined statistical tests on the basis of RP estimation. De Martini showed that RP estimation can be used for testing parametric hypotheses. Specially, it is shown that the point estimate of RP is greater than half if and only if a null hypthesis is rejected. De Martini considered two definitions of the RP of statistically significant results, which are the power of the test and the lower confidence bound of the power. De Capitani and De Martini [37] considered different RP estimators for the Wilcoxon rank sum test. In a recent study, De Capitani and De Martini [38] studied RP estimators for some nonparametric tests, e.g. binomial, sign and Wilcoxon signed rank tests.

The formulation of the reproducibility problem is not an estimation or a hypothesis testing problem, which are the main concepts that classical statistics tends to focus on [37, 38, 40]. It is naturally predictive, so NPI fits well with the core problem formulation of a repeat of the test in the future. The natural and straightforward approach of formulating inference on reproducibility probability as a predictive problem also follows the NPI framework presented in Coolen and Himd [25] and developed in this chapter.

Methods for addressing reproducibility, proposed in the literature since then, have mainly shown that the classical frequentist framework of statistics may not be immediately suitable for inference on test reproducibility. Recently, many aspects of reproducibility, including some attention to statistical methods, have been discussed in a volume dedicated to this topic [6]. NPI provides attractive approaches for inference on test reproducibility, as it is a predictive methodology, which is in line with the core problem formulation of a possible repeat of the test.

The reproducibility probability (RP) for a test is the probability for the event that, if the test is repeated based on an experiment performed in the same way as the original experiment, the test outcome, that is either rejection of the null-hypothesis or not, will be the same. In practice, focus may often be on reproducibility of tests in which the null-hypothesis is rejected, for example because significant effects tend to lead to new treatments in medical applications. However, also if the null-hypothesis is not rejected it is important to have a meaningful assessment of the reproducibility of the test. Note that RP is assessed knowing the outcome of the first, actual experiment, which consists of the actual observations, so not only the value of a sufficient test statistic or even just the conclusion on rejection of the null-hypothesis. This is important as the RP will vary with different experiment outcomes, which is logical and will lead to higher RP if the data supported the original test conclusion more strongly. A sufficient test statistic, if of reduced dimension compared to the full data set, does not provide suitable input for the NPI method, hence the use of the full data set is required for the inferences considered in this thesis.

Coolen and Bin Himd [25] introduced NPI for RP, denoted by NPI-RP, by considering some basic nonparametric tests: the sign test, Wilcoxon's signed rank test, and the two sample rank sum test [46]. For these inferences NPI for Bernoulli quantities [20] and for real-valued observations [7] were used. This did not lead to precise valued reproducibility probabilities but to NPI lower and upper reproducibility probabilities, denoted by $\underline{RP}$ and $\overline{RP}$, respectively. For these tests analytic methods were presented to calculate the NPI lower and upper probabilities for test reproducibility. To enable NPI for more complex test scenarios, the NPI-bootstrap method can be used, as introduced and illustrated by Bin Himd [16] for the Kolmogorov-Smirnov test.

This chapter presents NPI-RP for two classical tests which are based on order statistics, namely a one sample quantile test and a two sample precedence test. For these inferences, NPI for future order statistics presented in Chapter 2 is used. We assume that the first, actual experiment led to ordered real-valued observations $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$. As we consider an imaginary repeat of this experiment, we use NPI for $n = m$ future ordered observations, denoted by $X^f_{(1)} < X^f_{(2)} < \ldots < X^f_{(n)}$, with the superscript f used to emphasize that we consider future order statistics.

4.3 Quantile test

The quantile test is a basic nonparametric test for the value of a population quantile [46]. Let κ_p denote the $100 \times p$-th quantile of an unspecified continuous distribution, for $0 \leq p \leq 1$. On the basis of a sample of observations of independent and identically distributed random quantities $X_i, i = 1, \ldots, n$, we consider the one-sided test of null-hypothesis $H_0 : \kappa_p = \kappa_p^0$ versus alternative $H_1 : \kappa_p > \kappa_p^0$, for a specified value κ_p^0. Under H_0, κ_p^0 is the $100 \times p$-th quantile of the distribution function of the X_i, so $P(X_i \leq \kappa_p^0 | H_0) = p$. Define the random variable K as the number of X_i in the sample of size n that are less than or equal to κ_p^0, that is

$$K = \sum_{i=1}^{n} \mathbf{1}\{X_i \leq \kappa_p^0\}$$

A logical test rule is to reject H_0 if $X_{(r)} > \kappa_p^0$, so if $K \leq r - 1$, where $X_{(r)}$ is the r-th ordered observation in the sample (ordered from small to large), for a suitable value of r corresponding to a chosen significance level. For significance level α, r is the largest integer such that

$$P(X_{(r)} > \kappa_p^0 | H_0) = \sum_{i=0}^{r-1} \binom{n}{i} p^i (1-p)^{n-i} \leq \alpha \tag{4.1}$$

For a given data set $x_1, \ldots, x_n$, the test statistic of the one-sided quantile test as defined above is the number of observations less than or equal to κ_p^0, denoted by

$$k = \sum_{i=1}^{n} \mathbf{1}\{x_i \leq \kappa_p^0\}$$

For the value r derived as discussed above, H_0 is rejected if and only if $k \leq r - 1$.

For the two sided alternative $H_1 : \kappa_p \neq \kappa_p^0$, H_0 is rejected if and only if $k \leq r-1$ or $k \geq s$, where r is the largest integer such that $P(X_{(r)} > \kappa_p^0 | H_0) \leq \alpha/2$ and s is the smallest integer such that

$$P(X_{(s)} < \kappa_p^0 | H_0) = \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i} \leq \alpha/2 \tag{4.2}$$

For large sample sizes the Normal distribution approximation to the Binomial distribution can be used in order to determine the appropriate value of r (and s). In such case, the rejection region with $H_1 : \kappa_p > \kappa_p^0$ [46], is

$$k \leq -0.5 + np - z_\alpha \sqrt{np(1-p)}$$

and for the two sided alternative, H_0 is rejected if and only if

$$k \leq np - z_{\alpha/2}\sqrt{np(1-p)} - 0.5 \quad \text{or} \quad k \geq np + z_{\alpha/2}\sqrt{np(1-p)} + 0.5$$

Based on such data and the result of the actual hypothesis test, that is whether the null hypothesis is rejected in favour of the alternative hypothesis or not, NPI can be applied to study the reproducibility of the test. First we consider the case where $k \leq r-1$, so the original test leads to rejection of H_0. Reproducibility of this test result is therefore the event that, if the test were repeated, also with n observations, then that would also lead to rejection of H_0. Using the notation for future observations introduced in Section 4.2, this will occur if $X^f_{(r)} > \kappa^0_p$. The NPI lower and upper reproducibility probabilities for this event, as function of original test result $k \leq r-1$, are

$$\begin{aligned}
\underline{RP}(k) &= \underline{P}(X^f_{(r)} > \kappa^0_p | k) = \sum_{j=1}^{n+1} \mathbf{1}\{x_{j-1} > \kappa^0_p\} P(X^f_{(r)} \in I_j) \\
\overline{RP}(k) &= \overline{P}(X^f_{(r)} > \kappa^0_p | k) = \sum_{j=1}^{n+1} \mathbf{1}\{x_{j} > \kappa^0_p\} P(X^f_{(r)} \in I_j)
\end{aligned}$$

Note that the dependence of these lower and upper probabilities on the value k is not explicit in the notation used for the terms on the right-hand side, but is due to the number of data x_j that exceed κ^0_p. It is easily shown that $\underline{P}(X^f_{(r)} > \kappa^0_p | k) = \overline{P}(X^f_{(r)} > \kappa^0_p | k+1)$, which leads to $\underline{RP}(k) = \overline{RP}(k+1)$ for values of k leading to rejection of H_0.

If the original test does not lead to rejection of H_0, so if $k \geq r$, then reproducibility of the test is the event that the null hypothesis would also not get rejected in the future test. The NPI lower and upper reproducibility probabilities for this event, as functions of $k \geq r$, are

$$\begin{aligned}\underline{RP}(k) &= \underline{P}(X^f_{(r)} \leq \kappa^0_p | k) = \sum_{j=1}^{n+1} \mathbf{1}\{x_j \leq \kappa^0_p\} P(X^f_{(r)} \in I_j) \\ \overline{RP}(k) &= \overline{P}(X^f_{(r)} \leq \kappa^0_p | k) = \sum_{j=1}^{n+1} \mathbf{1}\{x_{j-1} \leq \kappa^0_p\} P(X^f_{(r)} \in I_j)\end{aligned}$$

It is easily seen that $\underline{RP}(k) = \overline{RP}(k-1)$ for values of k such that $k-1$ leads to H_0 not being rejected. If an actual observation in the original test is exactly equal to the specified value κ^0_p, then the NPI method provides a precise reproducibility probability. We do not consider this further here as the test hypotheses must always be specified without consideration of the actual test data. Hence this case is extremely unlikely to occur, but we do briefly illustrate this case in Example 4.1.

The minimum value that can occur for the NPI lower reproducibility probabilities for this one-sided quantile test, following either rejection or non-rejection of the null hypothesis in the original test, is equal to 0.5. This follows directly from the formulae for the NPI lower reproducibility probabilities given above, together with the probability $P(X^f_{(r)} < x_r) = P(X^f_{(r)} > x_r) = 0.5$ as explained in Section 2.2. The NPI upper reproducibility probabilities can be equal to one. This occurs when all observations in the original test are less than κ^0_p, so $k = n$, in which case the original test let to H_0 not being rejected for all values of r (so for all order statistics considered), at any level of significance; this reflects that, with no evidence in the original data in favour of the possibility that the data values can actually exceed κ^0_p, one cannot exclude the possibility that no future observations could exceed this value. Note that the corresponding NPI lower reproducibility probability will be less than one, reflecting that the original data set only provides limited information, this lower probability will increase towards one as function of n. The upper reproducibility probability is also equal to one if all observations in the original test are greater than κ^0_p, so $k = 0$, for which case the reasoning is similar to that above but of course now with H_0 being rejected.

For the two sided test with $H_1 : \kappa_p \neq \kappa^0_p$, the original test led to rejection of H_0

if and only if $k \leq r-1$ or $k \geq s$. Then the NPI lower and upper reproducibility probabilities, given k, are

$$\begin{aligned}
\underline{RP}(k) &= \underline{P}(X^f_{(r)} > \kappa^0_p \cup X^f_{(s)} < \kappa^0_p|k) = \underline{P}(X^f_{(r)} > \kappa^0_p|k) + \underline{P}(X^f_{(s)} < \kappa^0_p|k) \\
\overline{RP}(k) &= \overline{P}(X^f_{(r)} > \kappa^0_p \cup X^f_{(s)} < \kappa^0_p|k) \\
&= \overline{P}(X^f_{(r)} > \kappa^0_p|k) + \overline{P}(X^f_{(s)} < \kappa^0_p|k) - \overline{P}(X^f_{(r)} > \kappa^0_p, X^f_{(s)} < \kappa^0_p|k)
\end{aligned}$$

Suppose $\kappa^0_p \in (x_{i-1}, x_i)$, then we can write the NPI lower and upper reproducibility probabilities, given either $k \leq r-1$ or $k \geq s$ as

$$\begin{aligned}
\underline{RP}(k) &= P(X^f_{(r)} > x_i \cup X^f_{(s)} < x_{i-1}|k) = P(X^f_{(r)} > x_i|k) + P(X^f_{(s)} < x_{i-1}|k) \\
\overline{RP}(k) &= P(X^f_{(r)} > x_{i-1} \cup X^f_{(s)} < x_i|k) \\
&= P(X^f_{(r)} > x_{i-1}|k) + P(X^f_{(s)} < x_i|k) - P(X^f_{(r)} \in I_i, X^f_{(s)} \in I_i|k)
\end{aligned}$$

where the value k affects the probabilities through the value of i. Note that the events $X^f_{(r)} > x_i$ and $X^f_{(s)} < x_{i-1}$ are mutually exclusive events, but the events $X^f_{(r)} > x_{i-1}$ and $X^f_{(s)} < x_i$ are not mutually exclusive, as $X^f_{(r)}$ and $X^f_{(s)}$ can both be in same interval $I_i = (x_{i-1}, x_i)$. If the original test does not lead to rejection of H_0, so if $r \leq k \leq s-1$, then reproducibility of this test is the event that H_0 would also not get rejected in the future test, which occurs if $X^f_{(r)} < \kappa^0_p$ and $X^f_{(s)} > \kappa^0_p$. The NPI lower and upper reproducibility probabilities for this event, as a function of k, are

$$\begin{aligned}
\underline{RP}(k) &= \underline{P}(X^f_{(r)} < \kappa^0_p \cap X^f_{(s)} > \kappa^0_p|k) & (4.3) \\
&= \sum_{j=1}^{n+1}\sum_{l=j}^{n+1} 1\{x_j < \kappa^0_p, x_{l-1} > \kappa^0_p\} P(X^f_{(r)} \in I_j, X^f_{(s)} \in I_l) \\
\overline{RP}(k) &= \overline{P}(X^f_{(r)} < \kappa^0_p \cap X^f_{(s)} > \kappa^0_p|k) & (4.4) \\
&= \sum_{j=1}^{n+1}\sum_{l=j}^{n+1} 1\{x_{j-1} < \kappa^0_p, x_l > \kappa^0_p\} P(X^f_{(r)} \in I_j, X^f_{(s)} \in I_l)
\end{aligned}$$

which is easy to compute using the joint probability distribution (2.6) derived in Section 2.3.

The minimum value that can occur for the NPI lower reproducibility probabilities for this two-sided quantile test, for the case of rejection of the null hypothesis in the original test given either $k = r - 1$ or $k = s$, is greater than 0.5. This follows directly from the formulae for the NPI lower reproducibility probabilities given either $k = r - 1$ or $k = s$, with $P(X^f_{(r)} > x_r \cup X^f_{(s)} < x_{r-1} | k = r - 1) > 0.5$ and $P(X^f_{(r)} > x_{s+1} \cup X^f_{(s)} < x_s | k = s) > 0.5$. This is because of the two rejection regions for H_0, therefore the event here is different than the one for the one-sided test, as we now sum up the probability masses for the two events which are $X^f_{(r)} > x_r$ and $X^f_{(s)} < x_{r-1}$ both given $k = r - 1$, where the probability for the event $X^f_{(r)} > x_r$ is equal to 0.5, as explained in Section 2.2, and $P(X^f_{(s)} < x_{r-1} | k = r - 1) > 0$. The maximum value that can occur for the NPI upper reproducibility probabilities for the case of rejecting H_0 in the initial test given either $k = 0$ or $k = n$, is equal to 1. This occurs if $k = 0$, so all observations in the original test are greater than κ^0_p, in which situation the original test led to H_0 being rejected for all values of r, and for $k = n$, so all observations in the original test are less than κ^0_p, in which case the original test led to H_0 being rejected for all values of s.

If H_0 is not rejected in the original test, so if $s - 1 \leq k \leq r$, then the minimum value that can occur for the NPI lower reproducibility probability for the event $X^f_{(r)} < \kappa^0_p \cap X^f_{(s)} > \kappa^0_p$ is less than 0.5. For $k = r$, so $\kappa^0_p \in (x_r, x_{r+1})$, we have $\underline{RP}(r) = \underline{P}(X^f_{(r)} < x_r \cap X^f_{(s)} > x_{r+1}) < \underline{P}(X^f_{(r)} < x_r) = 0.5$, and for $k = s - 1$, so $\kappa^0_p \in (x_{s-1}, x_s)$, we have $\underline{RP}(s-1) = \underline{P}(X^f_{(r)} < x_{s-1} \cap X^f_{(s)} > x_s) < 0.5$. The maximum value that can occur for the NPI reproducibility upper probabilities for the case where the original test led to not reject the null hypothesis, is less than 1. This occurs when κ^0_p is neither close to x_r nor to x_s.

Actually, there is an interesting issue about two-sided tests in such scenarios, that requires some further thought. One could argue that one should only consider reproducibility for one-sided tests. This is because, in two-sided tests if the original test leads to rejection of the null hypothesis due to a relatively large value of the test statistic, one may not consider the test result to be reproduced if a future test leads to rejection due to a relatively small value of the test statistic, so in the other tail of the statistic's distribution under H_0! On the basis of the combined evidence of the two tests in such a case, one would probably want to investigate the whole setting further and not regard the second test as confirming the results of the first test.

NPI-RP for the one-sided and two-sided quantile tests is illustrated in the following example.

Example 4.1. Suppose that an original quantile test has sample size $n = 15$ and we are interested in testing the null hypothesis that the third quartile, so the 75% quantile, of the underlying distribution is equal to a specified value $\kappa^0_{0.75}$, against the alternative hypothesis that this third quartile is greater than $\kappa^0_{0.75}$, tested at significance level $\alpha = 0.05$. Using the Binomial distribution for the classical quantile test, this leads to the rule that H_0 is rejected if $x_{(8)} > \kappa^0_{0.75}$ and H_0 is not rejected if $x_{(8)} < \kappa^0_{0.75}$.

Table 4.1 presents the NPI lower and upper reproducibility probabilities for all values of k, which is the number of observations in the original test which are less than $\kappa^0_{0.75}$. If $k \leq 7$ then the original test leads to H_0 being rejected while it is not rejected for $k \geq 8$. Hence, the NPI lower and upper reproducibility probabilities are for the events $X^f_{(8)} > \kappa^0_{0.75}$ in case of rejecting H_0 in the original test and $X^f_{(8)} < \kappa^0_{0.75}$ in case of not rejecting H_0 in the original test. This table illustrates the logical fact that the worst reproducibility is achieved for k at the threshold values 7 and 8, with increasing RP values when moving away from these values, leading to maximum NPI-RP values for $k = 0$ and $k = 15$. Because for this test the threshold between rejecting and not rejecting H_0 is between $k = 7$ and $k = 8$ out of $n = 15$ observations, the NPI-RP values are symmetric, that is the same for $k = j$ and $k = 15 - j$, for $j = 0, 1, \ldots, 7$, in Table 4.1.

Table 4.2 illustrates the case if the third quartile by coincidence is equal to one of the data observations i.e. $x_{(j)} = \kappa^0_{0.75}$, which is slightly different as now the NPI-RP approach leads to precise probabilities instead of lower and upper probabilities.

The results in Table 4.2 show that the worst NPI-RP for the case of rejecting H_0 in the original test is equal to 0.5. This occurs for the NPI reproducibility probability at the threshold value $k = 7$, so $\kappa^0_{0.75} = x_{(8)}$, i.e. $RP(7) = P(X_{(8)} > x_8) = 0.5$. Whereas for the case of non-rejection of H_0 in the original test, the minimum NPI-RP value is greater than a half, i.e. $RP(8) = P(X_{(8)} < x_9) > 0.5$. The maximum value that can occur for NPI-RP is not equal to one, in both cases $RP(0) = P(X_{(8)} > x_1) < 1$ and $RP(14) = P(X_{(8)} < x_{15}) < 1$. This is because the value of $\kappa^0_{0.75}$ is equal to either the minimum or maximum observation, so there is one interval left beyond that observation. Note that here k can only take values from 0 to 14 as again the value of $\kappa^0_{0.75}$ is equal to one of the data values x_j, for $j = 1, \ldots, 15$, so if $\kappa^0_{0.75} = x_{15}$ then $k = 14$; there are 14 observations less than $\kappa^0_{0.75} = x_{15}$.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	0.9989	1	6	0.6424	0.7689	12	0.9359	0.9749
1	0.9929	0.9989	7	0.5	0.6424	13	0.9749	0.9929
2	0.9749	0.9929	8	0.5	0.6424	14	0.9929	0.9989
3	0.9359	0.9749	9	0.6424	0.7689	15	0.9989	1
4	0.8682	0.9359	10	0.7689	0.8682			
5	0.7689	0.8682	11	0.8682	0.9359			

Table 4.1: NPI-RP for third quartile, $n = 15$ and $\alpha = 0.05$.

k	RP	k	RP	k	RP
0	0.9989	6	0.6424	12	0.9749
1	0.9929	7	0.5	13	0.9929
2	0.9749	8	0.6424	14	0.9989
3	0.9359	9	0.7689		
4	0.8682	10	0.8682		
5	0.7689	11	0.9359		

Table 4.2: NPI-RP for third quartile, $n = 15$ and $\alpha = 0.05$.

Table 4.3 presents NPI-RP values for the quantile test considering the median, so the 50% quantile, again with sample size $n = 15$ and testing the null hypothesis that the median is equal to a specified value $\kappa^0_{0.5}$ against the one-sided hypothesis that it is greater than $\kappa^0_{0.5}$, at level of significance $\alpha = 0.05$. This leads to the test rule that H_0 is rejected if the number k of observations that are smaller than $\kappa^0_{0.5}$ is less than or equal to 3, and H_0 is not rejected if $k \geq 4$. Note that throughout this chapter, precise values 0.5 and 1 are presented without additional decimals, so the values 1.0000 are actually less than 1 but rounded upwards. Of course, these NPI-RP values are not symmetric, and reproducibility becomes very likely for initial test results with a substantial number of observations less than $\kappa^0_{0.5}$. But rejection of H_0, which occurs for $k \leq 3$ and is often of main practical relevance, has relatively low NPI-RP values.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	0.9502	1	6	0.7865	0.8775	12	0.9986	0.9997
1	0.8352	0.9502	7	0.8775	0.9359	13	0.9997	0.9999
2	0.6743	0.8352	8	0.9359	0.9698	14	0.9999	1.0000
3	0.5	0.6743	9	0.9698	0.9873	15	1.0000	1
4	0.5	0.6592	10	0.9873	0.9954			
5	0.6592	0.7865	11	0.9954	0.9986			

Table 4.3: NPI-RP for median, $n = 15$ and $\alpha = 0.05$.

Tables 4.4 and 4.5 present the NPI-RP results for the same one-sided quantile test on the third quartile for $n = 30$, at significance levels $\alpha = 0.05$ and $\alpha = 0.01$, respectively. Using the Normal distribution approximation, the test rule for $\alpha = 0.05$ is to reject H_0 that this third quartile is equal to $\kappa^0_{0.75}$ in favour of the alternative hypothesis that it is greater than $\kappa^0_{0.75}$ if $k \leq 18$ and not to reject it if $k \geq 19$, where k is again the number of observations less than $\kappa^0_{0.75}$. For $\alpha = 0.01$, H_0 is rejected if $k \leq 16$ and not rejected if $k \geq 17$. The change in level of significance α leads obviously to change of the rejection threshold, with H_0 being rejected for a smaller range of values k in case of smaller value of α. Comparison of these tables with Table 4.1 shows that the larger sample size tends to lead to slightly less imprecision, that is the difference between corresponding upper and lower probabilities, this is e.g. shown by considering the upper probabilities $\overline{RP}(k)$ for the values of k next to the rejection thresholds, so corresponding to $\underline{RP}(k) = 0.5$.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	1.0000	1	11	0.9651	0.9811	22	0.7941	0.8666
1	1.0000	1.0000	12	0.9398	0.9651	23	0.8666	0.9210
2	1.0000	1.0000	13	0.9023	0.9398	24	0.9210	0.9580
3	1.0000	1.0000	14	0.8503	0.9023	25	0.9580	0.9805
4	0.9999	1.0000	15	0.7826	0.8503	26	0.9805	0.9923
5	0.9998	0.9999	16	0.6995	0.7826	27	0.9923	0.9976
6	0.9993	0.9998	17	0.6038	0.6995	28	0.9976	0.9995
7	0.9981	0.9993	18	0.5	0.6038	29	0.9995	0.9999
8	0.9956	0.9981	19	0.5	0.6054	30	0.9999	1
9	0.9905	0.9956	20	0.6054	0.7056			
10	0.9811	0.9905	21	0.7056	0.7941			

Table 4.4: NPI-RP for third quartile, $n = 15$ and $\alpha = 0.05$.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	1.0000	1	11	0.9023	0.9406	22	0.9101	0.9483
1	1.0000	1.0000	12	0.8493	0.9023	23	0.9483	0.9731
2	1.0000	1.0000	13	0.7805	0.8493	24	0.9731	0.9875
3	0.9999	1.0000	14	0.6971	0.7805	25	0.9875	0.9949
4	0.9995	0.9999	15	0.6019	0.6971	26	0.9949	0.9983
5	0.9986	0.9995	16	0.5	0.6019	27	0.9983	0.9995
6	0.9964	0.9986	17	0.5	0.6026	28	0.9995	0.9999
7	0.9916	0.9964	18	0.6026	0.6995	29	0.9999	1.0000
8	0.9824	0.9916	19	0.6995	0.7852	30	1.0000	1
9	0.9664	0.9824	20	0.7852	0.8559			
10	0.9406	0.9664	21	0.8559	0.9101			

Table 4.5: NPI-RP for third quartile, $n = 30$ and $\alpha = 0.01$.

Tables 4.6 and 4.7 illustrate the NPI lower and upper reproducibility probabilities for the one sample quantile test considering the median, with two-sided alternative hypothesis $H_1 : \kappa_{0.5} \neq \kappa^0_{0.5}$, for $n = 20$ at significance levels $\alpha = 0.05$ and $\alpha = 0.01$, respectively. For Table 4.6, with $\alpha = 0.05$, the null hypothesis is rejected in the original test if $k \leq 5$ or $k \geq 15$ while it is not rejected if $6 \leq k \leq 14$. Hence, the NPI lower and upper reproducibility probabilities are for the events $X^f_{(6)} > \kappa^0_{0.5} \cup X^f_{(15)} < \kappa^0_{0.5}$, given the value of k with either $k \leq 5$ or $k \geq 15$ in case of rejection of H_0 in the original test, and $X^f_{(6)} < \kappa^0_{0.5}, X^f_{(15)} > \kappa^0_{0.5}$, given the value of k with $6 \leq k \leq 14$ in case of non-rejection of H_0 in the original test. The results in Table 4.6 show that the minimum value for $\underline{RP}(k)$, for k leading to rejection of H_0, namely 0.5006 at both $k = 5$ and $k = 15$, is now indeed greater than 0.5, due to the two rejection areas for H_0, as discussed before. Moving away from these threshold values between rejecting and not rejecting H_0 leads to increasing values of NPI-RP, with the maximum value of the upper reproducibility probabilities being equal to one, which is achieved for $k = 0$ and $k = 20$. The minimum value for $\underline{RP}(k)$, in case H_0 was not rejected in the original test, which is 0.4948 for both $k = 6$ and $k = 14$, is less than 0.5, as discussed before. Tables 4.6 and 4.7 illustrate that, when we consider the two-sided test for the median, the NPI lower and upper reproducibility probabilities are symmetric about $k = n/2 = 10$. The maximum NPI lower and upper RP values, for k which leads to non-rejection of H_0, are 0.8093 and 0.8947 in Table 4.6, and 0.9593 and 0.9813 in Table 4.7, respectively, which are both achieved at $k = 10$.

For the situation in Table 4.7, with $\alpha = 0.01$, the null hypothesis is rejected in the original test if $k \leq 3$ or $k \geq 17$ while it is not rejected for $4 \leq k \leq 16$. So with small α the null hypothesis not rejected for a large range of values k, which is in line with the same feature as discussed for one-sided tests earlier. Comparing this table with Table 4.6, the NPI lower and upper reproducibility probabilities are smaller for values of k for which H_0 is rejected and larger for values of k for which H_0 is not rejected. This is logical as the change in the level of significance obviously leads to change of the rejection threshold.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	0.9899	1	7	0.6221	0.7445	14	0.4948	0.6325
1	0.9543	0.9899	8	0.7230	0.8274	15	0.5006	0.6437
2	0.8824	0.9543	9	0.7873	0.8778	16	0.6420	0.7753
3	0.7747	0.8826	10	0.8093	0.8947	17	0.7747	0.8826
4	0.6420	0.7753	11	0.7873	0.8778	18	0.8824	0.9543
5	0.5006	0.6437	12	0.7230	0.8274	19	0.9543	0.9899
6	0.4948	0.6325	13	0.6221	0.7445	20	0.9899	1

Table 4.6: NPI-RP for median, $n = 20$ and $\alpha = 0.05$.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	0.9470	1	7	0.8618	0.9219	14	0.7742	0.8632
1	0.8292	0.9470	8	0.9185	0.9573	15	0.6525	0.7747
2	0.6693	0.8292	9	0.9495	0.9757	16	0.5000	0.6526
3	0.5000	0.6693	10	0.9593	0.9813	17	0.5000	0.6693
4	0.5000	0.6526	11	0.9495	0.9757	18	0.6693	0.8292
5	0.6525	0.7747	12	0.9185	0.9573	19	0.8292	0.9470
6	0.7742	0.8632	13	0.8618	0.9219	20	0.9470	1

Table 4.7: NPI-RP for median, $n = 20$ and $\alpha = 0.01$.

Table 4.8 presents NPI-values for the quantile test considering the third quantile, with sample size of 20 and testing the null hypothesis $\kappa_{0.75} = \kappa^0_{0.75}$ against the two-sided hypothesis that $\kappa_{0.75} \neq \kappa^0_{0.75}$ at level of significance $\alpha = 0.05$. The original test led to rejection of H_0 if $k \leq 10$ or $k \geq 19$ and non-rejection if $11 \leq k \leq 18$. Thus the NPI lower and upper reproducibility probabilities are for the events $X^f_{(11)} > \kappa^0_{0.75} \cup X^f_{(19)} < \kappa^0_{0.75}$ in case of rejection H_0 in the original test and $X^f_{(11)} < \kappa^0_{0.75}, X^f_{(19)} > \kappa^0_{0.75}$ in case of not rejecting H_0 in the original test. The minimum value for $\underline{RP}(k)$, for k leading to rejection of H_0, is equal to 0.5006 at $k = 10$, and 0.5002 at $k = 19$. The maximum $\overline{RP}(k)$ is equal to one, achieved for both $k = 0$ and $k = 20$. The smallest values for $\underline{RP}(k)$ in case H_0 was not rejected in the original test, are 0.4958 at $k = 11$ and 0.4934 at $k = 18$, while the maximum $\overline{RP}(k)$ for this case is equal to 0.9063, which occurs at $k = 15$.

k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$	k	$\underline{RP}(k)$	$\overline{RP}(k)$
0	0.9999	1	7	0.8297	0.8999	14	0.7878	0.8828
1	0.9994	0.9999	8	0.7364	0.8299	15	0.8139	0.9063
2	0.9971	0.9994	9	0.6240	0.7370	16	0.7814	0.8887
3	0.9906	0.9971	10	0.5006	0.6255	17	0.6774	0.8223
4	0.9758	0.9907	11	0.4958	0.6227	18	0.4934	0.6958
5	0.9473	0.9758	12	0.6146	0.7343	19	0.5002	0.7578
6	0.8999	0.9473	13	0.7167	0.8236	20	0.7564	1

Table 4.8: NPI-RP for third quartile, $n = 20$ and $\alpha = 0.05$.

4.4 Precedence test

As a second example of NPI for reproducibility of a statistical test based on order statistics, we consider a basic nonparametric precedence test. Such a test, first proposed by Nelson [73], is typically used for comparison of two groups of lifetime data, where one wishes to reach a conclusion before all units on test have failed. The test is based on the order of the observed failure times for the two groups, and typically leads to, possibly many, right-censored observations at the time when the test is ended. Balakrishnan and Ng [10] present a detailed introduction and overview of precedence testing, including more sophisticated tests than the basic one considered in this chapter. NPI for precedence testing was presented by Coolen-Schrijner, et al. [35], without consideration of reproducibility. It should be emphasized that we consider here the NPI approach for reproducibility of a classical precedence test, so not of the NPI approach to precedence testing [35].

We consider the classical scenario with two independent samples. Let $X_{(1)} < X_{(2)} < \ldots < X_{(n_x)}$ be random quantities representing the ordered real-valued observations in a sample of size n_x, drawn randomly from a continuously distributed population, which we refer to as the X population, with a probability distribution depending on location parameter λ_x. Similarly, let $Y_{(1)} < Y_{(2)} < \ldots < Y_{(n_y)}$ be random quantities representing the ordered real-valued observations in a sample of size n_y, drawn randomly from another continuously distributed population, the Y population, with a probability distribution which is identical to that of the X population except for its location parameter λ_y. We consider the hypothesis test for the locations of these two populations is $H_0 : \lambda_x = \lambda_y$ versus $H_1 : \lambda_x < \lambda_y$, which is to be interpreted such that, under H_1, observations from the Y population tend to be larger than observations from the X population.

The precedence test considered in this section, for this specific hypothesis test scenario, is as follows. Given n_x and n_y, one specifies the value of r, such that the test is ended at, or before, the r-th observation of the Y population. For specific level of signifance α, one determines the value k (which therefore is a function of α and of r) such that H_0 is rejected if and only if $X_{(k)} < Y_{(r)}$. The critical value for k is the smallest integer which satisfies

$$P(X_{(k)} < Y_{(r)}|H_0) = \binom{n_x + n_y}{n_x}^{-1} \sum_{j=0}^{r-1} \binom{j+k-1}{j} \binom{n_y - j + n_x - k}{n_y - j} \leq \alpha$$

Note that the test is typically ended at the time $T = \min(X_{(k)}, Y_{(r)})$, with the conclusion that H_0 is rejected in favour of the one-sided alternative hypothesis H_1, specified above, if $T = X_{(k)}$, and H_0 is not rejected if $T = Y_{(r)}$. It is of interest to emphasize this censoring; continuing with the original test would make no difference at all to the test conclusion, but further observations would make a difference for the NPI reproducibility results, as will be discussed later.

The NPI approach for reproducibility of this two-sample precedence test considers again the same test scenario applied to future order statistics, and derives the NPI lower and upper probabilities for the event that the same overall test conclusion will be derived, given the data from the original test. This involves the NPI approach for inference on the r-th future order statistic $Y^f_{(r)}$ out of n_y future observations based on the data from the Y population, and similarly for the k-th future order statistic $X^f_{(k)}$ out of the n_x future observations based on the data from the X population, where the values of r and k are the same as used for the original test (as we assume also the same significance level for the future test). Note, however, that there is a complication: for full specification of the NPI probabilities for these future order statistics, we require the full data from the original test to be available. But, as mentioned, the data resulting from the original precedence test typically have right-censored observations for at least one, but most likely both populations, and these are all just known to exceed the time T at which the original test had ended.

Before we proceed, we discuss this situation in more detail as it is important for the general idea of studying reproducibility of tests. We should emphasize that we have not come across this important issue before in the literature. There are two perspectives on the study of reproducibility of such precedence tests. First, one can study the test outcome assuming that, actually, complete data were available, so all n_x and n_y observations of the X and Y populations, respectively, in the original test are assumed to be available. Secondly, one can consider inference for the realistic scenario with the actual data from the original test, so including right-censored observations at time T. The first scenario is the most straightforward for the development of NPI-RP, and we start with this scenario. Then we explain how this first scenario, without additional assumptions, leads to NPI-RP for the second scenario.

The starting point for NPI-RP for the precedence test is to apply NPI for n_x future observations, based on the n_x original test observations from the X population, which are assumed to be fully available, and similarly for n_y future observations based on the n_y observations from the Y population. Using the results presented in Chapter 2, with notation adapted to indicate the specific populations, the following NPI lower and upper reproducibility probabilities are derived. First, if H_0 is rejected in the original test, so $x_{(k)} < y_{(r)}$, then

$$\underline{RP} = \underline{P}(X^f_{(k)} < Y^f_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x)} < y_{(j_y-1)}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \tag{4.5}$$

$$\overline{RP} = \overline{P}(X^f_{(k)} < Y^f_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x-1)} < y_{(j_y)}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \tag{4.6}$$

Note that these $\underline{RP}$ and $\overline{RP}$ are conditional on the orderings of the full data from the original test, which for simplicity we do not include in the notation. If H_0 is not rejected in the original test, so $x_{(k)} > y_{(r)}$, then

$$\underline{RP} = \underline{P}(X^f_{(k)} > Y^f_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x-1)} > y_{(j_y)}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \tag{4.7}$$

$$\overline{RP} = \overline{P}(X^f_{(k)} > Y^f_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x)} > y_{(j_y-1)}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \tag{4.8}$$

Both in case of rejecting and not rejecting H_0, the maximum possible value of the NPI upper reproducibility probability is 1. If H_0 was rejected this occurs if $x_{(n_x)} < y_{(1)}$, while if H_0 was not rejected this occurs if $x_{(1)} > y_{(n_y)}$, so both cases lead to maximum reproducibility if the original test data were entirely separated in the sense that either all observations from the X population occurred before all observations from the Y population, or the other way around.

In both cases of rejecting or not rejecting H_0 in the original test, the minimum value of the NPI lower reproducibility probability is 0.25. If H_0 was rejected, this occurs if $y_{(r-1)} < x_{(1)}$ and $x_{(k)} < y_{(r)}$ and $y_{(n_y)} < x_{(k+1)}$. If H_0 was not rejected, this occurs if $x_{(k-1)} < y_{(1)}$ and $y_{(r)} < x_{(k)}$ and $x_{(n_x)} < y_{(r+1)}$. Both these smallest possible values for $\underline{RP}$ result from data orderings that, whilst leading to a test conclusion, are least supportive for it, together with the fact that $P(X^f_{(k)} < x_{(k)}) = P(X^f_{(k)} > x_{(k)}) = 0.5$, and similar for $Y^f_{(r)}$, as discussed in Section 2.2.

The effect of local changes to the combined ordering of the data of the two populations in the original test is important. Suppose that, for given data for the X and Y populations for the original test, observations $y_{(u)}$ and $x_{(v)}$ are such that $y_{(u)} < x_{(v)}$ and in the combined ordering of all $n_x + n_y$ data they are consecutive. Now suppose that we change these observations, and denote them by $\tilde{y}_{(u)}$ and $\tilde{x}_{(v)}$, respectively, such that they keep their order in the data from their own population but between them change their order, so $\tilde{x}_{(v)} < \tilde{y}_{(u)}$. From Equations (4.5) and (4.6), the difference between the NPI lower and upper probabilities for the event $X^f_{(k)} < Y^f_{(r)}$ given $y_{(u)} < x_{(v)}$ and the NPI lower and upper probabilities for the same event given $\tilde{x}_{(v)} < \tilde{y}_{(u)}$, are

$$
\begin{aligned}
&\underline{P}(X^f_{(k)} < Y^f_{(r)} | y_{(u)} < x_{(v)}) - \underline{P}(X^f_{(k)} < Y^f_{(r)} | \tilde{x}_{(v)} < \tilde{y}_{(u)}) \\
&= \sum_{j_x \neq v}^{n_x+1} \sum_{j_y \neq u+1}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \\
&\quad - \Big[\sum_{j_x \neq v}^{n_x+1} \sum_{j_y \neq u+1}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \\
&\quad + P(X^f_{(k)} \in I^x_{v_x}) P(Y^f_{(r)} \in I^y_{u_y+1}) \Big] = -P(X^f_{(k)} \in I^x_{v_x}) P(Y^f_{(r)} \in I^y_{u_y+1})
\end{aligned}
$$

$$
\begin{aligned}
&\overline{P}(X^f_{(k)} < Y^f_{(r)} | y_{(u)} < x_{(v)}) - \overline{P}(X^f_{(k)} < Y^f_{(r)} | \tilde{x}_{(v)} < \tilde{y}_{(u)}) \\
&\quad = \sum_{j_x \neq v+1}^{n_x+1} \sum_{j_y \neq u}^{n_y+1} 1\{x_{j_x-1} < y_{j_y}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \\
&\qquad - \Big[\sum_{j_x \neq v}^{n_x+1} \sum_{j_y \neq u+1}^{n_y+1} 1\{x_{j_x} < y_{j_y-1}\} P(X^f_{(k)} \in I^x_{j_x}) P(Y^f_{(r)} \in I^y_{j_y}) \\
&\qquad + P(X^f_{(k)} \in I^x_{v_x+1}) P(Y^f_{(r)} \in I^y_{u_y}) \Big] = -P(X^f_{(k)} \in I^x_{v_x+1}) P(Y^f_{(r)} \in I^y_{u_y})
\end{aligned}
$$

Then this local change to the combined ordering of the data leads to increase of both the NPI lower and upper probabilities for the event $X^f_{(k)} < Y^f_{(r)}$, that is

$$
\underline{P}(X^f_{(k)} < Y^f_{(r)} | y_{(u)} < x_{(v)}) \;<\; \underline{P}(X^f_{(k)} < Y^f_{(r)} | \tilde{x}_{(v)} < \tilde{y}_{(u)}) \tag{4.9}
$$

$$
\overline{P}(X^f_{(k)} < Y^f_{(r)} | y_{(u)} < x_{(v)}) \;<\; \overline{P}(X^f_{(k)} < Y^f_{(r)} | \tilde{x}_{(v)} < \tilde{y}_{(u)}) \tag{4.10}
$$

This implies that the NPI-RP inferences for the precedence test depend monotonically on the combined ordering of the original test data, which is an important property to derive such inference for actual tests including right-censored observations, as discussed after Example 4.2.

Example 4.2. Nelson [74] presents data consisting of six groups of times (in minutes) to breakdown of an insulating fluid subjected to different levels of voltage. To illustrate NPI-RP for the basic precedence test as discussed above, we assume that sample 3 provides data from the X population and sample 6 from the Y population. These times are presented in Table 4.9. Both samples are of size 10, and we assume that the precedence testing scenario discussed in this section is followed, so we assume that the population distributions may only differ in location parameters, with $H_0 : \lambda_x = \lambda_y$ tested versus $H_1 : \lambda_x < \lambda_y$. We assume that $r = 6$, so the test is set up to end at the observation of the sixth failure time for the Y population. We discuss both significance levels $\alpha = 0.05$ and $\alpha = 0.1$. The missing values in Table 4.9 are only known to exceed 3.83.

X sample	0.94	0.64	0.82	0.93	1.08	1.99	2.06	2.15	2.57	*
Y sample	1.34	1.49	1.56	2.10	2.12	3.83	*	*	*	*

Table 4.9: Times to insulating fluid breakdown.

For significance level $\alpha = 0.05$, the critical value is $k = 10$, while for $\alpha = 0.1$ this is $k = 9$. Therefore, the provided data will lead, in this precedence test, to rejection of H_0 at 10% level of significance but not to rejection of H_0 at 5% level of significance. For both scenarios, the NPI lower and upper reproducibility probabilities are presented in Table 4.10, for all of the possible orderings of the right-censored observations. Note that in total 15 observations are available, with 1 value of the X sample and 4 values of the Y sample only known to exceed 3.83. In Table 4.10, we give the rank, from the combined ordering of all 20 observations, of the right-censored observation $x_{(10)}$, for example when this is 17 it implies that $y_{(7)} < x_{(10)} < y_{(8)}$. Table 4.10 presents both the results for $\alpha = 0.05$, in which case H_0 was not rejected in the original test, hence reproducibility is achieved if H_0 is also not rejected in the future test, and the results for $\alpha = 0.01$, in which case H_0 was rejected so reproducibility also implies rejection of H_0 in the future test. Note that for $\alpha = 0.1$ we still assume that $y_{(6)} = 3.83$ was actually observed, even though the test could have been concluded at time $x_{(9)} = 2.57$ because $x_{(9)} < y_{(6)}$ was conclusive for the test in this case. Table 4.10 shows that the NPI-RP values are increasing in the combined rank of $x_{(10)}$ for $\alpha = 0.05$ and decreasing for $\alpha = 0.1$, which illustrates the monotonicity of these inferences with regard to changes in ranks of the data as discussed above, as increasing combined rank of $x_{(10)}$ provides more evidence in support of H_0, hence in favour of reproducing the original test result for $\alpha = 0.05$ but against doing so for $\alpha = 0.1$. We notice that the actual rank that $x_{(10)}$ would have among the 20 combined observations has substantial influence on the NPI-RP values.

$\Diamond$

	$\alpha = 0.05$		$\alpha = 0.1$	
rank of $x_{(10)}$	$\underline{RP}$	$\overline{RP}$	$\underline{RP}$	$\overline{RP}$
16	0.3871	0.7814	0.3885	0.7079
17	0.4746	0.8209	0.3490	0.6665
18	0.5496	0.8484	0.3215	0.6309
19	0.6019	0.8627	0.3072	0.6062
20	0.6290	0.8669	0.3029	0.5934

Table 4.10: NPI-RP for precedence test on insulating fluid breakdown data.

Thus far, we have studied reproducibility of the basic precedence test from the perspective of having the complete data available. In Example 4.2 this was illustrated by considering all possible orderings for the right-censored data in the two samples. However, a more realistic perspective is to only use the actual test outcome, without any assumptions on the ordering of the right-censored observations. Using lower and upper probabilities, this can be easily achieved by defining $\underline{RP}$ as the minimum of all NPI lower probabilities for reproducibility over all possible orderings for the right-censored observations, and similarly by defining $\overline{RP}$ as the maximum of all NPI upper probabilities for reproducibility over all possible orderings for the right-censored observations. Hence, in Example 4.2, this leads to $\underline{RP} = 0.3871$ and $\overline{RP} = 0.8669$ for $\alpha = 0.05$, and $\underline{RP} = 0.3029$ and $\overline{RP} = 0.7079$ for $\alpha = 0.1$. Of course, this leads to increased imprecision compared to every possible specific ordering of the right-censored observations, but it is convenient as no further assumptions about those right-censored observations are required. Furthermore, to derive the NPI-RP values for this perspective one does not need to calculate the corresponding values for each possible combined ordering of right-censored observations, due to the above discussed monotonicity of these inferences and given in Equations (4.9) and (4.10). Hence, we always know for which specific ordering of right-censored observations these NPI-RP values are obtained, that is either with all right-censored observations from the

X sample occurring before all right-censored observations from the Y sample, or the other way around, depending on the actual outcome of the original test. This perspective is illustrated in Example 4.3.

Example 4.3. We consider again NPI-RP for the precedence test as presented in this section, so with one-sided alternative hypothesis $H_1 : \lambda_x < \lambda_y$. Suppose that $n_x = 10$ units of the X population and $n_y = 8$ units of the Y population are put on a life test, where one wants at most two Y units to actually fail, so the value $r = 2$ is chosen. Testing at significance level $\alpha = 0.05$, the critical value is $k = 7$, so H_0 is rejected if $x_{(7)} < y_{(2)}$ while H_0 is not rejected if $y_{(2)} < x_{(7)}$. Note that, with the test ending at time $\min(x_{(7)}, y_{(2)})$, there are at least 3 right-censored X observations and at least 6 right-censored Y observations; this leads to large imprecision in the NPI-RP values.

Table 4.11 presents the NPI lower and upper reproducibility probabilities for this test, for all possible data in the original test, which are indicated through the rankings of all observations until the test is ended, in the combined ranking of the X and Y samples. As indicated, the columns to the left relate to the cases where H_0 is not rejected while the columns to the right relate to the cases where H_0 is rejected. All these NPI-RP values are calculated using the monotonicity with regard to the combined ranks of the right-censored observations, as explained above. These results illustrate the earlier discussed maximum value 1 for $\overline{RP}$ and minimum value 0.25 for $\underline{RP}$. It is particularly noticeable that the NPI lower reproducibility probabilities for this test tend to be small, which is not really surprising due to the large number of right-censored observations resulting from the choice $r = 2$.

Table 4.12 presents NPI-RP for the precedence test with alternative hypothesis $H_1 : \lambda_x < \lambda_y$, considering $r = 5$, with $n_x = 10$, $n_y = 8$ and testing at significance level $\alpha = 0.05$, the critical value is $k = 10$. So H_0 is rejected in the original test if all observations in X are observed before $y_{(5)}$, i.e. $x_{(10)} < y_{(5)}$, whereas H_0 is not rejected if $y_{(5)} < x_{(10)}$. In case of not rejecting H_0 in the original test, there are

H_0 not rejected				H_0 rejected			
X ranks	Y ranks	$\underline{RP}$	$\overline{RP}$	X ranks	Y ranks	$\underline{RP}$	$\overline{RP}$
-	1,2	0.4992	1	1-7	-	0.3833	1
1	2,3	0.4951	0.9988	1-6,8	7	0.3367	0.8833
2	1,3	0.4970	0.9992	1-5,7,8	6	0.2993	0.8425
1,2	3,4	0.4826	0.9924	1-4,6-8	5	0.2739	0.8098
1,3	2,4	0.4884	0.9946	1-3,5-8	4	0.2593	0.7875
2,3	1,4	0.4903	0.9951	1,2,4-8	3	0.2526	0.7748
1-3	4,5	0.4553	0.9733	1,3-8	2	0.2504	0.7690
1-4	5,6	0.4075	0.9314	2-8	1	0.25	0.7670
1-5	6,7	0.3375	0.8582				
1-6	7,8	0.25	0.7509				
2-7	1,8	0.3663	0.8375				

Table 4.11: NPI-RP for precedence test with $n_x = 10$, $n_y = 8$, $r = 2$, $k = 7$ and $\alpha = 0.05$.

at most 3 right-censored Y observations and at least 1 right-censored X observation, while if the original test led to rejection of H_0, there are no right-censored observations in the X sample and at least 3 right-censored observations in Y sample. So in the case of rejecting H_0, we have complete data available from the X sample, but at least 1 right-censored observation in Y sample, and it is straightforward to derive the NPI lower and upper reproducibility probabilities. However, if the original test did not lead to rejection of H_0, both samples have right-censored observations, so the lower and upper reproducibility probabilities are the minimum and the maximum over all lower and upper reproducibility probabilities respectively, of all possible orderings of the right-censored observations. Comparing Table 4.11 and Table 4.12 illustrates that increasing the value of r, which leads to a decrease of the numbers of right-censored observations, leads to an increase of the values of the NPI lower and upper reproducibility probabilities.

H_0 not rejected $[\underline{P},\overline{P}](X^f_{(10)} > Y^f_{(5)})$				H_0 rejected $[\underline{P},\overline{P}](X^f_{(10)} < Y^f_{(5)})$			
X ranks	Y ranks	$\underline{RP}$	$\overline{RP}$	X ranks	Y ranks	$\underline{RP}$	$\overline{RP}$
-	1,5	0.5000	1	1-10	-	0.4936	1
1	2-6	0.5000	1.0000	1-9,11	10	0.4813	0.9936
2	1,3-6	0.5000	1.0000	1-8,10,11	9	0.4752	0.9902
1,2	3-7	0.4998	0.9999	1-7,9-11	8	0.4723	0.9885
1,3	2,4-7	0.4998	1.0000	1-6,8-11	7	0.4710	0.9877
2,3	1,4-7	0.4998	0.9999	1-5,7-11	6	0.4705	0.9874
1-3	4-8	0.4992	0.9996	1-4,6-11	5	0.4704	0.9872
1-4	5-9	0.4973	0.9985	1-3,5-11	4	0.4703	0.9872
1-5	6-10	0.4919	0.9947	1,2,4-11	3	0.4703	0.9872
1-6	7-11	0.4783	0.9840	1,3-11	2	0.4703	0.9872
2-7	1,8-11	0.4048	0.8816	2-11	1	0.4703	0.9872
1-9	10-14	0.25	0.7662	5-14	1-4	0.25	0.6904

Table 4.12: NPI-RP for precedence test with $n_x = 10$, $n_y = 8$, $r = 5$, $k = 10$ and $\alpha = 0.05$.

$\Diamond$

4.5 Concluding remarks

The NPI approach to reproducibility of tests provides many research challenges. It can be developed for many statistical tests, while for some data types, e.g. multivariate data, NPI requires to be developed further. The test scenarios studied for particular tests may require careful attention, as illustrated by the different perspectives discussed for the precedence test in Section 4.4. As mentioned, the precedence test scenarios discussed in Section 4.4 are very basic. Balakrishnan and Ng [10] present a detailed introduction and overview of precedence testing, including more sophisticated tests than the basic one considered in this chapter. In practice, it is important for such tests, and also in general, to also consider the power of the test; thus far this has not yet been considered in the NPI approach for reproducibility of testing. With further development of this approach, we are aiming at guidance on selection of test methods which, for specified level of significance, have good power and good reproducibility properties. This may often require more test data than needed following traditional guidance, but the assurance of good reproducibility is important for many applications and may lead to savings in the longer run by reducing processes, such as development of new medication, to continue on the basis of test results which may later turn out not to be reproduced in repeated tests under similar circumstances.

Chapter 5

Robustness of NPI

5.1 Introduction

As every statistical inference has underlying assumptions about models and specific methods used, one important field in statistics is the study of robustness of inferences. Statistical inferences are based on the data observations as well as the underlying assumptions, e.g. about randomness, independence and distributional models [62]. Since the middle of the twentieth century, many theoretical efforts have been dedicated to develop statistical procedures that are resistant with regard to outliers and robust with regard to small deviations from the assumed parametric model [15]. Huber [60] is a significant contributor to the development of robust statistical procedures. He provided in his graduate level textbook the basic robustness theory [60]. Further, Hampel [50, 51, 52] made significant contributions to the theory of robust estimation. Hampel, et al. [53] discussed some properties of robust estimators, test statistics and linear models. In these developments the primary focus has been on estimating location, scale, and regression parameters [63]. It is well known that some classical procedures are not robust to slight contamination of the strict model assumptions [15]. From this perspective robustness against small deviations from the assumed model and existence of outliers or contamination, have all been identified as principal issues [63].

In classical statistics, there are several tools used in robust statistics to describe robustness: the influence function, the sensitivity curve and the breakdown point (see Section 5.2 for a brief discussion of such tools). These three measures of robustness are not the only existing ones but they are commonly used methods.

In this chapter, robustness of NPI is presented. This involves adopting some of the concepts of classical robust statistics within the NPI setting, namely sensitivity curve and breakdown point. These concepts fit well with the NPI setting as they depend on the actual data at hand rather than on a hypothetical underlying assumption. Data may be subject to error occurring during the measurement process, so a reported value can differ from the true one [41]. Data contamination refers to errors in data which happen due to human error in collecting and recording the data, which introduces unintended changes to the original data. The concept of robust inference is usually aimed at development of inference methods which are not too sensitive to data errors or to deviations from the model assumptions. In this chapter, we use it in a slightly narrower sense, as for our aims robustness indicates insensitivity to a small change in the data and existence of outliers.

Outliers are generally described as observations that appear to be discordant with the rest of the data and are not appropriate for the assumed model [11]. One of the important reasons for detecting the presence of outliers is that they can have a crucial impact on parameter estimates of the model, which may cause erroneous inference. As a result, removing outliers may improve the accuracy of the estimators. However, outliers sometimes provide information that may be lost if the outlier is removed. Bellio and Ventura [15] stated that empirical evidence shows that good robust procedures behave relatively better than techniques based on the rejection of outliers. Robustness is a wide topic, and dealing with outliers is just one aspect of it. We could interpret robustness in two different ways, namely robustness with regard to small changes in data and robustness with regard to outliers. Our aim is to explore these two aspects for NPI for future order statistics as presented in this thesis.

This chapter is organized as follows. Section 5.2 provides a brief overview of some concepts used in robust statistics, namely influence function, sensitivity curve and breakdown point. In Section 5.3 we introduce the sensitivity curve and breakdown point in the NPI framework. Section 5.4 presents the use of these tools for NPI for events involving the r-th future observation. In Section 5.5 we use these tools to explore the robustness of the inferences involving the median and the mean of the m future observations. In Section 5.6, we briefly present NPI robustness of further inferences, namely pairwise comparisons and reproducibility of statistical tests, as presented in the previous chapters.

5.2 Classical concepts for evaluating robustness

In the literature of robustness, many measures of robustness of an estimator have been introduced [51, 53]. In this section, we review some concepts from classical theory of robust statistics, namely the influence function (IF), sensitivity curve (SC), empirical influence function (EIF) and breakdown point (BP). First, we consider the influence function (IF), an approach that is due to Hampel [51]. IF measures the influence of infinitesimal perturbations in a distribution on a statistic. Suppose we have a basic model with CDF F, and G_ξ is the CDF with a point mass at ξ.

Definition 5.1. For an estimator T based on data from a population with CDF F, the influence function of T at basic distribution F is

$$IF_{T,F}(\xi) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon G_\xi) - T(F)}{\epsilon} \tag{5.1}$$

Here $((1-\epsilon)F + \epsilon G_\xi)$ with $0 < \epsilon < 1$ is a mixture distribution of F and G_ξ. This definition of the IF depends on the assumed distribution as it assesses the effect of an infinitesimal perturbation in a distribution on the value of the estimator. There are several finite sample versions of (5.1), the most important being the sensitivity curve [81], the empirical influence function [53] and the jackknife [71, 72, 75, 80].

The first two of these we will also consider.

Tukey [81] proposed the sensitivity curve (SC), designed to assess the sensitivity of an estimator with regard to replacement or addition of an observation in the sample. SC also illustrates Tukey's emphasis on the nonprobabilistic aspects of statistics and data analysis. He preferred to rely on the actual data at hand rather than on a hypothetical underlying population of which it might be a sample [61]. This is in line with the NPI method, as in NPI we do not consider the hypothetical underlying population. For given data, we do not focus our inferences on a population from which the data are assumed to come, but on future data that are exchangeable with given data. There are two versions of SC, one with an additional observation and one with replacement of an observation [53]. First, we consider the case where we add an additional observation to the data. Many of the common estimators and test statistics depend on the sample $x_1, \ldots, x_n$ only through the empirical distribution function [62]

$$F_n(x) = \frac{\sum 1\{x_i < x\}}{n}$$

Thus we can write $T(F_n) = T(x_1, \ldots, x_n)$ for some estimator T [62].

Definition 5.2. Let $T_n(X) = T_n(x_1, .., x_n)$ denote a statistic of the sample $X = (x_1, .., x_n)$ and let $T_{n+1}(X, \xi)$ denote the corresponding statistic of the sample $x_1, .., x_n$, ξ. If we replace F by the empirical function F_n and ϵ by $1/(n+1)$ in (5.1), then the sensitivity curve (SC) is defined as [62]

$$\begin{aligned} SC_n(\xi, T_n, X) &= \frac{T\left(\frac{n}{n+1}F_n + \frac{1}{n+1}G_\xi\right) - T(F_n)}{1/(n+1)} \\ &= (n+1)\left(T_{n+1}(x_1, \ldots, x_n, \xi) - T_n(x_1, \ldots, x_n)\right) \\ &= (n+1)\left(T_{n+1}(X, \xi) - T_n(X)\right) \end{aligned} \tag{5.2}$$

$SC_n(\xi, T_n, X)$ measures the sensitivity of T_n to the addition of one observation with value ξ [62]. The sensitivity curve measures sensitivity of an estimator to a change in the sample. By contrast, the IF in Definition 5.1 shows what happens to an estimator when we change the distribution of the data slightly. There also exists a version of SC where instead of adding an observation ξ to the sample, one observation x_i is replaced by ξ.

Definition 5.3. We define the SC, if we replace F by the empirical function F_n and ϵ by $1/n$ in (5.1) [62]

$$\begin{aligned} SC_i(\xi, T_n, X) &= \frac{T\left(\frac{n-1}{n}F_n + \frac{1}{n}G_\xi\right) - T(F_n)}{1/n} \\ &= n\left(T_n(x_1 \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n) - T_n(x_1 \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)\right) \\ &= n\left(T_n(X, \xi, i) - T_n(X)\right) \end{aligned} \tag{5.3}$$

This version of SC measures the sensitivity of T_n to replacing the i-th value in the sample by an arbitrary value.

Definition 5.4. The empirical influence function of T_n for the sample X at ξ is

$$EIF_i(\xi, T_n, X) = T_n(x_1 \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n) \tag{5.4}$$

This EIF_i is defined by replacing the i-th value in the sample X by an arbitrary value ξ and looking at the output of the estimator [53]. Alternatively, one can define it by adding an obervation, i.e. when the original sample consists of n observations one can add an arbitrary value ξ, $EIF(\xi, T_n, X) = T_n(x_1, \ldots, x_n, \xi)$. The EIF gives us an idea of how an estimator behaves when we change one point in the sample X and no model assumptions are made [67].

In NPI we will adopt the two versions of the sensitivity curve (SC), as these give more insight than the EIF concept. The concepts defined above are local measurements, as they in principle examine the effect on an estimator of substituting a single contaminant for one of the n observations, or of adding a data point to the sample. In contrast, the breakdown point is a global measurement, as it gives the highest fraction of outliers one may have in the data before the estimator goes to infinity [63]. Let $X = (x_1, \ldots, x_n)$ be a fixed sample of size n.

We can contaminate this sample in many ways [62]. We consider the following two; λ_a replacement and λ_b contamination. These will also be considered in the NPI setting in Section 5.3.

1. λ_a replacement: we replace an arbitrary subset of size l of the sample by arbitrary values $y_1, \ldots, y_l$, so $1 \leq l \leq n$ [62]. Let $X^{'}$ denote the contaminated sample. The fraction of contaminated values in the contaminated sample $X^{'} = (x_1 \ldots, x_{l-1}, y_l, \ldots, y_n)$, is $\lambda_a = \frac{l}{n}$

2. λ_b contamination: we add l arbitrary additional values $Y = (y_1, \ldots, y_l)$ to the sample X [62]. Let $X^{''}$ denote the contaminated sample by adding l arbitrary additional values. Thus, the fraction of contaminated values in the contaminated sample $X^{''} = X \cup Y$ is $\lambda_b = \frac{l}{l+n}$

Definition 5.5. Let $T = (T_n)$ be an estimator and $T(X)$ be its value at the sample X. The maximum bias which might be caused by general λ, which is either λ_a or λ_b, is [62]

$$b(\lambda; X, T) = \sup |\{T(X, Y) - T(X)\}| \tag{5.5}$$

where the supremum is taken over the set of all λ-contaminated samples which is either $X^{'}$ or $X^{''}$.

Definition 5.6. The definition of the breakdown point is

$$\lambda^*(X, T) = \inf\{\lambda | b(\lambda; X, T) = \infty\} \tag{5.6}$$

The breakdown point $\lambda^*(X, T)$ of an estimator T at sample X is the smallest value of λ for which the estimator $T(X, Y)$ can have values arbitrarily far from $T(X)$.

For example, a contaminant in a sample of size n will, if large enough, shift the sample mean to be large, so the sample mean has unbounded influence function and breakdown point $1/n$. In contrast, the influence of a contaminant on the sample median is bounded, because $2k-2$ contaminants, whatever their magnitudes, added to a sample of size $n = 2k-1$, where k is the position of the median, can at most shift the sample median from x_k to x_1 or x_{2k-1} [11]. However, as soon as the number of contaminants added exceeds $2k-2$, the sample median can take any value. Thus the median has breakdown point $\frac{2k-1}{4k-2} = \frac{1}{2}$ and this is the highest value of the breakdown point which an estimator can achieve [11].

5.3 Robustness concepts in NPI

This section presents a first exploration of the use, and where needed adaptation, of the robustness concepts discussed in Section 5.2, within the NPI setting. It turns out that we need to adapt the concepts, but we can keep similar interpretations.

A simple way to study NPI robustness is to contaminate the given data and then explore its effect on our predictive inference. This approach is straightforward, gives an intuitive analysis, and is in line with the classic nonparametric robustness concepts, as they typically assess the influence on statistical inference of an arbitrary data value either added to the data or replacing an original observation.

We do not look at IF for NPI, as IF depends on the assumed distribution and in the NPI approach we do not assume any underlying distribution. Also we do not consider EIF and we consider SC as it gives more insight than EIF. In our study of the robustness of NPI, we will focus on the sensitivity curve and breakdown point as they typically rely on the actual data at hand rather than on a hypothetical underlying population.

From robustness for the NPI perspective, questions of interest include the following. Does contamination of the given observations influence the predictive inferences on future order statistics? Is the effect on the predictive inference proportional to the number of contaminants present? How does a single contaminant effect relate to its magnitude? These questions are similar as in the classical theory, where they are answered by the SC and BP of an estimator.

5.3.1 Notation

Let $\underline{x} = \{x_1, \ldots, x_n\}$ be a given sample of real-valued observations and let $I(\underline{x})$ be a predictive inference for future observations, based on the sample $\underline{x}$. Such a sample $\underline{x}$ can be contaminated in many ways, and we consider two of them.

1. Replacement: We replace a subset of size l of the data $\underline{x}$, $x_{j_1}, \ldots, x_{j_l}$, by $x_{j_1} + \delta, \ldots, x_{j_l} + \delta$, where $1 \leq l \leq n$. We denote these contaminated data by $\underline{x}(j_1, \ldots, j_l, \delta)$. Let $I(\underline{x}(j_1, \ldots, j_l, \delta))$ denote the inference of interest based on the contaminated data. The fraction of contaminant values in the contaminated sample $\underline{x}(j_1, \ldots, j_l, \delta)$ is $\lambda_a = \frac{l}{n}$. This includes the special case $l = 1$, where we replace one observation x_j of the sample by the value $x_j + \delta$, for any real-valued δ. The inference, with this replacement, is denoted by $I(\underline{x}(j, \delta))$.

2. Additional: We add l arbitrary additional observations $y_1, \ldots, y_l$ to the past data $\underline{x}$. We denote these contaminated data by $(\underline{x}, y_1, \ldots, y_l)$. The inference is denoted by $I(\underline{x}, y_1, \ldots, y_l)$. The fraction of contaminant values in the contaminated sample $(\underline{x}, y_1, \ldots, y_l)$, is $\lambda_b = \frac{l}{l+n}$.

These two ways of contaminating the sample will be studied separately in the NPI framework. We first focus on the effect of adding δ to one of the observations in the past data, as it is convenient and logical to do this in the NPI method.

The ordered observations of the contaminated sample are denoted by $\tilde{x}_1 < \ldots < \tilde{x}_n$. Adding δ to x_j might change its rank from j to l, where $l \geq j$ for $\delta > 0$ and $l \leq j$ for $\delta < 0$ and denote by $\tilde{x}_l$.

5.3.2 NPI concepts for evaluating robustness

In this subsection we introduce and discuss robustness concepts for NPI. We illustrate their use in the following sections for events involving the r-th ordered future observation, the median and the mean of the future observations, and further inferences involving the future observations that have been introduced in Chapters 2, 3 and 4.

To begin with we consider the case of replacement, substituting a contaminant for one of the n observations. A simple way to assess the effect of an arbitrary value δ added to one of the data observations on a particular predictive inference is to compute the difference between the value of the predictive inference with and without δ added.

Definition 5.7. The NPI-sensitivity curve (NPI-SC) for a predictive inference $I(\underline{x})$, in case of replacing one observation x_j by $x_j + \delta$, is defined by

$$SC_I(\underline{x}(j,\delta)) = I(\underline{x}(j,\delta)) - I(\underline{x}) \tag{5.7}$$

It can also be of interest to consider $nSC_I(\underline{x}(j,\delta))$, corresponding to the classical definition of the sensitivity curve as given in Definition 5.3. We may multiply $SC_I(\underline{x}(j,\delta))$ by n, but in our case Equation (5.7) is more straightforward, and it depends on n, so when n is large we expect $SC_I(\underline{x}(j,\delta))$ to become smaller. However, if one wants to compare sensitivity for different values of n, then one may need to multiply SC by n.

Example 5.1. To illustrate the use of $nSC_I(\underline{x}(j,\delta))$ and $SC_I(\underline{x}(j,\delta))$ for the NPI lower and upper probabilities for the event $X_{(r)} \geq z$, we consider artificial data sets consisting of the ordered numbers from 1 to n, for $n = 8, 20, 30, 40, \ldots, 100$, and $m = 3$ future observations. The effect of replacing $x_1 < z$ by $x_1 + \delta = \tilde{x}_l > z$ on the probability of the event that the second ordered future observation $X_{(2)}$, out of $m = 3$ future observations, is greater than z, where $z \in I_{\frac{n}{2}+1}$, and n is an even number, is reflected by the NPI-SC

$$\begin{aligned}
SC_{\underline{P}(X_{(2)}>z)}(\underline{x}(1,\delta)) &= \underline{P}_{\underline{x}(j,\delta)}(X_{(2)} > z) - \underline{P}_{\underline{x}}(X_{(2)} > z) \\
&= P(X_{(2)} \in I_{\frac{n}{2}+1}) \quad \text{if} \quad x_1 < z \quad \text{and} \quad \tilde{x}_l > z \\
SC_{\overline{P}(X_{(2)}>z)}(\underline{x}(1,\delta)) &= \overline{P}_{\underline{x}(j,\delta)}(X_{(2)} > z) - \overline{P}_{\underline{x}}(X_{(2)} > z) \\
&= P(X_{(2)} \in I_{\frac{n}{2}}) \quad \text{if} \quad x_1 < z \quad \text{and} \quad \tilde{x}_l > z
\end{aligned}$$

We plot the NPI-SC of $[\underline{P}, \overline{P}](X_{(2)} > z)$ as function of different values of n. Figure 5.1 shows $n \times SC_{[\underline{P},\overline{P}](X_{(2)} \geq z)}(\underline{x}(1, \delta))$ and $SC_{[\underline{P},\overline{P}](X_{(2)} \geq z)}(\underline{x}(1, \delta))$, for $m = 3$, where the lower probability is denoted by LP and the upper probability denoted by UP. The results clearly illustrate that the effect of replacing x_1 by $x_1 + \delta$ on the lower and upper probabilities for event $X_{(2)} > z$ decreases as the value of n increases.

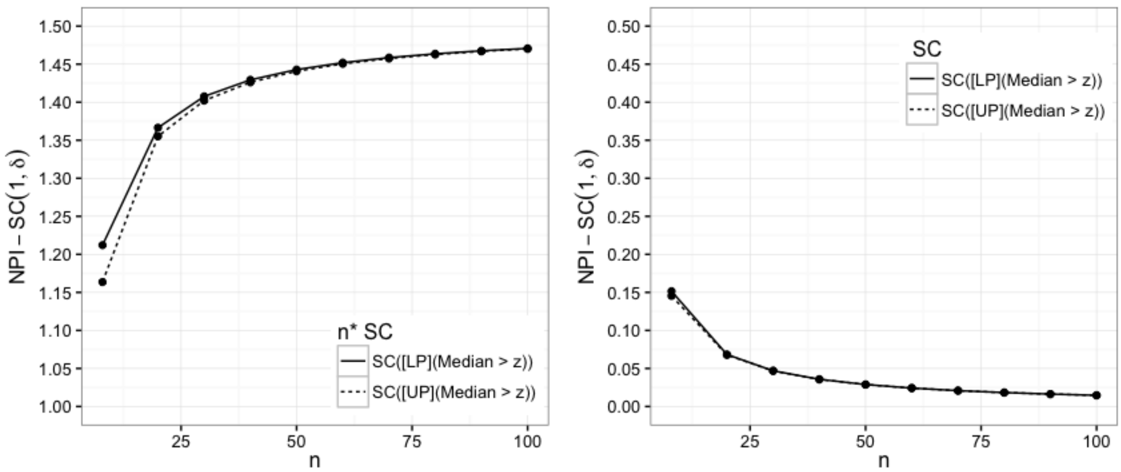

Figure 5.1: $nSC_{P(X_{(2)} \geq z)}(\underline{x}(1, \delta))$ and $SC_{P(X_{(2)} \geq z)}(\underline{x}(1, \delta))$

$\Diamond$

Definition 5.8. The NPI-SC, in the case of adding an additional observation y to the data, is

$$SC_I(\underline{x}, y) \quad = \quad I(\underline{x}, y) - I(\underline{x}) \tag{5.8}$$

This NPI-$SC_I(\underline{x}, y)$ assesses the sensitivity of an inference to the position of an additional observation, so it illustrates the impact of adding an additional observation y to the sample on the inferences involving future observations. In Example 5.2, we will illustrate NPI-SC for the event that $X_{(r)} \in I_j$, as a function of y where $y \in I^*_j$, for all $j^* = 1, \ldots, n+1$.

A finite sample breakdown point (BP) was first proposed by Hodges [59], as "tolerance of extreme values" in the situation of the location parameter problems, and it was generalized for a variety of cases by Hampel [50]. However, it has not been applied to situations of predictive inferences where the range of the inferences for the future observations is bounded, but it can easily be extended to such situations. We will modify the concept of BP to fit with the NPI approach. The maximum value of predictive inferences in terms of lower and upper probabilities is 1. We introduce a new definition of BP, which we call *the c-breakdown point*, and denote by $\lambda_c^*(\lambda, I, \underline{x}(\delta, j_1, \ldots, j_l))$.

Definition 5.9. The maximum bias which might be caused by λ_a-replacement, is

$$\begin{aligned} b(\lambda_a; \underline{x}, I) &= \sup |(I(\underline{x}(j_1, \ldots, j_l, \delta)) - I(\underline{x}))| \\ &= \sup |(SC_I(\underline{x}(j_1, \ldots, j_l, \delta))| \end{aligned} \tag{5.9}$$

where the supremum is taken over the set of all λ_a-replacement samples $\underline{x}(j_1, \ldots, j_l, \delta)$, with $\{j_1, \ldots, j_l\} \subset \{1, \ldots, n\}$ for fixed δ and given data $\underline{x}$. Alternatively, one can define the maximum bias by adding l contaminated values to the sample $\underline{x}$, so the maximum bias which might be caused by λ_b-contamination is

$$\begin{aligned} b(\lambda_b; \underline{x}, I) &= \sup |(I(\underline{x}, y_1, \ldots, y_l) - I(\underline{x}))| \\ &= \sup |(SC_I(\underline{x}, y_1, \ldots, y_l)| \end{aligned} \tag{5.10}$$

where the supremum is taken over the set of all λ_b-contaminated samples $(\underline{x}, y_1, \ldots, y_l)$, with $y_1, \ldots, y_l \in \mathbb{R}$ of given data $\underline{x}$.

Definition 5.10. The *c-breakdown point*, where $c \in [0, 1]$, for the case of λ_a-replacement, is defined as

$$\lambda_c^*(I, \underline{x}(\delta, j_1, \ldots, j_l)) = \inf\{\lambda_a | b(\lambda_a; \underline{x}, I) > c\} \tag{5.11}$$

where the infimum is taken over all possible fractions $\lambda_a = \frac{l}{n}$. Alternatively, the *c-breakdown point* for the case of adding l observations to the original sample (λ_b-contamination), is

$$\lambda_c^*(I, (\underline{x}, y_{j_1}, \ldots, y_{j_l})) = \inf\{\lambda_b | b(\lambda_b; \underline{x}, I) > c\} \tag{5.12}$$

where the infimum is taken over all possible fractions $\lambda_b = \frac{l}{l+n}$.

The *c-breakdown point* is the smallest fraction of contamination in the past data that could cause a predictive inference for future order statistics to take a value at least c away from the value of the initial predictive inference. This definition includes for, $c = 0$, the case when any change in the inference caused by l contaminated observations, is considered as breakdown point of the inference of interest. The value c determines how much we allow the inference to change before its breakdown.

5.4 Robustness of NPI for the rth future order statistic

To illustrate the use of the robustness concepts for NPI, namely NPI-SC and NPI-BP as defined in Section 5.3, we first consider the probabilities for events involving the r-th ordered future observation. We illustrate both ways that the sample can be contaminated.

5.4.1 NPI-SC for data replacement

To begin with, we explore how a contamination in the data affects the NPI probability for the event that $X_{(r)} \in I_k$ in Equation (2.2). The probability (2.2) is only affected by replacing contamination if the indices, $k = 1, \ldots, n+1$, differ. The effect of replacing an observation x_j by $x_j + \delta = \tilde{x}_l$, with $\delta \in \mathbb{R}$, on the probability for the event $X_{(r)} \in I_k$ is

$$
\begin{aligned}
SC_{\underline{P}(X_{(r)} \in I_k)}(\underline{x}(j,\delta)) &= \underline{P}_{\underline{x}(j,\delta)}(X_{(r)} \in (\tilde{x}_{k-1}, \tilde{x}_k)) - \underline{P}_{\underline{x}}(X_{(r)} \in I_k) \\
&= \begin{cases}
0 & \text{if} \quad x_j < x_k \quad \text{and} \quad \tilde{x}_l < x_k \\
P(X_{(r)} \in I_{k-1}) - P(X_{(r)} \in I_k) & \text{if} \quad x_j < x_k \quad \text{and} \quad \tilde{x}_l > x_k \\
\sum_{i=k+1}^{l} P(X_{(r)} \in I_i) & \text{if} \quad x_j = x_k \quad \text{and} \quad \tilde{x}_l > x_k \\
\sum_{i=l+1}^{k-1} P(X_{(r)} \in I_i) - P(X_{(r)} \in I_k) & \text{if} \quad x_j = x_k \quad \text{and} \quad \tilde{x}_l < x_k \\
P(X_{(r)} \in I_{k+1}) & \text{if} \quad x_j > x_k \quad \text{and} \quad \tilde{x}_l \in (x_{k-1}, x_k) \\
0 & \text{if} \quad x_j > x_k \quad \text{and} \quad \tilde{x}_l > x_k
\end{cases}
\end{aligned}
$$

The NPI lower and upper probabilities for the event $X_{(r)} > z$ are, in some cases, affected slightly by changing x_j to $x_j + \delta$. Let $z \in I_k = (x_{k-1}, x_k)$, then the effect of replacing an observation x_j by $x_j + \delta = \tilde{x}_l$, with $\delta \in \mathbb{R}$, on the NPI lower and upper probabilities for the event $X_{(r)} > z$, is

$$
\begin{aligned}
SC_{\underline{P}(X_{(r)}>z)}(\underline{x}(j,\delta)) &= \underline{P}_{\underline{x}(j,\delta)}(X_{(r)} > z) - \underline{P}_{\underline{x}}(X_{(r)} > z) \\
&= \begin{cases} 0 & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l < z \\ P(X_{(r)} \in I_k) & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l > z \\ -P(X_{(r)} \in I_{k-1}) & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l < z \\ 0 & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l > z \end{cases} \\
SC_{\overline{P}(X_{(r)}>z)}(\underline{x}(j,\delta)) &= \overline{P}_{\underline{x}(j,\delta)}(X_{(r)} > z) - \overline{P}_{\underline{x}}(X_{(r)} > z) \\
&= \begin{cases} 0 & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l < z \\ P(X_{(r)} \in I_{k-1}) & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l > z \\ -P(X_{(r)} \in I_k) & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l < z \\ 0 & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l > z \end{cases}
\end{aligned}
$$

This NPI-SC depends on the value of r and which interval it falls in, and will be illustrated in Example 5.2 in Subsection 5.4.4.

5.4.2 NPI-SC for additional data

Suppose we are interested in assessing the effect of an additional observation on the probability for the event that the rth ordered future observation falls in interval I_j, by considering

$$SC_{P(X_{(r)} \in I_j)}(\underline{x}, y) = P_{(\underline{x},y)}(X_{(r)} \in I_j) - P_{\underline{x}}(X_{(r)} \in I_j) \tag{5.13}$$

We let j^* be such that $y \in I_{j^*}$ If the method is robust to the new observation then $P(X_{(r)} \in I_j | y \in I_{j^*})$ should be close to $P(X_{(r)} \in I_j)$ for all r, j, j^*. The intuitive question we should investigate is when the influence is larger, if $j^* < j$, or $j^* = j$, or $j^* > j$? Thus, this $P(X_{(r)} \in I_j | y \in I_{j^*})$ needs to be studied with respect to the position of j^* and j. The $P(X_{(r)} \in I_j | y \in I_{j^*})$ can be derived using

Equation (2.2), as given in Appendix C.

Theorem 5.1. $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) > 0$ for $j^* < j$ if and only if $j \leq \frac{(r-1)(n+1)}{m}$ and for $j^* > j$ if and only if $j \geq \frac{r(n+1)}{m}+1$.

The proof of Theorem 5.1 is given in the Appendix D.

The SC for the event that $X_{(r)} \in I_j$, when we add an additional observation $y \in I_{j^*}$ where $j^* < j$ and $\tilde{I}_{j+1} = (\tilde{x}_j, \tilde{x}_{j+1}) = (x_{j-1}, x_j)$, is

$$\begin{aligned} SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) &= P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_j^*) - P(X_{(r)} \in I_j) \\ &= P(X_{(r)} \in I_j) \left[\frac{(j+r-1)(n+1)}{j(n+1+m)} - 1 \right] \\ &= P(X_{(r)} \in I_j) \left[\frac{(r-1)(n+1) - jm}{j(n+1+m)} \right] \end{aligned} \tag{5.14}$$

If $j^* > j$, so $\tilde{I}_j = I_j = (\tilde{x}_{j-1}, \tilde{x}_j)$, then

$$\begin{aligned} SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) &= P(X_{(r)} \in I_j | y \in I_j^*) - P(X_{(r)} \in I_j) \\ &= P(X_{(r)} \in I_j) \left[\frac{(n-j+2+m-r)(n+1)}{(n-j+2)(n+m+1)} - 1 \right] \\ &= P(X_{(r)} \in I_j) \left[\frac{m(j-1) - r(n+1)}{(n-j+2)(n+m+1)} \right] \end{aligned} \tag{5.15}$$

If $j^* > j$ and $j = \frac{r(n+1)}{m} + 1$ is an integer number then $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) = 0$, as illustrated in Example 5.2. If $j^* = j$, so I_j now becomes $\tilde{I}_j \cup \tilde{I}_{j+1}$ where $\tilde{I}_j = (x_{j-1}, y)$ and $\tilde{I}_{j+1} = (y, x_j)$, then NPI-SC for $P(X_{(r)} \in I_j)$ is

$$\begin{aligned} SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) &= \left[P(X_{(r)} \in \tilde{I}_j | y \in I_j) + P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_j) \right] - P(X_{(r)} \in I_j) \\ &= P(X_{(r)} \in I_j) \left[\frac{(r-1)(n+1) - jm}{j(n+1+m)} + \frac{m(j-1) - r(n+1)}{(n-j+2)(n+m+1)} \right] \end{aligned}$$

The NPI-SC measures how a single contaminant, whether added or substituted, affects an inference of interest, which is in line with SC in classical robustness.

5.4.3 NPI-BP for data replacement and adding

We illustrate the NPI-BP for the lower and upper probabilities for the event that $X_{(r)} > z$, where $z \in (x_{k-1}, x_k)$. Suppose we keep $x_1, \ldots, x_{k-1}$ fixed and let

$x_k, \ldots, x_n$ go to infinity, then the NPI lower and upper probabilities for the event that $X_{(r)} > z$, will not change at all. However, when we only keep $x_1, \ldots, x_{k-2}$ fixed and let $x_{k-1}, \ldots, x_n$ go to infinity then $[\underline{P}, \overline{P}](X_{(r)} > z)$ will increase. For $c = 0$ the minimum fraction of the contaminated values in the contaminated sample that can cause $b(\lambda_a; \underline{x}, [\underline{P}, \overline{P}](X_{(r)} > z)) > 0$, is

$$\lambda_0^*([\underline{P}, \overline{P}](X_{(r)} > z), \underline{x}(\delta, k-1, \ldots, n)) = \frac{n-k+2}{n} \tag{5.16}$$

An effect on such an inference occurs only when the contaminated values lead to change of the number of the observations that are greater than z. The value of the c-breakdown point decreases as the value of k increases, where I_k is the interval that z falls in. Similarly, the c-breakdown point for the probability for the event that $X_{(r)} \in I_k$ is $\frac{n-k+2}{n}$.

In the case of adding observations to the data, the c-breakdown point for the probability for the event that $X_{(r)} \in I_i$, for $c = 0$, is

$$\begin{aligned} \lambda_0^*(P(X_{(r)} \in I_i), (\underline{x}, y_{j_1}, \ldots, y_{j_l})) &= \inf\{\lambda_b | b(\lambda_b; \underline{x}, P(X_{(r)} \in I_i)) > 0\} \\ \lambda_0^*(P(X_{(r)} \in I_i), (\underline{x}, y_{j_1})) &= \frac{1}{n+1} \end{aligned} \tag{5.17}$$

Thus, adding a single data observation will change the probability for the event that $X_{(r)} \in I_j$. The size of the change varies depending on which order statistic is considered and in which interval it is, which will be illustrated in Example 5.2. Similarly, in the case of additional observations to the sample, the c-breakdown point for the event that $X_{(r)} > z$, for $c = 0$ is

$$\lambda_0^*([\underline{P}, \overline{P}](X_{(r)} > z), (\underline{x}, y_{j_1})) = \frac{1}{n+1} \quad \text{for} \quad z \in I_k$$

Thus, if we add a single data observation, it will change the probability of the event that $X_{(r)} > z$.

We have only considered the NPI-BP for $c = 0$ here. In Example 5.2, we will also illustrate NPI-BP for $c > 0$.

5.4.4 Example

We illustrate the NPI-SC and NPI-BP presented in this section by the following example.

Example 5.2. We consider data set $\underline{x} = \{-9, -7, 0, 2, 5, 7, 10, 16\}$, and the corrupted sample $\underline{x}(2, \delta)$, where we replace $x_2 = -7$ by $-7 + \delta$ for $\delta \in \mathbb{R}$.

Table 5.1 presents the NPI-SC for the lower and upper probabilities for the event $X_{(r)} \geq 1$, for $m = 5$ and $r = 1, \ldots, 5$. These inferences are not affected by adding $\delta < 8$ to x_2, as $x_2 + \delta < 1$, whereas for $\delta \geq 8$ the value $x_2 + \delta > 1$, which changes the values of the lower and upper probabilities by an amount $P(X_{(r)} \in I_4)$, and $P(X_{(r)} \in I_3)$, respectively. The results illustrate that the largest effect of replacing $x_2 = -7$ by $-7 + \delta$, for $\delta \geq 8$, occurs for $r = 2$ and the smallest effect occurs for $r = 5$.

	$\delta < 8$		$\delta \geq 8$	
r	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
1	0	0	0.09790	0.16317
2	0	0	0.17405	0.19580
3	0	0	0.16317	0.13054
4	0	0	0.09324	0.05439
5	0	0	0.02720	0.01166

Table 5.1: $SC_{P(X_{(r)} \geq 1)}(\underline{x}(j, \delta))$ for $m = 5$

To illustrate the NPI-BP, we consider the data set $\underline{x}$ and the case with $m = 5$ and interest in event $X_{(r)} \geq 1$. Figure 5.2 and Table 5.2 present the NPI-SC for the NPI lower and upper probabilities for $X_{(r)} \geq 1$ for the values $r = 1, \ldots, 5$, in the case where we keep $x_1, \ldots, x_{8-l}$ and we added $\delta = 100$ to $x_{9-l}, \ldots, x_8$ for $l = 1, \ldots, 8$. The results clearly show that as the value of r increases the effect of replacing l observations by contaminated values on the NPI lower and upper probabilities for $X_{(r)} \geq 1$ is decreasing. If we chose $c = 0.15$, as presented by the red line in Figure 5.2, then the maximum NPI-BP for the event $X_{(r)} \geq 1$ is when $r = 5$, whereas the minimum NPI-BP is for which $r = 2$. The higher the breakdown point of an inference, the more robust it is. $\lambda_0^*(\underline{P}(X_{(1)} \geq 1), \underline{x}(2, \ldots, 8, 100)) = \lambda_0^*(\overline{P}(X_{(3)} \geq 1), \underline{x}(2, \ldots, 8, 100)) = \frac{7}{8}$ whereas the NPI-BP for the lower and upper probabilities for $X_{(2)} \geq 1$ and the lower probability for $X_{(3)} \geq 1$ is $\frac{6}{8}$ and for the lower probability for $X_{(4)} \geq 1$ is one whereas for upper probability for $X_{(4)} \geq 1$ does not breakdown. For $r = 5$ the inferences did not breakdown.

	$r=1$		$r=2$		$r=3$		$r=4$		$r=5$	
l	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
6	0.0979	0.1632	0.1740	0.1958	0.1632	0.1305	0.0932	0.0544	0.0272	0.0117
7	0.2611	0.4196	0.3699	0.3823	0.2937	0.2145	0.1476	0.0793	0.0389	0.0155
8	0.5175	0.8042	0.5563	0.5105	0.3776	0.2494	0.1725	0.0862	0.0427	0.0163

Table 5.2: $SC_{P(X_{(r)}\geq 1)}(\underline{x}(9-l,\ldots,8,100))$ for $m=5$

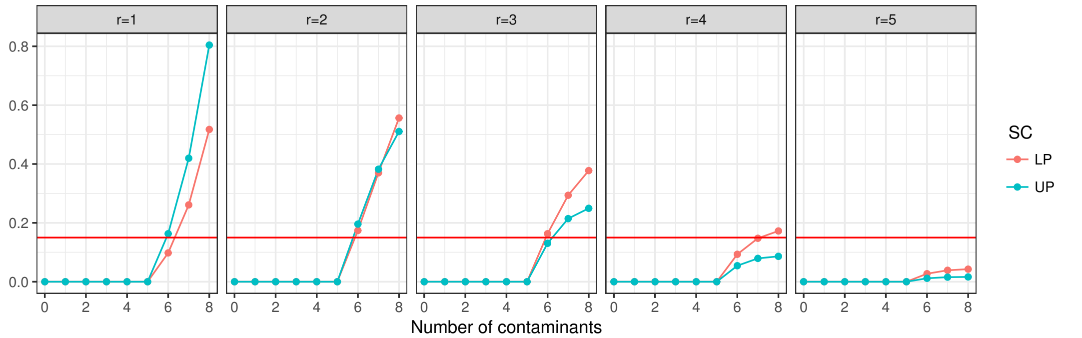

Figure 5.2: $SC_{P(X_{(r)}\geq 1)}(\underline{x}(9-l,\ldots,8,100))$ for $m=5$

To illustrate the possible effect on $P(X_{(r)} \in I_j)$ of an additional value y to the original sample, we consider $n = 8$ data observations and $m = 3$ future observations. Table 5.3 illustrates the NPI-SC for the event that $X_{(1)} \in I_j$, for $j = 1,\ldots,9$, and for all possible values of y that might fall in $I_{j^*} = (x_{j^*-1}, x_{j^*})$ for $j^* = 1,\ldots,9$. The results clearly illustrate that the effect of adding y to any intervals to the left of interval I_j, on the NPI probability for $X_{(1)} \in I_j$, is the same, and also if y is added to any intervals to the right of I_j.

Figures 5.3, 5.4 and 5.5 illustrate the NPI-SC for the event $X_{(r)} \in I_j$, for $r = 1,2,3$, $j = 1,\ldots,9$ and $j^* < j$, $j^* = j$ and $j^* > j$. These figures illustrate that $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y)$ is symmetric, i.e. $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y) = SC_{P(X_{(m+1-r)}\in I_{n+2-j})}(\underline{x}, y)$, so e.g. $SC_{P(X_{(1)}\in I_9)}(\underline{x}, y) = SC_{P(X_{(3)}\in I_1)}(\underline{x}, y)$. For all r, the NPI-SC for $X_{(r)} \in I_j$ is unimodal in j. Note that the values 0 in Table 5.3, for $j = 4$ and $j^* > 4$, illustrate the property mentioned immediately before Equation (5.15).

	j^*								
j	1	2	3	4	5	6	7	8	9
1	**0.1818**	-0.0227	-0.0227	-0.0227	-0.0227	-0.0227	-0.0227	-0.0227	-0.0227
2	-0.0545	**0.1500**	-0.0136	-0.0136	-0.0136	-0.0136	-0.0136	-0.0136	-0.0136
3	-0.0424	-0.0424	**0.1212**	-0.0061	-0.0061	-0.0061	-0.0061	-0.0061	-0.0061
4	-0.0318	-0.0318	-0.0318	**0.0955**	0	0	0	0	0
5	-0.0227	-0.0227	-0.0227	-0.0227	**0.0727**	0.0045	0.0045	0.0045	0.0045
6	-0.0152	-0.0152	-0.0152	-0.0152	-0.0152	**0.0530**	0.0076	0.0076	0.0076
7	-0.0091	-0.0091	-0.0091	-0.0091	-0.0091	-0.0091	**0.0364**	0.0091	0.0091
8	-0.0045	-0.0045	-0.0045	-0.0045	-0.0045	-0.0045	-0.0045	**0.0227**	0.0091
9	-0.0015	-0.0015	-0.0015	-0.0015	-0.0015	-0.0015	-0.0015	-0.0015	**0.0121**

Table 5.3: $SC_{P(X_{(1)} \in I_j)}(\underline{x}, y)$

j	$j^* < j$	$j^* > j$	$j^* = j$
1	-	-0.02273	0.18182
2	-0.05455	-0.01364	0.15000
3	-0.04242	-0.00606	0.12121
4	-0.03182	0	0.09545
5	-0.02273	0.00455	0.07273
6	-0.01515	0.00758	0.05303
7	-0.00909	0.00909	0.03636
8	-0.00455	0.00909	0.02273
9	-0.00152	-	0.01212

Table 5.4: $SC_{P(X_{(1)} \in I_j)}(\underline{x}, y)$

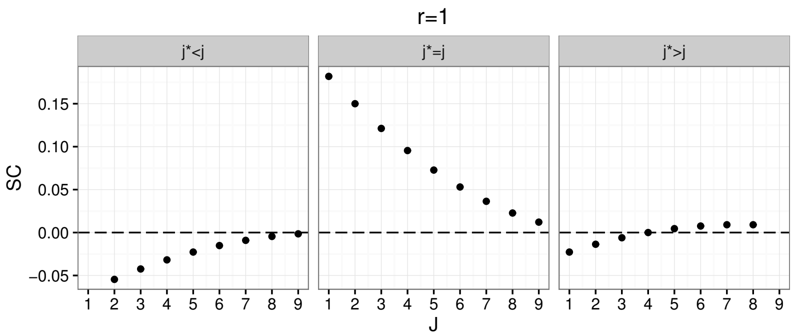

Figure 5.3: $SC_{P(X_{(1)} \in I_j)}(\underline{x}, y)$ for $n = 8$ and $m = 3$.

To illustrate the c-breakdown point λ_c^* for the event $X_{(r)} \in I_j$, we choose $c = 0.05$ and plot the absolute value of $SC_{P(X_{(r)} \in I_j)}(\underline{x}, y_1, \ldots, y_l)$ as function of l, where l is the number of the contaminated values that have been added to the data set of size $n = 8$. These are given in Figures 5.6, 5.7 and 5.8 and Tables 5.5 and 5.6 for $r = 1, 2, 3$. For $r = 1$ and $j \geq 3$, the probability for the event $X_{(1)} \in I_j$ does not break down, whereas for $j = 1$, $\lambda^*_{0.05}(P(X_{(1)} \in I_1), (\underline{x}, y_9, y_{10}, y_{11})) = \frac{3}{11} = 0.2727$

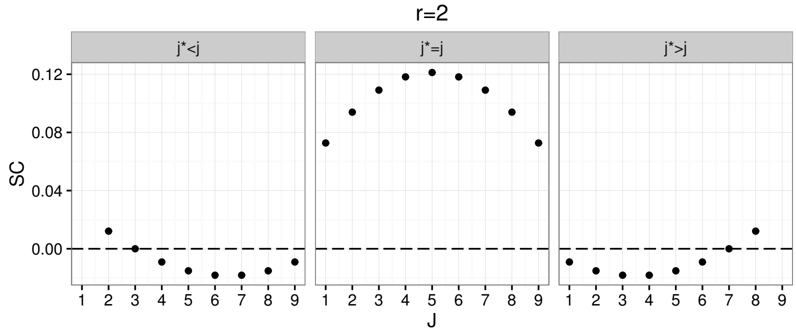

Figure 5.4: $SC_{P(X_{(2)}\in I_j)}(\underline{x}, y)$ for $n = 8$ and $m = 3$.

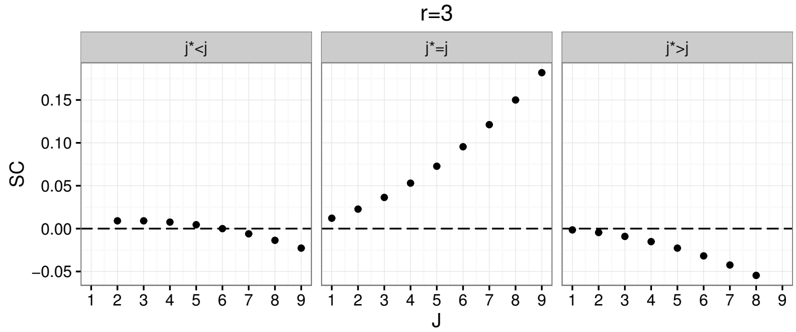

Figure 5.5: $SC_{P(X_{(3)}\in I_j)}(\underline{x}, y)$ for $n = 8$ and $m = 3$.

r	l	$j=1$	l	$j=2$	l	$j=3$	l	$j=4$	l	$j=5$
1	2	0.0420	4	0.0468	7	0.0410	7	0.0158	7	0.0047
	3	0.0584	5	0.0557	8	0.0459	8	0.0189	8	0.0030
2	7	0.0349	4	0.0442	3	0.0449	3	0.0466	3	0.0416
	8	0.0370	5	0.0505	4	0.0547	4	0.0575	4	0.0526
3	7	0.0048	7	0.0145	7	0.0290	7	0.0484	3	0.0497
	8	0.0050	8	0.0151	8	0.0302	8	0.0503	4	0.0579

Table 5.5: $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y_1, \ldots, y_l)$ for $m = 3$

and for $j = 2$, $\lambda^*_{0.05}(P(X_{(1)} \in I_2), (\underline{x}, y_9, \ldots, y_{13})) = \frac{5}{13} = 0.3846$. Figure 5.7 presents the absolute value of the NPI-SC for $X_{(2)} \in I_j$, where for $j = 3, 4, 5$ the NPI-BP is $\frac{4}{12} = 0.3333$, for $j = 2$ is $\frac{5}{13} = 0.3846$ and for $j = 6$ is $\frac{6}{14} = 0.4286$. The probability for the event $X_{(2)} \in I_j$ for $j = 1, 7, 8, 9$, does not break down as $SC_{P(X_{(2)}\in I_j)}(\underline{x}, y_1, \ldots, y_8) < 0.05$. Figure 5.8 shows that for $r = 3$, and $j = 4$, $\lambda^*_{0.05}(P(X_{(3)} \in I_j), (\underline{x}, y_9, \ldots, y_{16})) = \frac{8}{16} = 0.5$, whereas as j increases the NPI-BP decreases, such that for $j = 8, 9$, $\lambda^*_{0.05}(P(X_{(3)} \in I_j), (\underline{x}, y_9)) = 1/9$.

r	l	$j=6$	l	$j=7$	l	$j=8$	l	$j=9$
1	7	0.0203	7	0.0310	7	0.0370	7	0.0381
	8	0.0199	8	0.0317	8	0.0386	8	0.0404
2	5	0.0490	7	0.0415	7	0.0087	7	0.0337
	6	0.0572	8	0.0478	8	0.0144	8	0.0290
3	1	0.0318	1	0.0424	1	0.0545	1	0.0682
	2	0.0538	2	0.0718	2	0.0923	2	0.1154

Table 5.6: $SC_{P(X_{(r)}\in I_j)}(\underline{x}, y_1, \ldots, y_l)$ for $m=3$

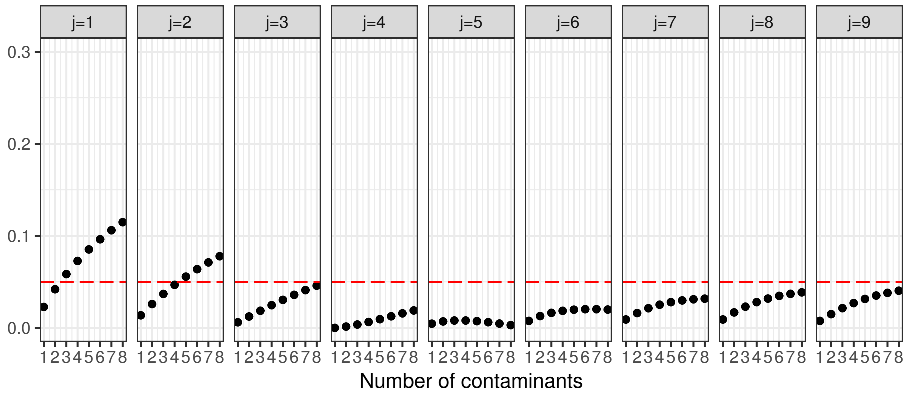

Figure 5.6: $SC_{P(X_{(1)}\in I_i)}(\underline{x}, y_1, \ldots, y_l)$ for $n=8$ and $m=3$.

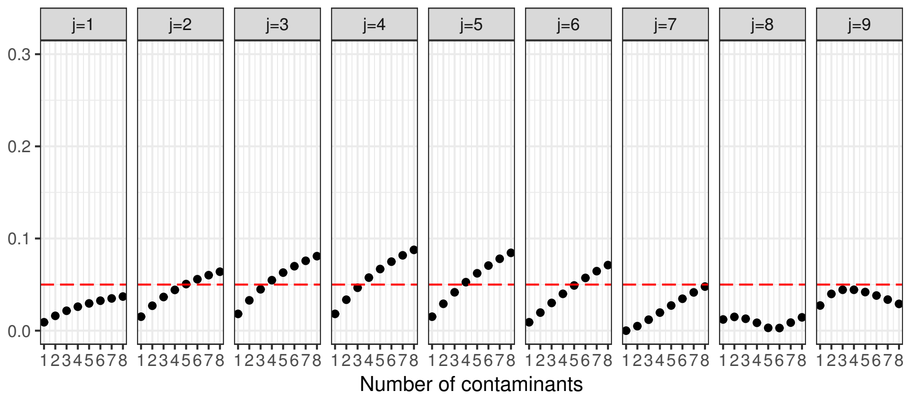

Figure 5.7: $SC_{P(X_{(2)}\in I_i)}(\underline{x}, y_1, \ldots, y_l)$ for $n=8$ and $m=3$.

$\Diamond$

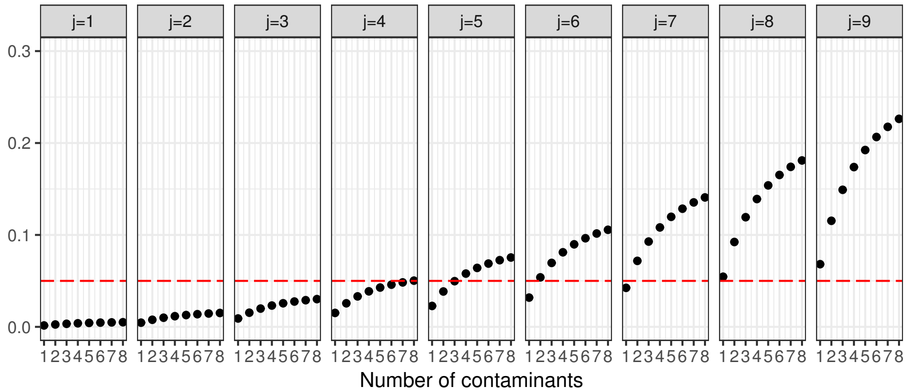

Figure 5.8: $SC_{P(X_{(3)} \in I_i)}(\underline{x}, y_1, \ldots, y_l)$ for $n = 8$ and $m = 3$.

5.5 Robustness of the median and mean of the future observations

In the classical robustness literature there has been quite a lot of emphasis on robust estimation of a location parameter, where typically they compare the robustness of the mean and the median. In this section, we illustrate the use of the robustness concepts for NPI, namely NPI-SC and NP-BP, by considering events involving the median and the mean of the m future observations.

5.5.1 Median of the m future observations

We first examine how contamination in the data affects NPI for an event involving the median of the m future observations, for m is odd. We consider the NPI-SC for the lower and upper probabilities for the event $M_m < z$. We wish to examine the effect on $[\underline{P}, \overline{P}](M_m < z)$ of adding a contaminant δ to one of the observations x_j with $j = 1, \ldots, n$. Let $z \in I_k = (x_{k-1}, x_k)$, if we add δ to x_j this becomes $\tilde{x}_l = x_j + \delta$, where $\delta \in \mathbb{R}$. The NPI-SC for event $M_m < z$ is

$$SC_{\underline{P}(M_m<z)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if } x_j > z \text{ and } \tilde{x}_l > z \\ 0 & \text{if } x_j < z \text{ and } \tilde{x}_l < z \\ P(M_m \in I_k) & \text{if } x_j > z \text{ and } \tilde{x}_l < z \\ -P(M_m \in I_{k-1}) & \text{if } x_j < z \text{ and } \tilde{x}_l > z \end{cases}$$

$$SC_{\overline{P}(M_m<z)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l > z \\ 0 & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l < z \\ P(M_m \in I_{k+1}) & \text{if} \quad x_j > z \quad \text{and} \quad \tilde{x}_l < z \\ -P(M_m \in I_k) & \text{if} \quad x_j < z \quad \text{and} \quad \tilde{x}_l > z \end{cases}$$

The NPI-SC for lower and upper probabilities for the event $M_m < z$ is a step function, with the step occurring when the contamination value changes the number of intervals to the right of z.

Next we consider the NPI-SC for the lower and upper probability for the event that $M_m \in (z_1, z_2)$. Let $z_1 \in I_k$ and $z_2 \in I_d$ where $k \leq d$. If we add δ to one of the data observations, i.e. x_j is replaced by $\tilde{x}_l$, then there are three possible situations. The effect of adding δ to x_j is to change the value of the NPI lower and upper probabilities for the event $M_m \in (z_1, z_2)$, by an amount NPI-SC as specified for each case below. First, if $x_j < z_1$

$$\begin{aligned} SC_{\underline{P}(M_m\in(z_1,z_2))}(\underline{x}(j,\delta)) &= \underline{P}_{\underline{x}(j,\delta)}(M_m \in (z_1,z_2)) - \underline{P}_{\underline{x}}(M_m \in (z_1,z_2)) \\ &= \begin{cases} 0 & \text{if} \quad \tilde{x}_l < z_1 \\ P(M_m \in I_k) & \text{if} \quad \tilde{x}_l \in (z_1,z_2) \\ P(M_m \in I_k) - P(M_m \in I_{d-1}) & \text{if} \quad \tilde{x}_l > z_2 \end{cases} \end{aligned} \tag{5.18}$$

$$\begin{aligned} SC_{\overline{P}(M_m\in(z_1,z_2))}(\underline{x}(j,\delta)) &= \overline{P}_{\underline{x}(j,\delta)}(M_m \in (z_1,z_2)) - \overline{P}_{\underline{x}}(M_m \in (z_1,z_2)) \\ &= \begin{cases} 0 & \text{if} \quad \tilde{x}_l < z_1 \\ P(M_m \in I_{k-1}) & \text{if} \quad \tilde{x}_l \in (z_1,z_2) \\ P(M_m \in I_{k-1}) - P(M_m \in I_d) & \text{if} \quad \tilde{x}_l > z_2 \end{cases} \end{aligned} \tag{5.19}$$

Secondly, if $x_j > z_2$

$$SC_{\underline{P}(M_m\in(z_1,z_2))}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad \tilde{x}_l > z_2 \\ P(M_m \in I_d) & \text{if} \quad \tilde{x}_l \in (z_1,z_2) \\ P(M_m \in I_d) - P(M_m \in I_{k+1}) & \text{if} \quad \tilde{x}_l < z_1 \end{cases}$$

$$SC_{\overline{P}(M_m \in (z_1,z_2))}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if } \tilde{x}_l > z_2 \\ P(M_m \in I_{d+1}) & \text{if } \tilde{x}_l \in (z_1,z_2) \\ P(M_m \in I_{d+1}) - P(M_m \in I_k) & \text{if } \tilde{x}_l < z_1 \end{cases}$$

Thirdly, if $x_j \in (z_1, z_2)$

$$SC_{\underline{P}(M_m \in (z_1,z_2))}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if } \tilde{x}_l \in (z_1,z_2) \\ -P(M_m \in I_{d-1}) & \text{if } \tilde{x}_l > z_2 \\ -P(M_m \in I_{k+1}) & \text{if } \tilde{x}_l < z_1 \end{cases} \quad (5.20)$$

$$SC_{\overline{P}(M_m \in (z_1,z_2))}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if } \tilde{x}_l \in (z_1,z_2) \\ -P(M_m \in I_d) & \text{if } \tilde{x}_l > z_2 \\ -P(M_m \in I_k) & \text{if } \tilde{x}_l < z_1 \end{cases} \quad (5.21)$$

So, when the data are contaminated and that contamination does not affect the number of intervals in (z_1, z_2) then there is no effect on this inference at all, which is an attractive property. But this is not the same if m is even. The probability for the event involving the median of the m future observations, where m is even, given in Equations (2.29) and (2.30), depends on the actual observations and not only on the ordering of the observations. For example, if we consider the lower and upper probabilities for the event that $M_m \geq z$, where $M_m = \frac{X_{(\frac{m}{2})}+X_{(\frac{m}{2}+1)}}{2}$, then the NPI-SC for this event, if we add δ to x_j, are

$$\begin{aligned} &SC_{\underline{P}(M_m \geq z)}(\underline{x}(j,\delta)) = \\ &\left[\sum_{i=1}^{n+1}\sum_{k=i}^{n+1} 1\{(\frac{\tilde{x}_{i-1}+\tilde{x}_{k-1}}{2}, \frac{\tilde{x}_i+\tilde{x}_k}{2}) \geq z\} P(X_{(\frac{m}{2})} \in \tilde{I}_i, X_{(\frac{m}{2}+1)} \in \tilde{I}_k)\right] \\ &-\left[\sum_{i=1}^{n+1}\sum_{k=i}^{n+1} 1\{(\frac{x_{i-1}+x_{k-1}}{2}, \frac{x_i+x_k}{2}) \geq z\} P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_k)\right] \end{aligned} \quad (5.22)$$

$$\begin{aligned} &SC_{\overline{P}(M_m \geq z)}(\underline{x}(j,\delta)) = \\ &\left[\sum_{i=1}^{n+1}\sum_{k=i}^{n+1} 1\{(\frac{\tilde{x}_{i-1}+\tilde{x}_{k-1}}{2}, \frac{\tilde{x}_i+\tilde{x}_k}{2}) \cap (z,\infty) \neq \emptyset\} P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_k)\right] \\ &-\left[\sum_{i=1}^{n+1}\sum_{k=i}^{n+1} 1\{(\frac{x_{i-1}+x_{k-1}}{2}, \frac{x_i+x_k}{2}) \cap (z,\infty) \neq \emptyset\} P(X_{(\frac{m}{2})} \in I_i, X_{(\frac{m}{2}+1)} \in I_k)\right] \end{aligned} \quad (5.23)$$

We also consider the c-breakdown point for the NPI lower and upper probabilities for the event $M_m > z$, where $z \in I_k$. The NPI lower and upper probabilities for such an event depend only on the number of observations that are greater than z, so in the sample of n observations, only $n-k+2$ outliers cause these probabilities to change. The c breakdown point for the event $M_m > z$ is similar as presented in Section 5.4, if we replace $X_{(r)}$ by M_m, where m is odd, in Equation (5.16). For m even, the c-breakdown point for the NPI lower and upper probabilities for the event $M_m > z$ and $c = 0$ is

$$\lambda_0^*([\underline{P}, \overline{P}](M_m > z), \underline{x}(n, \delta)) = \frac{1}{n} \quad \text{for} \quad z \in I_k$$

5.5.2 Mean of the m future observations

We consider the NPI-SC for the mean of the m future observations. It is well known that the mean of the population in classical statistics is more sensitive than the median to any change in the data [62]. In this subsection we investigate the robustness of inferences involving the mean of the m future observations.

To begin with, we consider the NPI-SC for the NPI lower and upper expected values for μ_m, as given in Equations (2.33) and (2.34). Suppose that one observation x_j is replaced by $x_j + \delta$, then the NPI-SC for the lower and upper expectation for the mean of the m future observations is $SC_{\underline{E}(\mu)}(\underline{x}(j,\delta)) = SC_{\overline{E}(\mu)}(\underline{x}(j,\delta)) = \frac{\delta}{n+1}$. The lower and upper bounds for the mean of the m future observations given the ordering O_i, as given in Equations (2.31) and (2.32), depend on the value of s_j^i. The NPI-SC for the lower and upper bounds of the μ_m^i, for the case that $x_j + \delta$ did not shift from its rank among the observations, i.e. $x_{j-1} < x_j + \delta < x_{j+1}$, are

$$SC_{\underline{\mu_m^i}}(\underline{x}(j,\delta)) = \frac{1}{m} s_{j+1}^i \delta \tag{5.24}$$

$$SC_{\overline{\mu_m^i}}(\underline{x}(j,\delta)) = \frac{1}{m} s_j^i \delta \tag{5.25}$$

If the value of $s_j^i = s_{j+1}^i = 0$ then there is no influence at all on the lower and upper μ_m^i, whereas for $s_j^i = m$ or $s_{j+1}^i = m$ then NPI-SC of the lower or the upper bounds for μ_m^i, will exceed any bound for δ large or small enough.

The NPI-SC for the lower and upper mean of the m future observations given the ordering O_i, for the case that $x_j + \delta$ gets another rank than x_j, i.e. x_j becomes $\tilde{x}_l$, for $\delta > 0$ and $l > j$ or for $\delta < 0$ and $l < j$, are

$$SC_{\underline{\mu^i_m}}(\underline{x}(j,\delta)) = \frac{1}{m}\left[\sum_{k=j}^{l} s^i_{k+1}[\tilde{x}_k - x_k]\right] \tag{5.26}$$

$$SC_{\overline{\mu^i_m}}(\underline{x}(j,\delta)) = \frac{1}{m}\left[\sum_{k=j}^{l} s^i_{k}[\tilde{x}_k - x_k]\right] \tag{5.27}$$

The NPI-SC for $\mu_m \geq z$, if x_j becomes $x_j + \delta = \tilde{x}_l$ and $\delta \in \mathbb{R}$, is

$$\begin{aligned}
SC_{\underline{P}(\mu_m \geq z)}(\underline{x}(j,\delta)) &= \underline{P}_{\underline{x}(j,\delta)}(\mu_m \geq z) - \underline{P}_{\underline{x}}(\mu_m \geq z) \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\left[1\{\underline{\mu^i_m}(\underline{x}(j,\delta)) \geq z\} - 1\{\underline{\mu^i_m}(\underline{x}) \geq z\}\right] \\
SC_{\overline{P}(\mu_m \geq z)}(\underline{x}(j,\delta)) &= \overline{P}_{\underline{x}(j,\delta)}(\mu_m \geq z) - \overline{P}_{\underline{x}}(\mu_m \geq z) \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\left[1\{\overline{\mu^i_m}(\underline{x}(j,\delta)) \geq z\} - 1\{\overline{\mu^i_m}(\underline{x}) \geq z\}\right]
\end{aligned}$$

Special case if $x_{j-1} < x_j + \delta < x_j$, so $\underline{\mu^i_m}(\underline{x}(j,\delta)) = \underline{\mu^i_m}(\underline{x}) + \frac{s^i_{j+1}\delta}{m}$, then the NPI-SC for $\mu_m \geq z$ is

$$\begin{aligned}
SC_{\underline{P}(\mu_m \geq z)}(\underline{x}(j,\delta)) &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\left[1\{\underline{\mu^i_m}(\underline{x}(j,\delta)) \geq z\} - 1\{\underline{\mu^i_m}(\underline{x}) \geq z\}\right] \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i)\left[1\{[\underline{\mu^i_m} + \frac{s^i_{j+1}\delta}{m}] \geq z\} - 1\{\underline{\mu^i_m} \geq z\}\right] \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) 1\{\underline{\mu^i_m} \leq z \leq \underline{\mu^i_m} + \frac{s^i_{j+1}\delta}{m}\} \qquad (5.28) \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) 1\{0 \leq z - \underline{\mu^i_m} \leq \frac{s^i_{j+1}\delta}{m}\} \\
&= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) 1\{0 \leq \frac{m(z - \underline{\mu^i_m})}{s^i_{j+1}} \leq \delta\}
\end{aligned}$$

Similarly, the NPI-SC of the upper probability for this event, is

$$\begin{aligned} SC_{\overline{P}(\mu_m \geq z)}(\underline{x}(j,\delta)) &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) \left[1\{[\overline{\mu^i_m} + \frac{s^i_j \delta}{m}] \geq z\} - 1\{\overline{\mu^i_m} \geq z\} \right] \\ &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) 1\{0 \leq \frac{m(z - \overline{\mu^i_m})}{s^i_j} \leq \delta\} \end{aligned} \quad (5.29)$$

For $\delta > 0$, the NPI-SC for the lower and upper probabilities for the event $\mu_m \geq z$, as given in Equations (5.28) and (5.29), has three different scenarios. For some orderings O_i, both $\mu^i_m(\underline{x}) \geq z$ and $\mu^i_m(\underline{x}(j,\delta)) \geq z$, for other orderings both $\mu^i_m(\underline{x}) < z$ and $\mu^i_m(\underline{x}(j,\delta)) < z$, and for some orderings $\mu^i_m(\underline{x}) \geq z$ and $\mu^i_m(\underline{x}(j,\delta)) < z$, or $\mu^i_m(\underline{x}) < z$ and $\mu^i_m(\underline{x}(j,\delta)) \geq z$. The latter orderings lead to changes in the inference for the mean. This is reflected by Equations (5.28) and (5.29), only the affected orderings are considered.

The NPI-SC of the lower and upper probabilities for the event $\mu_m \in (z_1, z_2))$, are

$$\begin{aligned} &SC_{\underline{P}(\mu_m \in (z_1,z_2))}(\underline{x}(j,\delta)) = \underline{P}_{\underline{x}(j,\delta)}(\mu_m \in (z_1,z_2)) - \underline{P}_{\underline{x}}(\mu_m \in (z_1,z_2)) \\ &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) \Big[1\{z_1 \leq \underline{\mu^i_m}(\underline{x}(j,\delta)) \leq \overline{\mu^i_m}(\underline{x}(j,\delta)) \leq z_2\} \\ &\quad -1\{z_1 \leq \underline{\mu^i_m}(\underline{x}) \leq \overline{\mu^i_m}(\underline{x}) \leq z_2\} \Big] \end{aligned}$$

and

$$\begin{aligned} &SC_{\overline{P}(\mu_m \in (z_1,z_2))}(\underline{x}(j,\delta)) = \overline{P}_{\underline{x}(j,\delta)}(\mu_m \in (z_1,z_2)) - \overline{P}_{\underline{x}}(\mu_m \in (z_1,z_2)) \\ &= \sum_{i=1}^{\binom{n+m}{n}} P(O_i) \Big[1\{(\underline{\mu^i_m}(\underline{x}(j,\delta)), \overline{\mu^i_m}(\underline{x}(j,\delta))) \cap (z_1,z_2) \neq \emptyset\} \\ &\quad -1\{(\underline{\mu^i_m}(\underline{x}), \overline{\mu^i_m}(\underline{x})) \cap (z_1,z_2) \neq \emptyset\} \Big] \end{aligned} \quad (5.30)$$

These NPI-SC will be illustrated in Example 5.3 in Subsection 5.5.3.

The c-breakdown points of the lower and upper bounds of μ^i_m, are

$$\begin{aligned} \lambda^*_c(\lambda_a, \underline{\mu^i_m}, \underline{x}(\delta, j_1, \ldots, j_l)) &= \frac{1}{n} \quad \text{if} \quad s_{l+1} \neq 0 \\ \lambda^*_c(\lambda_a, \overline{\mu^i_m}, \underline{x}(\delta, j_1, \ldots, j_l)) &= \frac{1}{n} \quad \text{for} \quad s^i_l \neq 0 \quad \text{and} \quad k = 0, .., l \end{aligned}$$

because if we hold $x_1, \ldots, x_{n-1}$ fixed and let x_n go to infinity then μ_m^i also goes to infinity if $s_{l+1}^i \neq 0$ or $s_l^i \neq 0$, corresponding to $\underline{\mu_m^i}$ and $\overline{\mu_m^i}$. However, when we consider inference involving the mean, we will not let x_n go to infinity, as we have bounds for the data observations $L < x_1 < \ldots < x_n < R$, so will let x_n go large but within these bounds. So $\lambda_c^*(\lambda_a, \underline{\mu_m^i}, \underline{x}(\delta, j_1, \ldots, j_l))$ may not be equal to $\frac{1}{n}$. This will be illustrated in Example 5.3 in Subsection 5.5.3.

5.5.3 Comparison of robustness of the median and the mean of the future observations

A main topic in the classical theory of robustness is comparison of the robustness of the mean and the median. The mean is typically very sensitive to small changes in the data whereas the median is more robust.

In our case the inferences that involve the median of the m future observations depend on the event of interest, for example, the lower and upper probabilities for the event $M_m > z$ might slightly be affected if the contaminant changes the number of observations that are less than z, and its effect is a step function, as will be illustrated in Example 5.3. The c-breakdown point for $M_m > z$, where $z \in (x_{k-1}, x_k)$, is $\frac{n-k+2}{n}$, so the value of NPI-BP for the median decreases as the value of k increases.

If we replace x_j by $\tilde{x}_l$, then the inferences of events involving the mean of the m future observations might be affected by a small change in the data, if s_l^i the number of future observations in I_l given the ordering O_i is not equal to zero. Example 5.3 illustrates the NPI-SC and NPI-BP for inferences involving the mean and the median of the m future observations.

Example 5.3. To illustrate the NPI-SC for different inferences involving the median and mean of the $m = 3$ future observations, we consider the data set $\underline{x} = \{-9, -7, 0, 2, 5, 7, 10, 16\}$, so $n = 8$, and the contaminated sample $\underline{x}(2, \delta)$, where we add δ to $x_2 = -7$ and $\delta \in \mathbb{R}$. When we consider the mean of the 3 future observations, we set $x_0 = -17$ and $x_9 = 18$ as bounds for the observations.

Figure 5.9 shows the NPI-SC for the NPI lower and upper probabilities for the events $\mu_3 \geq 1$, $\mu_3 \in (1, 9)$, $M_3 \geq 1$ and $M_3 \in (1, 9)$ given $\underline{x}$, and the contaminated sample $\underline{x}(2, \delta)$. Note that the NPI lower probability for such an event of interest in these figures is denoted by LP represented by line, and the NPI upper probability by UP and represented by dotted line. The NPI-SC for $\mu_3 \geq 1$ increases as the value of $-7 + \delta$ increases, and the maximum NPI-SC for the lower and upper probabilities for $\mu_3 \geq 1$ are 0.1576 and 0.1333 respectively, which occur at $-7+\delta = 16$ which is the largest contaminate value, as δ can not go to 25 as we set $R = 18$ as upper bound for the observations. The inferences involving the median of the $m = 3$ future observations depend on the ranks of the observations, which are only affected if the number of the observations that are greater than 1, or in $(1, 9)$, changes, so NPI-SC is a step function. The NPI-SC for the NPI lower and upper probabilities for $M_3 \geq 1$ are 0.1454 and 0.1273 respectively, which occur at $\delta > 8$. So it is less than NPI-SC for $\mu_3 \geq 1$. The NPI-SC for the event $\mu_3 \in (1, 9)$ increases till $\delta \geq 12.3$ then for $\delta > 12.3$ it decreases to be close to zero. The maximum NPI-SC for the lower and upper probabilities for $\mu_3 \in (1, 9)$ are 0.0667 and 0.0909 respectively, and it occurred at $\delta = 10.8$. The maximum NPI-SC for the NPI lower and upper probabilities for $M_3 \in (1, 9)$ are 0.1454 and 0.1273 respectively, so it is greater than NPI-SC for $\mu_3 \in (1, 9)$. Table 5.7 presents that for $\delta < 7$ and $\delta > 19$, the inference involving the mean is more sensitive than the inference involving the median. In contrast for $8 < \delta \leq 15.3$ the inferences involving the mean are more robust.

Figure 5.10 and Table 5.8 present the NPI-SC for the NPI lower and upper probabilities for the events that $\mu_3 \geq 1$, $\mu_3 \in (1, 9)$, $M_3 \geq 1$ and $M_3 \in (1, 9)$, as function of an additional value y added to the data $\underline{x}$. The results illustrate that both inferences are sensitive to any additional value where the NPI-SC for the inference involving the mean increases as function of y, whereas the NPI-SC for the inference involving the median is a step function. The maximum NPI-SC for the NPI lower and upper probabilities for the event $\mu_3 \geq 1$ are 0.0909 and 0.0545, which occur at $y = 17.39$, and for $M_3 \geq 1$ are 0.0606 and 0.0424, which occur at

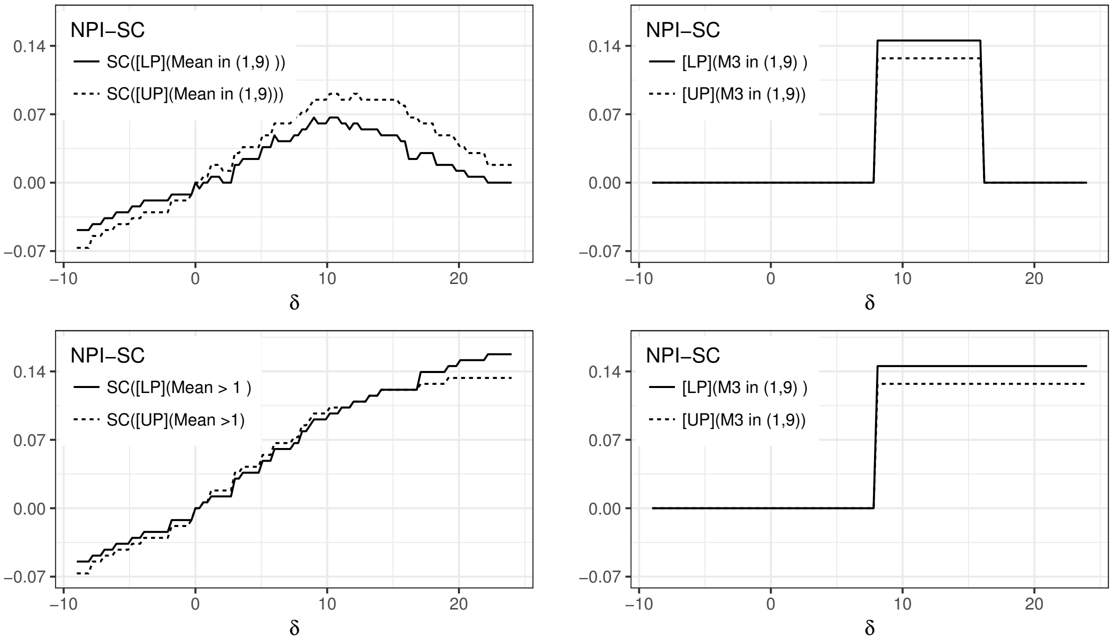

Figure 5.9: $SC_I(\underline{x}(2,\delta))$ for the events $\mu_3 \geq 1$, $\mu_3 \in (1,9)$, $M_3 \geq 1$ and $M_3 \in (1,9)$

	$\mu_3 \geq 1$		$M_3 \geq 1$		$\mu_3 \in (1,9)$		$M_3 \in (1,9)$	
δ	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
-9	-0.0545	-0.0667	0	0	-0.0485	-0.0667	0	0
-7.41	-0.0485	-0.0545	0	0	-0.0424	-0.0545	0	0
-5.82	-0.0364	-0.0424	0	0	-0.0303	-0.0424	0	0
-4.23	-0.0303	-0.0364	0	0	-0.0242	-0.0364	0	0
-2.64	-0.0242	-0.0303	0	0	-0.0182	-0.0303	0	0
-1.05	-0.0121	-0.0182	0	0	-0.0121	-0.0182	0	0
0.54	0.0061	0.0061	0	0	0.0000	0.0061	0	0
2.13	0.0121	0.0182	0	0	0.0000	0.0121	0	0
3.72	0.0364	0.0424	0	0	0.0242	0.0364	0	0
5.31	0.0485	0.0545	0	0	0.0364	0.0485	0	0
6.9	0.0606	0.0667	0	0	0.0424	0.0606	0	0
8.49	0.0788	0.0848	0.1455	0.1273	0.0545	0.0727	0.1455	0.1273
10.08	0.0970	0.1030	0.1455	0.1273	0.0667	0.0909	0.1455	0.1273
11.67	0.1030	0.1030	0.1455	0.1273	0.0545	0.0848	0.1455	0.1273
12.9	0.1091	0.1091	0.1455	0.1273	0.0545	0.0848	0.1455	0.1273
13.26	0.1152	0.1152	0.1455	0.1273	0.0545	0.0848	0.1455	0.1273
14.85	0.1212	0.1212	0.1455	0.1273	0.0485	0.0848	0.1455	0.1273
15.3	0.1212	0.1212	0.1455	0.1273	0.0485	0.0848	0.1455	0.1273
16.44	0.1212	0.1212	0.1455	0.1273	0.0242	0.0667	0.0000	-0.0000
18.03	0.1394	0.1273	0.1455	0.1273	0.0182	0.0485	0.0000	-0.0000
19.62	0.1455	0.1333	0.1455	0.1273	0.0121	0.0424	0.0000	-0.0000
21.21	0.1515	0.1333	0.1455	0.1273	0.0061	0.0303	0.0000	-0.0000
22.8	0.1576	0.1333	0.1455	0.1273	0.0000	0.0182	0.0000	-0.0000
24.39	0.1576	0.1333	0.1455	0.1273	-0.0121	0.0061	0.0000	-0.0000

Table 5.7: $SC_I(\underline{x}(2,\delta))$ for $m = 3$

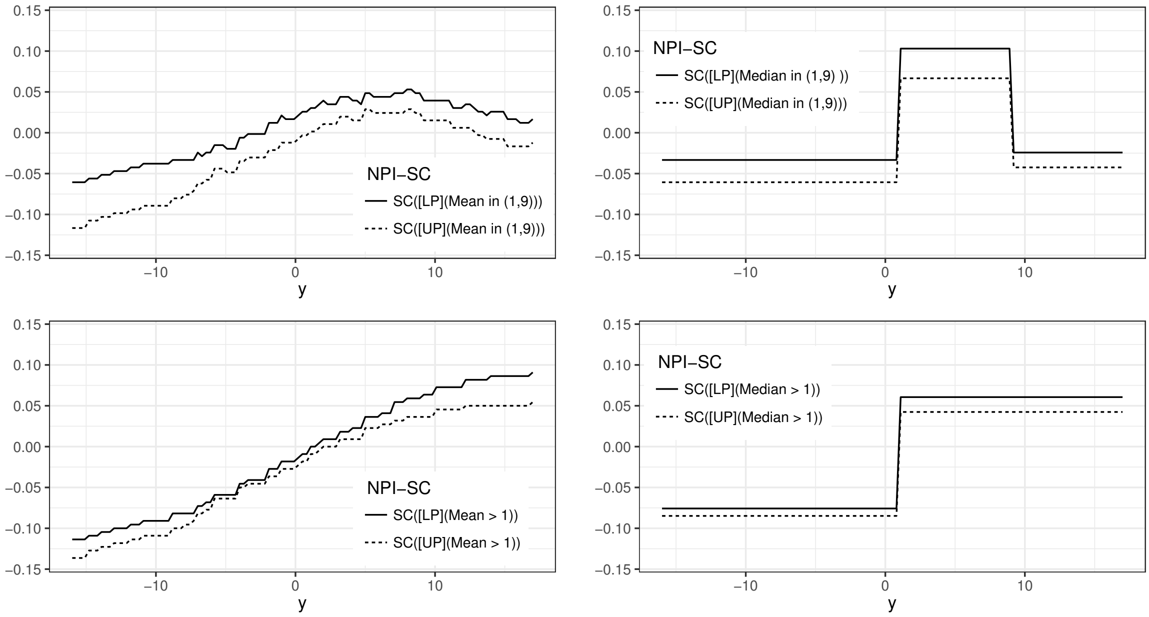

Figure 5.10: $SC_I(\underline{x}, y)$ for the events $\mu_3 \geq 1$, $\mu_3 \in (1, 9)$, $M_3 \geq 1$ and $M_3 \in (1, 9)$

$y > 1$. The maximum NPI-SC for the NPI lower and upper probabilities for the event $\mu_3 \in (1, 9)$ are 0.0530 and 0.0288, which occurs at $y = 8$. For $y \leq -8.05$ the effects on the lower and upper probabilities for the events $\mu_3 \geq 1$ and $\mu_3 \in (1, 9)$ are greater than the effects on the lower and upper probabilities for the events $M_3 \geq 1$ and $M_3 \in (1, 9)$. Table shows that for $y = -6.46, \ldots, 8.3$ the NPI-SC for $\mu_3 \geq 1$ is less than the NPI-SC for $M_3 \geq 1$. From adding $y \leq -8.05$ and $y > 9$ the inferences for $\mu_3 \geq 1$ and $\mu_3 \in (1, 9)$ are more sensitive than $M_3 \geq 1$ and $M_3 \in (1, 9)$.

To illustrate the c-breakdown point, we consider NPI-SC as function of the number of contaminants present in the data, starting by replacing x_8 by $x_8 + 1$, then x_8 and x_7 by $x_8 + 1$ and $x_7 + 5$, and so on, until all observations have been contaminated. Figure 5.11 shows $SC_{[\underline{P},\overline{P}](\mu_3 \geq 1)}(\underline{x}(j_1, \ldots, j_l, \delta))$ and $SC_{[\underline{P},\overline{P}](M_3 \geq 1)}(\underline{x}(j_1, \ldots, j_l, \delta))$ as functions of the number of the observations that have been contaminated by adding different value of δ to them. The results clearly show that when we contaminate up to 5 observations, which are $2, 5, 7, 10, 16$ in the data, to become $11.5, 12, 12.5, 13, 17$, the inference involving the median $X_{(2)} \geq 1$ is not affected at all whereas the inference involving the mean of the future ob-

	$\mu_3 \geq 1$		$M_3 \geq 1$		$\mu_3 \in (1,9)$		$M_3 \in (1,9)$	
y	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
-16	-0.1136	-0.1364	-0.0758	-0.0848	-0.0606	-0.1167	-0.0333	-0.0606
-14.41	-0.1091	-0.1273	-0.0758	-0.0848	-0.0561	-0.1076	-0.0333	-0.0606
-12.82	-0.1000	-0.1182	-0.0758	-0.0848	-0.0470	-0.0985	-0.0333	-0.0606
-11.23	-0.0955	-0.1136	-0.0758	-0.0848	-0.0424	-0.0939	-0.0333	-0.0606
-9.64	-0.0909	-0.1091	-0.0758	-0.0848	-0.0379	-0.0894	-0.0333	-0.0606
-8.05	-0.0818	-0.1000	-0.0758	-0.0848	-0.0333	-0.0803	-0.0333	-0.0606
-6.46	-0.0682	-0.0773	-0.0758	-0.0848	-0.0242	-0.0576	-0.0333	-0.0606
-4.87	-0.0591	-0.0636	-0.0758	-0.0848	-0.0197	-0.0485	-0.0333	-0.0606
-3.28	-0.0409	-0.0455	-0.0758	-0.0848	-0.0015	-0.0303	-0.0333	-0.0606
-1.69	-0.0273	-0.0364	-0.0758	-0.0848	0.0121	-0.0212	-0.0333	-0.0606
-0.1	-0.0182	-0.0273	-0.0758	-0.0848	0.0167	-0.0121	-0.0333	-0.0606
1.49	0.0000	-0.0091	0.0606	0.0424	0.0303	0.0015	0.1030	0.0667
3.08	0.0182	0.0091	0.0606	0.0424	0.0439	0.0197	0.1030	0.0667
4.67	0.0227	0.0091	0.0606	0.0424	0.0348	0.0152	0.1030	0.0667
5.9	0.0364	0.0227	0.0606	0.0424	0.0439	0.0242	0.1030	0.0667
6.26	0.0409	0.0273	0.0606	0.0424	0.0439	0.0242	0.1030	0.0667
7.85	0.0545	0.0318	0.0606	0.0424	0.0485	0.0242	0.1030	0.0667
8.3	0.0591	0.0364	0.0606	0.0424	0.0530	0.0288	0.1030	0.0667
9.44	0.0636	0.0364	0.0606	0.0424	0.0394	0.0152	-0.0242	-0.0424
11.03	0.0727	0.0455	0.0606	0.0424	0.0303	0.0061	-0.0242	-0.0424
12.62	0.0818	0.0500	0.0606	0.0424	0.0303	0.0015	-0.0242	-0.0424
14.21	0.0864	0.0500	0.0606	0.0424	0.0258	-0.0076	-0.0242	-0.0424
15.8	0.0864	0.0500	0.0606	0.0424	0.0167	-0.0167	-0.0242	-0.0424
17.39	0.0909	0.0545	0.0606	0.0424	0.0030	-0.0258	-0.0242	-0.0424

Table 5.8: $SC_I(\underline{x}, y)$ for $m = 3$

servations is affected. If we choose $c = 0.15$, then the c-breakdown points for the lower and upper probabilities for $M_3 \geq 1$ and for the upper probability for $\mu_3 \geq 1$, are all equal to 0.875, so breakdown occurs when we change 7 observations out of 8, whereas the c- breakdown point for the NPI lower probability for $\mu_3 \geq 1$ is 0.625, so breakdown occurs if 5 out of 8 observations are contaminated.

Figure 5.12 illustrates the NPI-SC for the events $\mu_3 \geq 1$ and $M_3 \geq 1$ as function of the number of contaminants added to the data. The contaminant values that have been added to the data sequentially are $Y = \{16.2, 16.3, 16.4, 16.5, 16.6, 16.9, 17, 17.9\}$. The results show that both inferences involving the mean and the median are affected by any number of contaminates that have been added to the data. Again if we choose $c = 0.15$, the smallest fraction of contaminations added to the sample that cause $SC_I(\underline{x}, y_1, \ldots, y_l)$ to be greater than 0.15, are 4/12 and 6/14 for the NPI lower and upper probabilities for $M_3 \geq 11$, and 3/11 and 5/13 for $\mu_3 \geq 1$. Thus the 0.15-breakdown point for this inference involving

	$M_3 > 1$		$\mu_3 > 1$	
	$\lambda^*_{0.15}(\underline{P})$	$\lambda^*_{0.15}(\underline{P})$	$\lambda^*_{0.15}(\underline{P})$	$\lambda^*_{0.15}(\underline{P})$
λ_a	0.875	0.875	0.625	0.875
λ_b	0.3333	0.4286	0.2727	0.3846

Table 5.9: c-breakdown point for $c = 0.15$.

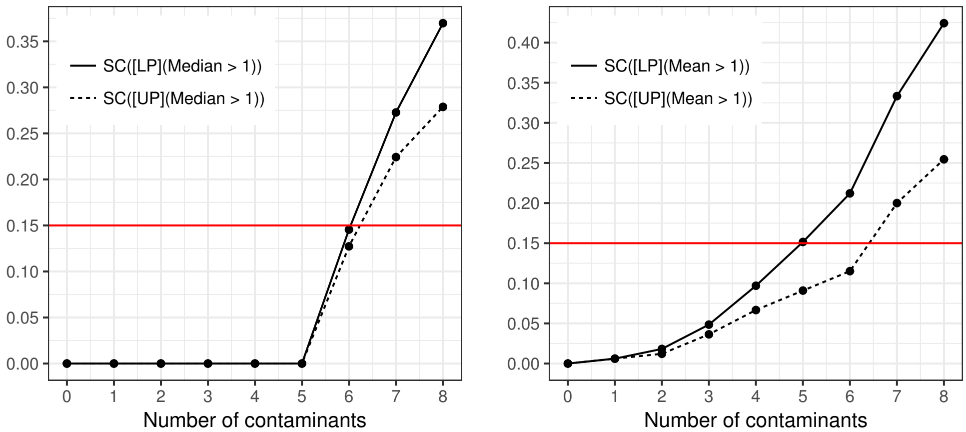

Figure 5.11: $SC_I(\underline{x}(j_1, \ldots, j_l, \delta))$ for the events $\mu_3 \geq 1$ and $M_3 \geq 1$

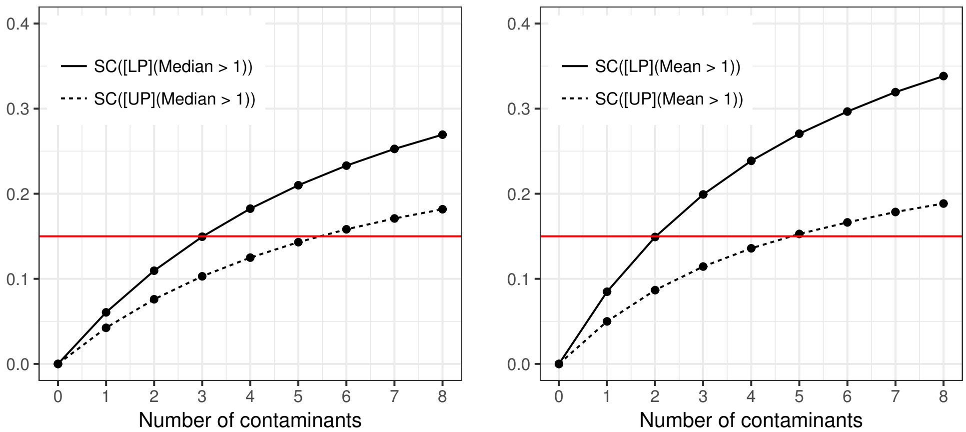

Figure 5.12: $SC_I(\underline{x}, y_1, \ldots, y_l)$ for the events $\mu_3 \geq 1$ and $M_3 \geq 1$

the median is higher than for this inference involving the mean, as illustrated in Table 5.9. Comparing Figures 5.11 and 5.12 we can see that contaminating the data by replacing observations by contaminated values tends to be less affecting than by adding contamination, as the latter affects the sample size which plays an important role in our inferences.

$\Diamond$

5.6 Robustness of other inferences

In this section we consider the use of the presented tools for robustness, namely NPI-SC and NPI-BP, for pairwise comparisons and for reproducibility of tests, as presented in Sections 3.2, 4.3 and 4.4.

5.6.1 Robustness of pairwise comparisons

We investigate the robustness of one of the applications of NPI for future order statistics for statistical inference problems, as presented in Section 3.2. The NPI-SC of the lower and upper probabilities for the event that $X_{(r)} < Y_{(r)}$, if we replace y_j by $y_j + \delta$, which we denote by $\tilde{y}_l$, are

$$SC_{\underline{P}(X_{(r)}<Y_{(r)})}(\underline{y}(j,\delta))$$
$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ P(Y^f_{(r)} \in I^y_{l+1}) \times P(X_{(r)} \in I^x_d) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ P(Y^f_{(r)} \in I^y_{l+1}) \times [P(X_{(r)} \in I^x_d)P(X_{(r)} \in I^x_{d+1})] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

$$SC_{\overline{P}(X_{(r)}<Y_{(r)})}(\underline{y}(j,\delta))$$
$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ P(Y_{(r)} \in I^y_{l}) \times P(X_{(r)} \in I^x_{d+1}) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ P(Y^f_{(r)} \in I^y_{l}) \times [P(X_{(r)} \in I^x_{d+1})P(X_{(r)} \in I^x_{d+2})] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

The NPI pairwise comparisons for such an event are not sensitive to a small change in the data, as they only are affected if the change to an observation has changed the order of the X and Y observations. In the next example we will illustrate the NPI-SC and NPI-BP for such NPI pairwise comparisons.

Example 5.4. To illustrate the NPI-SC for pairwise comparisons, we consider the data set of a study of the effect of ozone environment on rats growth, as given in Example 3.4. We use this dataset to illustrate the effect of replacing $x_2 = -14.7$ by $-14.7 + \delta$, for δ from -50 to 100, on the pairwise comparisons based on the events $X_{(r)} < Y_{(r)}$, $r = 1, \ldots, m$, and $m = 5$.

Figure 5.13 illustrates what happens to the NPI lower and upper probabilities for the event $X_{(r)} < Y_{(r)}$, if observation $x_2 = -14.7$ in the X sample is replaced by $-14.7 + \delta$. Increasing the value -14.7 to $-14.7 + \delta$ leads to decreasing $SC_{P(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta))$ for δ such that the rank of this observation among the Y group changes. However, if the contaminated value $-14.7 + \delta$ does not change its rank among Y observations then $SC_{\underline{P}(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta)) = 0$ and $SC_{\overline{P}(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta)) = 0$. For $\delta \leq -30$ the NPI-SC for $X_{(1)} < Y_{(1)}$ has large effect where the other NPI-SC for the other inferences, for $r = 2, \ldots, 5$, are close to zero. For $-1.5 \leq \delta \leq 27$ the $SC_{\underline{P}(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta)) = 0$ and $SC_{\overline{P}(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta)) = 0$ for all r, as the value $-14.7 + \delta$ does not change its rank among Y observations. For $\delta > 27$, the effect of the contaminated value $-14.7 + \delta$ increases as the value of r increase. The inferences involving $r = 4$ and 5 have large NPI-SC when the value $x_2 + \delta$ exceeds all the Y observations.

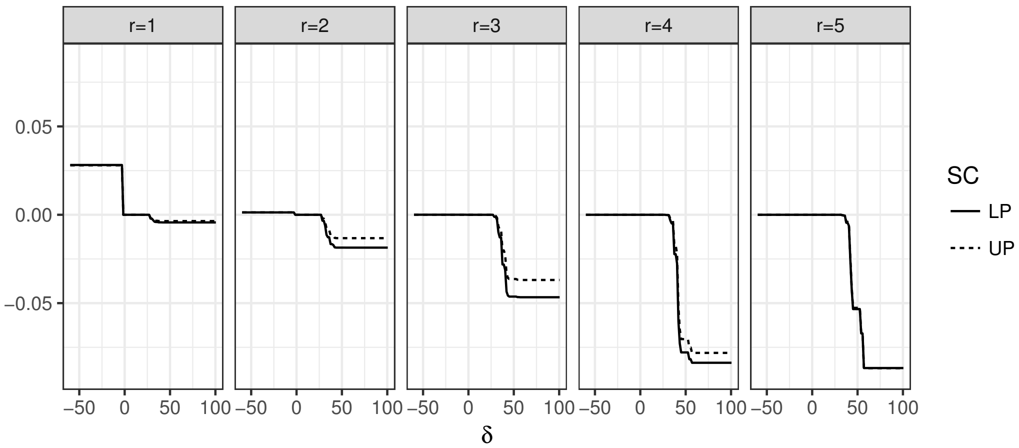

Figure 5.13: $SC_{P(X_{(r)}<Y_{(r)})}(\underline{x}(2, \delta))$ for $m = 5$

To illustrate the c-breakdown point of these NPI pairwise comparisons, we consider NPI-SC for $X_{(r)} < Y_{(r)}$ for $m = 3$ and $r = 1, 2, 3$, for the case of adding the value 100 to l observations in group X or group Y. This is shown in Figures 5.14 and 5.15 and Tables 5.10 and 5.11. Figure 5.14 illustrates that the absolute value of the NPI-SC increases as the value of l, the number of contaminations in the X sample, increases. If we choose $c = 0.05$, then the NPI-BP for $r = 1$ is $10/22$, for $r = 2$ it is $6/22$ and for $r = 3$ it is $5/22$, so as the value of r increases the NPI-BP decreases. Thus the probability for the event $X_{(r)} < Y_{(r)}$ based on the given data is more robust if we consider $r = 1$, as it has the highest 0.05-breakdown point.

Figure 5.15 illustrates the NPI-SC for $X_{(r)} < Y_{(r)}$, for $m = 3$ and $r = 1, 2, 3$, if we replace l observations from the Y sample by $y_{24-l} + 100, \ldots, y_{23} + 100$. For $c = 0.05$, the c-breakdown point for the lower and upper probabilities for the event $X_{(1)} < Y_{(1)}$ is 1. This is because the inference $X_{(1)} < Y_{(1)}$ is most affected in case of contamination of all y values such that these exceed all X observations. For $r = 2$ the NPI-BP is $\lambda^*_{0.05}(\underline{P}(X_{(2)} < Y_{(2)}), \underline{x}(\delta, 13, \ldots, 23)) = 11/23$ and $\lambda^*_{0.05}(\overline{P}(X_{(2)} < Y_{(2)}), \underline{x}(\delta, 12, \ldots, 23)) = 13/23$. However, for $r = 3$, $\lambda^*_{0.05}([\underline{P}, \overline{P}](X_{(3)} < Y_{(3)}), \underline{x}(\delta, 22, 23)) = 2/23$, so if we let y_{22} and y_{23} exceed all the X observations, the inference breaks down.

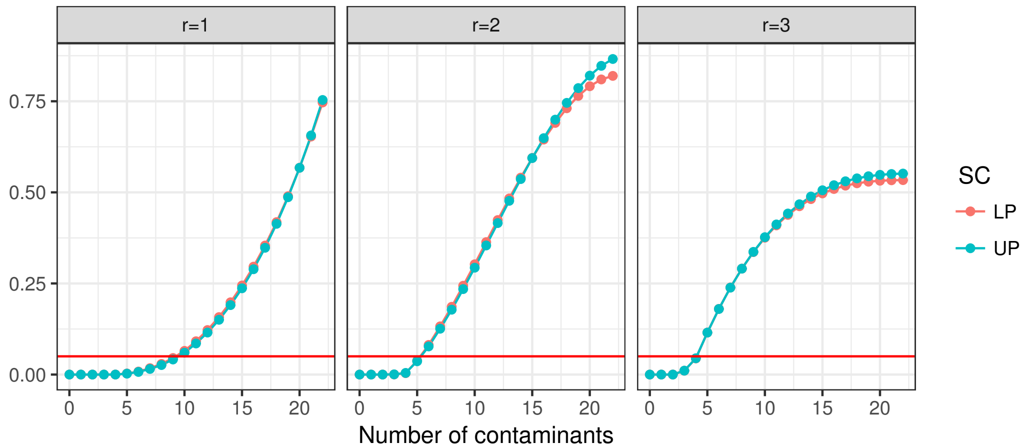

Figure 5.14: $|SC_{P(X_{(r)}<Y_{(r)})}(\underline{x}(23-l, \ldots, 22, 100))|$ for $m = 3$

$r = 1$			$r = 2$			$r = 3$		
l	$SC_{\underline{P}}$	$SC_{\overline{P}}$	l	$SC_{\underline{P}}$	$SC_{\overline{P}}$	l	$SC_{\underline{P}}$	$SC_{\overline{P}}$
9	0.0454	0.0413	5	0.0382	0.0365	4	0.0442	0.0447
10	0.0649	0.0599	6	0.0813	0.0772	5	0.1151	0.1151

Table 5.10: The absolute value of $SC_{P(X_{(r)}<Y_{(r)})}(\underline{x}(j_1, \ldots, j_l, 100))$ for $m = 3$ and $n = 22$.

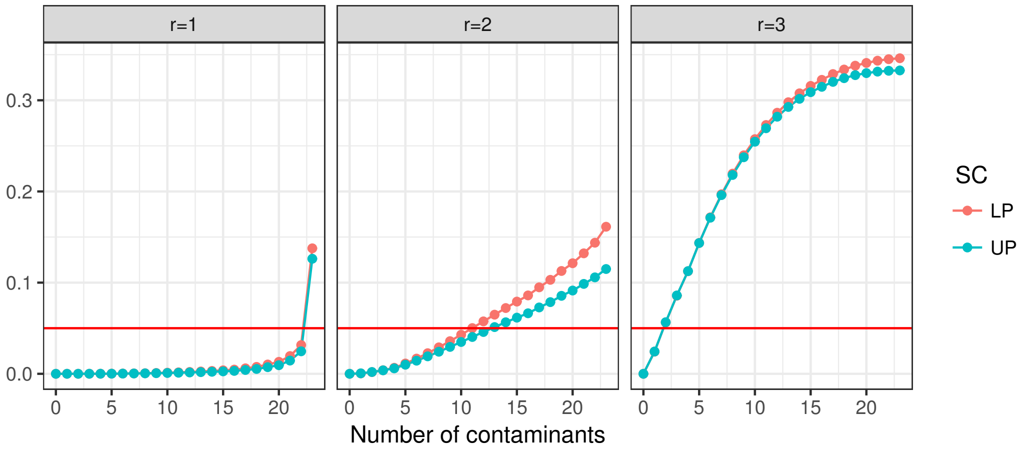

Figure 5.15: $SC_{P(X_{(r)}<Y_{(r)})}(\underline{y}(24-l,\ldots,23,100))$ for $m=3$

$r=1$			$r=2$				$r=3$		
l	$SC_{\underline{P}}$	$SC_{\overline{P}}$	l	$SC_{\underline{P}}$	l	$SC_{\overline{P}}$	l	$SC_{\underline{P}}$	$SC_{\overline{P}}$
22	0.0315	0.0247	10	0.0429	12	0.0458	1	0.0243	0.0244
23	0.1376	0.1262	11	0.0501	13	0.0513	2	0.0563	0.0566

Table 5.11: $SC_{P(X_{(r)}<Y_{(r)})}(\underline{y}(j_1,\ldots,j_l,100))$ for $m=3$ and $n=23$.

$\Diamond$

5.6.2 Robustness of NPI reproducibility probability of two statistical tests

Based on the given data in the original test and the result of the actual hypothesis test, that is whether the null hypothesis is rejected or not, NPI can be applied to study the reproducibility of the test, assuming that the sample size of the actual test and future test are the same $n=m$. This seems to be a natural assumption in order to reflect reproducibility.

To study the robustness of NPI reproducibility of classical statistical tests, we will only consider one way of contaminating the data which is by replacing one of the observations by a small contaminant. We do not consider contamination by adding a value to the data as this would make a substantial change to the test statistic and could require a different threshold value, which would complicate the study.

We consider how a small change in the original data would influence the NPI reproducibility probability of the test outcome given the original data. But we have to emphasize the assumption that we only consider a small change to the data which would not lead to a different result of the underlying test. So to keep everything simple we assume that we only consider a small change which does not change the outcome of the original test.

Most of the literature [54, 68] considers the robustness of the test result, so that if a test is robust then small variations in the data should not be able to reverse the test decision. In our study, we are interested in exploring the robustness of the NPI reproducibility probability of the test conclusion, not the robustness of the original test result. Thus, we will not consider the case where adding δ to one of the observations could change the original test decision from rejecting to not rejecting the null hypotheses, or the other way around.

First we consider the robustness of the reproducibility of the one-sided quantile test of $H_0 : \kappa_p = \kappa_p^0$ versus $H_1 : \kappa_p > \kappa_p^0$. The original test leads to rejection of H_0 if and only if $k \leq r - 1$, where k is the number of observations in the original sample $\underline{x}$ of size n that are less than κ_p^0. Reproducibility of this test result is therefore the event that, if the test were repeated, also with n observations, then that would also lead to rejection of H_0. Let $\kappa_p^0 \in I_t = (x_{t-1}, x_t)$, then the effect of adding δ to any of the data observations, say x_j which becomes $\tilde{x}_l$, on the reproducibility of the quantile test for that event is

$$SC_{\underline{P}(X_{(r)}>\kappa_p^0|k)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ P(X_{(r)} \in I_t) & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \end{cases}$$

$$SC_{\overline{P}(X_{(r)}>\kappa_p^0|k)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ P(X_{(r)} \in I_{t-1}) & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \end{cases}$$

If the original test led to not reject H_0, so if $k \geq r$, then reproducibility of the test is the event that H_0 would also not get rejected in the future test. The NPI-SC for the NPI lower and upper reproducibility probabilities for $X_{(r)} \leq \kappa_p^0$

are

$$SC_{\underline{P}(X_{(r)}<\kappa_p^0|k)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad x_j > \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \\ 0 & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ P(X_{(r)} \in I_t) & \text{if} \quad x_j > \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ -P(X_{(r)} \in I_{t-1}) & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \end{cases}$$

The NPI-SC for the NPI upper probability

$$SC_{\overline{P}(X_{(r)}<\kappa_p^0|k)}(\underline{x}(j,\delta)) = \begin{cases} 0 & \text{if} \quad x_j > \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \\ 0 & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ P(X_{(r)} \in I_{t+1}) & \text{if} \quad x_j > \kappa_p^0 \quad \text{and} \quad \tilde{x}_l < \kappa_p^0 \\ -P(X_{(r)} \in I_t) & \text{if} \quad x_j < \kappa_p^0 \quad \text{and} \quad \tilde{x}_l > \kappa_p^0 \end{cases}$$

So the NPI-RP for the quantile test is only affected if the change in the data changes the value of k, which is the number of observations less than κ_p^0.

Example 5.5. We consider the same case as presented in Example 4.1, to illustrate the NPI-SC for the NPI-RP for the quantile test considering the third quantile, with sample size of 15, and 5% significance level. Let $\tilde{k}$ denote the number of observations that are less than κ_p^0 based on the contaminated sample $\underline{x}(j,\delta)$.

Table 5.12 presents, in the first column, the NPI-SC for the NPI-RP for the event that the future test would also reject H_0 if $X_{(8)} \geq \kappa_{0.75}^0$ given all possible value of k in the original test. This NPI-RP for this event is only affected if k, the number of observations less than $\kappa_{0.75}^0$, changes, otherwise $SC_{[\underline{RP}(k),\overline{RP}(k)]}(\underline{x}(j,\delta)) = 0$. The size of the effect for such an inference increases as the value of k increases.

Table 5.12 presents, in the second column, the NPI-SC for the test reproducibility if the original test did not reveal a significance affect, which is the event that the future test would also lead to not reject H_0, if $X_{(8)} < k_{0.75}^0$. The RP for $X_{(8)} < k_{0.75}^0$ is only affected if $x_j < \kappa_{0.75}^0$ becomes $x_j + \delta > \kappa_{0.75}^0$. The NPI-SC for such an inference decreases as the value of k increase.

$\diamondsuit$

		$P(X_{(8)} \geq \kappa^0_{0.75}\|k)$				$P(X_{(8)} \leq \kappa^0_{0.75}\|k)$	
k	$\tilde{k}$	$SC_{\underline{RP}}$	$SC_{\overline{RP}}$	k	$\tilde{k}$	$SC_{\underline{RP}}$	$SC_{\overline{RP}}$
0	1	-0.00600	-0.00110	9	8	-0.14238	-0.12656
1	2	-0.01799	-0.00600	10	9	-0.12656	-0.09930
2	1	0.01799	0.00600	11	10	-0.09930	-0.06770
3	2	0.03898	0.01799	12	11	-0.06770	-0.03898
4	3	0.06770	0.03898	13	12	-0.03898	-0.01799
5	4	0.09930	0.06770	14	13	-0.01799	-0.00600
6	5	0.12656	0.09930	15	14	-0.00600	-0.00110
7	6	0.14238	0.12656				

Table 5.12: $SC_{RP(k)}(\underline{x}(j,\delta))$ for $n = 15$.

We next consider NPI-SC for the NPI-RP of the precedence test presented in Section 4.4. As the NPI-RP inferences for the precedence test depend monotonically on the combined ordering of the original test data, so the local change to the combined ordering of the data of the two populations in the original test leads to change both the NPI lower and upper probabilities for the event of interest.

First we will consider the RP for the case that H_0 is rejected in the original test, so $x_k < y_r$, then $\underline{RP} = \underline{P}(X_{(k)} < Y_{(r)})$ and $\overline{RP} = \overline{P}(X_{(k)} < Y_{(r)})$. The effects of adding δ to one of the observations in group Y, say y_j which becomes $y_j + \delta = \tilde{y}_l$, on $\underline{RP}$ and $\overline{RP}$ are

$$SC_{\underline{P}(X_{(k)}<Y_{(r)})}(\underline{y}(j,\delta))$$

$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ P(Y^f_{(r)} \in I^y_{l+1}) \times P(X_{(k)} \in I^x_d) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ P(Y^f_{(r)} \in I^y_{l+1}) \times \left[P(X_{(k)} \in I^x_d) + P(X_{(k)} \in I^x_{d+1})\right] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and} \quad x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

$$SC_{\overline{P}(X_{(k)}<Y_{(r)})}(\underline{y}(j,\delta))$$

$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ P(Y_{(r)} \in I^y_l) \times P(X_{(k)} \in I^x_{d+1}) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ P(Y^f_{(r)} \in I^y_l) \times \left[P(X_{(k)} \in I^x_{d+1}) + P(X_{(k)} \in I^x_{d+2})\right] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and} \quad x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

If H_0 is not rejected in the original test, so $x_{(k)} > y_{(r)}$, then $\underline{RP} = \underline{P}(X_{(k)} >$

$Y_{(r)})$ and $\overline{RP} = \overline{P}(X_{(k)} > Y_{(r)})$. The effects of adding δ to y_j in group Y, so y_j becomes $\tilde{y}_l$, on $\underline{RP}$ and $\overline{RP}$ are

$$SC_{\underline{P}(X_{(k)}>Y_{(r)})}(\underline{y}(j,\delta))$$
$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ -P(Y^f_{(r)} \in I^y_l) \times P(X_{(k)} \in I^x_{d+1}) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ -P(Y^f_{(r)} \in I^y_l) \times [P(X_{(k)} \in I^x_{d+1}) + P(X_{(k)} \in I^x_{d+2})] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and} \quad x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

$$SC_{\overline{P}(X_{(k)}>Y_{(r)})}(\underline{y}(j,\delta))$$
$$= \begin{cases} 0 & \text{if} \quad y_j < x_d \quad \text{and} \quad \tilde{y}_l < x_d \\ -P(Y_{(r)} \in I^y_l) \times P(X_{(k)} \in I^x_{d}) & \text{if} \quad y_j < x_d \quad \text{and} \quad x_d < \tilde{y}_l \\ -P(Y^f_{(r)} \in I^y_{l+1}) \times [P(X_{(k)} \in I^x_{d}) + P(X_{(k)} \in I^x_{d+1})] & \text{if} \quad y_j < x_d < x_{d+1} \\ & \text{and} \quad x_d < x_{d+1} < \tilde{y}_l \end{cases}$$

In Example 5.6 we illustrate the effect of a contaminant in X sample on the NPI lower and upper reproducibility probabilities.

Example 5.6. To illustrate the NPI-SC for the NPI-RP for the precedence test as presented in Section 4.4, we consider the same data set as in Example 4.2, with one-sided alternative hypothesis $H_1 : \lambda_x < \lambda_y$. We consider both significance levels $\alpha = 0.05$ and $\alpha = 0.1$, and assume $r = 6$. The critical value is $k = 10$ for $\alpha = 0.05$ whereas for $\alpha = 0.1$ it is $k = 9$. This precedence test does not lead to rejection of H_0 at 5% significance level and it leads to rejection H_0 at 10%. As discussed after Example 4.2, the NPI lower and upper reproducibility probabilities, by using only the actual outcome without any assumption on the ordering of the right-censored observations, are $\underline{RP} = \underline{P}(X_{(10)} > Y_{(6)}) = 0.3871$ and $\overline{RP} = \overline{P}(X_{(10)} > Y_{(6)}) = 0.8669$ for $\alpha = 0.05$. While for $\alpha = 0.1$, $\underline{RP} = \underline{P}(X_{(9)} < Y_{(6)}) = 0.3029$ and $\overline{RP} = \overline{P}(X_{(9)} < Y_{(6)}) = 0.7079$. Let us now assume that we added an increasing value of δ to $x_2 = 0.64$, then we examine its effect on the NPI lower and upper reproducibility probabilities.

	$\alpha = 0.05$		$\alpha = 0.1$	
δ	$SC_{\underline{RP}}$	$SC_{\overline{RP}}$	$SC_{\underline{RP}}$	$SC_{\overline{RP}}$
< 0.176	0	0	0	0
$0.176 - 0.846$	0.00006	0.00009	-0.00044	-0.00023
$0.872 - 0.898$	0.00031	0.00031	-0.00149	-0.00120
$0.924 - 1.444$	0.00092	0.00069	-0.00337	-0.00355
1.470	0.00711	0.00457	-0.01501	-0.01748
$1.496 - 3.186$	0.01598	0.00923	-0.02897	-0.03744
> 3.83	0.06121	0.07502		

Table 5.13: $SC_{RP}(\underline{x}(2, \delta))$ for $X_{(10)} > Y_{(6)}$ and $X_{(9)} < Y_{(6)}$.

Figure 5.16 presents, in the first column, the NPI-SC for the NPI-RP for the event that $X_{(10)} > Y_{(6)}$, as a function of δ. The results clearly illustrates that NPI-SC for the NPI-RP for precedence test is a step function, so the NPI-RP is only affected if $x_2 + \delta$ changes its rank among the Y observations. If $x_2 + \delta > 3.83 = y_6$ then $x_2 + \delta$ is treated as right-censored observation in the x group, and the lower and upper reproducibility probabilities are achieved by taking the minimum and the maximum NPI lower and upper probability respectively, for reproducibility over all possible orderings for the right-censored. The maximum NPI-SC for $X_{(10)} > Y_{(6)}$ is achieved when $x_2 + \delta$ becomes very large and exceeds y_6.

Figure 5.16 presents, in the second column, the NPI-SC for the lower and upper reproducibility probabilities for the event $X_{(9)} < Y_{(6)}$, as a function of δ. Increasing the values of δ such that it affects the $x_2 + \delta$ rank among the Y observations leads to decrease of the value of the NPI-SC. We consider only a small value of δ, as if $x_2 + \delta$ exceeds y_6 that will change the original test conclusion and also the reproducibility probability.

Table 5.13 shows the NPI-SC for the NPI-RP for this precedence test. The values of NPI-SC are increasing as the value of $x_2 + \delta$ increases for $\alpha = 0.05$, and decreasing for $\alpha = 0.1$. That again illustrates the monotonicity of these inference with regard to changes in ranks of the data as discussed in Section 4.4.

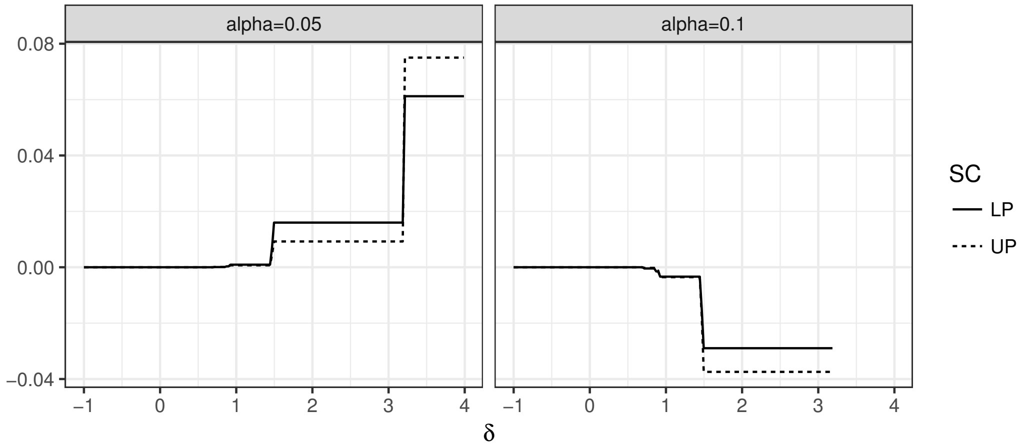

Figure 5.16: $SC_{RP}(\underline{x}(2, \delta))$ for $X_{(10)} > Y_{(6)}$ and $X_{(9)} < Y_{(6)}$

$\Diamond$

5.7 Concluding remarks

This chapter is a first step towards robustness theory for the NPI setting, and we looked at some examples involving inferences on future order statistics. We found that some of the concepts from classical statistics cannot immediately be applied, because we do not use estimators but specific inferences which are limited in value between $[0, 1]$. So, inspired by the classical concepts we have defined new concepts which are related to NPI. We then explored their use for some inferences presented in the earlier chapters of this thesis. We had an investigation of the mean and the median for the m future observations. The inference that involving the median of the m future observation is a step function, whereas the mean is affected by a single contaminate, but the size of the effect is close to the median or less in some cases. For future research it will be of interest to consider other robustness concepts and questions for NPI, and also, of course, robustness of other NPI methods.

Chapter 6

Concluding Remarks

In this thesis we have contributed to the development of NPI by considering multiple future observations for real valued random quantities. We have presented core predictive inferences of several events of interest involving order statistics of the future observations. Although some of these inferences agree with classical order statistics result [3, 36, 49], their derivation is often more straightforward than for the classical approach due to the use of the $A_{(.)}$ assumptions for prediction of the m future observations. In the classical approach, typically both the data and future observations are considered to be random quantities sampled from an unknown population, and the predictive inference is derived through conditioning on the given data observations. The NPI framework enables us to use lower and upper probabilities which extend the range of possible inferences compared to the classical method. We have also showed how pairwise and multiple comparisons can be based on such future order statistics.

We have also developed NPI for the reproducibility probability (RP) of statistical tests by considering two basic tests based on order statistics, namely a quantile test and a precedence test. Some tests may have quite poor RP, the minimum NPI lower RP is 0.5 for the quantile test and 0.25 for the precedence test. The precedence test scenarios studied in this thesis require careful attention as it typically has right-censored observations. Therefore, we have introduced two points of view for NPI-RP for such a precedence test. First we have assumed that all the data were available in the original test, considering all possible orderings of the right-censored observations. Secondly, we defined the lower and upper RP as the minimum and maximum over all RP values for all possible orderings of the right-censored data. Obviously, this perspective leads to high imprecision but it is convenient as there are no assumptions on the right-censored data. The NPI approach to reproducibility of tests can be developed for many statistical test, including more sophisticated precedence tests as presented by Balakrishnan and Ng [10].

A first exploration of robustness of NPI has been presented in this thesis. We have introduced some well-known concepts of robust tools, namely sensitivity curve and breakdown point, to the NPI setting. We have modified these two concepts to suit our context of NPI inferences but we have kept similar interpretations. We have introduced the NPI- sensitivity curve and c- breakdown point for evaluating the robustness of the predictive inferences. For future research it will be of interest to consider other robustness concepts and questions for NPI, and also, of course, robustness of other NPI methods.

Bibliography

[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] Ahmadi, J., and Balakrishnan, N. (2012). Outer and inner prediction intervals for order statistics intervals based on current records. *Statistical Papers*, **53**, 789-802.

[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] Atmanspacher, H. and Maasen, S. (eds) (2016). *Reproducibility: Principles, Problems, Practices and Prospects*. Wiley, Hoboken, New Jersey.

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

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

[9] Baker, R. (2010) *Multinomial Nonparametric Predictive Inference: Selection, Classification and Subcategory Data*. PhD Thesis, University of California. USA.

[10] Balakrishnan, N. and Ng, H.K.T. (2006). *Precedence-Type Tests and Applications*. Wiley, Hoboken, New Jersey.

[11] Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Wiley, Chichester.

[12] Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics*, 16–39.

[13] Bechhofer, R.E., Santner, T.J. and Goldsman, D.M. (1995). *Design and Analysis of Experiment for Statistical Selection, Screening, and Multiple Comparisons*. Wiley, Chichester.

[14] Becker, W.A. (1961). Comparing entries in random sample tests. *Poultry Science*, **40**(6), 1507-1514.

[15] Bellio, R. and Ventura, L. (2005). An introduction to robust estimation with R functions. *Proceedings of 1st International Work*, 1-57.

[16] Bin Himd, S. (2014) *Nonparametric Predictive Methods for Bootstrap and Test Reproducibility*. PhD Thesis, Durham University (available from *www.npi-statistics.com*).

[17] Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. Wiley, Chichester.

[18] Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1987). *Statistics for Experimenters*. Wiley, Chichester.

[19] Coolen, F.P.A. (1996). Comparing two populations based on low stochastic structure assumptions. *Statistics and Probability Letters*, **29**, 297-305.

[20] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics and Probability Letters*, **36**, 349-357.

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

[22] 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.

[23] Coolen, F.P.A. and Alqifari, H.N. (2017). Nonparametric predictive inference for reproducibility of two basic tests based on order statistics. *REVSTAT - Statistical Journal*.

[24] Coolen, F.P.A. and Augustin, T. (2005). Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model. In *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*, vol. 5, 125-134.

[25] 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.

[26] Coolen, F.P.A. and Coolen-Schrijner, P. (2006). Nonparametric predictive subset selection for proportions. *Statistics and Probability Letters*, **76**, 1675-1684.

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

[28] 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.

[29] Coolen, F.P.A., Coolen-Maturi, T. and Alqifari, H.N. (2017) Nonparametric predictive inference for future order statistics. *Journal of Communications in Statistics – Theory and Methods*, to appear.

[30] 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.

[31] 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.

[32] Coolen, F.P.A. and Yan, K.J. (2003). Nonparametric predictive comparison of two groups of lifetime data. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, vol. 3, 148-161.

[33] Coolen-Maturi, T., Coolen, F.P.A. and Muhammad, N. (2016). Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. *Journal of Statistical Theory and Practice*, **10**, 515-538.

[34] 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.

[35] Coolen-Schrijner, P., Maturi, T.A. and Coolen, F.P.A. (2009). Nonparametric predictive precedence testing for two groups. *Journal of Statistical Theory and Practice*, **3**, 273-287.

[36] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*. Wiley, Hoboken.

[37] De Capitani, L. and De Martini, D. (2011). On stochastic orderings of the Wilcoxon rank sum test statistic—with applications to reproducibility probability estimation testing. *Statistics and Probability Letters*, **81**, 937-946.

[38] De Capitani, L. and De Martini, D. (2016). Reproducibility probability estimation and RP-testing for some nonparametric tests. *Entropy*, **18**(4), 142.

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

[40] De Martini, D. (2008). Reproducibility probability estimation for testing statistical hypotheses. *Statistics and Probability Letters*, **78**, 1056-1061.

[41] De Waal, T., Pannekoek, J. and Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation.* Wiley, New York.

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

[43] Epstein, B. (1955). Comparison of some non-parametric tests against normal alternatives with an application to life testing. *Journal of the American Statistical Association*, **50**, 894-900.

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

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

[46] Gibbons, J.D. and Chakraborti, S. (2011). *Nonparametric Statistical Inference* (5th ed.). Chapman and Hall, Boca Raton, Florida.

[47] Gibbons, J.D., Olkin, I. and Sobel, M. (1999). *Selecting and Ordering Populations: a New Statistical Methodology.* Society for Industrial and Applied Mathematics, Philadelphia.

[48] Goodman, S.N. (1992). A comment on replication, p-values and evidence. *Statistics in Medicine*, **11**, 875–879.

[49] Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press, New York.

[50] Hampel, F.R. (1968) *Contributions to the Theory of Robust Estimation*. PhD Thesis, University of California. USA.

[51] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383-393.

[52] Hampel, F. R. (1997). What can the foundations discussion contribute to data analysis? And what may be some of the future directions in robust methods and data analysis? *Journal of Statistical Planning and Inference*, **57**, 7-19.

[53] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: the Approach Based on Influence Functions*. Wiley, Chichester.

[54] He, X., Simpson, D. G., and Portnoy, S. L. (1990). Breakdown robustness of tests. *Journal of the American Statistical Association*, 85(410), 446-452.

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

[56] Hill, B.M. (1988). De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). In J.M. Bernardo, et al. (Eds.), *Bayesian Statistics*, **3**, 211-241. Oxford University Press.

[57] Hill, B.M. (1992). Bayesian nonparametric prediction and statistical inference. *Bayesian Analysis in Statistics and Econometrics*, **75**, 43-94.

[58] Hill, B.M. (1993). Parametric models for A_n: splitting processes and mixtures. *Journal of the Royal Statistical Society.*, **55**, 423-433.

[59] Hodges, J.L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 163-186. University of California Press.

[60] Huber, P.J. (1996). *Robust Statistical Procedures.* Society for Industrial and Applied Mathematics, Philadelphia.

[61] Huber, P.J. (2002). John W. Tukey's contributions to robust statistics. *Annals of Statistics*, **137**, 1640–1648.

[62] Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics* (Second Edition). Wiley, Chichester.

[63] Jurecková, J. and Sen, P.K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* Wiley, Chichester.

[64] Kamps, U. (1981). *Generalized Order Statistics.* Wiley, Chichester.

[65] Kochar, S. (2012). Stochastic comparisons of order statistics and spacings: a review. *International Scholarly Research Notices: Probability and Statistics.*

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

[67] Mahalik, N.P. (2007). *Sensor Networks and Configuration.* Springer, Berlin.

[68] Markatou, M., and He, X. (1994). Bounded influence and high breakdown point testing procedures in linear models. *Journal of the American Statistical Association*, 89(426), 543-549.

[69] Maturi, T. (2010) *Nonparametric Predictive Inference for Multiple Comparisons.* PhD Thesis, Durham University. UK.

[70] Maturi, T.A., Coolen-Schrijner, P. and Coolen, F.P.A. (2009). Nonparametric predictive pairwise comparison for real-valued data with terminated tails. *International Journal of Approximate Reasoning*, **51**, 141-150.

[71] Miller, R.G. (1964). A trustworthy jackknife. *The Annals of Mathematical Statistics*, **35**, 1594-1605.

[72] Miller, R.G. (1974). The Jackknife–A Review. *Biometrika*, **61**, 1-15.

[73] Nelson, L.S. (1963). Tables for a precedence life test. *Technometrics*, **124**, 491-499.

[74] Nelson, W.B. (1982). *Applied Life Data Analysis.* Wiley, Hoboken, New Jersey.

[75] Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, **43 (3/4)**, 353-360.

[76] Senn, S. (2002). Comment on 'A comment on replication, p-values and evidence' by S.N. Goodman (letter to the editor). *Statistics in Medicine*, **21**, 2437-2444.

[77] Shao, J. and Chow, S.C. (2002). Reproducibility probability in clinical trials. *Statistics in Medicine*, **21**, 1727-1742.

[78] Steele, S. (December 21, 1997), *Babies by the Dozen for Christmas: 24-Hour Baby Boom.* The Sunday Mail.

[79] 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.

[80] Tukey, J.W. (1958). Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, **29**, 614-623.

[81] Tukey, J.W. (1977). *Exploratory data analysis.* Pearson, Reading, Massachusetts.

[82] 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.

[83] 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.

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

[85] Weichselberger, K. (1995). Axiomatic foundations of the theory of interval-probability. In: V. Mammitzsch and H. Schneeweiss (eds.), *Proceedings of the Second Gauß symposion, Section B*. De Gruyter, Berlin, 47-64.

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

[87] Weichselberger, K. (2001). *Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept*. Physika, Heidelberg.

[88] Wetherill, G. B., and Ofosu, J. B. (1974). Selection of the best of k normal populations. *Applied statistics*, 253-277.

Appendix A

Proof properties of Equation (2.2)

For all $r = 1, \ldots, m$ and $j = 1, \ldots, n+1$ we have the obvious symmetry

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

This can be proved straightforward by applying Equation (2.2) to the right hand side of Equation (A.1),

$$P(X_{(m+1-r)} \in I_{n+2-j}) = \frac{\binom{(m+1-r)+(n+2-j)-2}{(n+2-j)-1}\binom{n-(n+2-j)+1+m-(m+1-r)}{n-(n+2-j)+1}}{\binom{n+m}{n}}$$

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

$\square$

Further, one of the properties we want to consider is whether $P(X_{(r)} \in I_j)$ is unimodal. The probability distribution function given by Equation (2.2) is unimodal if and only if it has a mode $j^* \in \{1, \ldots, n+1\}$ such the function is non-decreasing on $(-\infty, j^*]$, while on $[j^*, \infty)$ it is non-increasing. If j^* is a mode for $P(X_{(r)} \in I_j)$ then $P(X_{(r)} \in I_j) \leq P(X_{(r)} \in I_j^*)$ for all j. The probability for $X_{(r)} \in I_j$ is unimodal in j, with the maximum probability of $X_{(r)} \in I_{j^*}$ with,

$$(\frac{r-1}{m-1})(n+1) \leq j^* \leq (\frac{r-1}{m-1})(n+1) + 1 \tag{A.2}$$

To prove that the probability in Equation (2.2) is unimodal with the single local maximum in interval I_j^*, we divide the proof into two cases. For $j \leq j^*$, we show that $P(X_{(r)} \in I_{j-1}) \leq P(X_{(r)} \in I_j)$, and for $j \geq j^*$, we show that $P(X_{(r)} \in I_j) \geq P(X_{(r)} \in I_{j+1})$.

Case 1. For $1 \leq r \leq m$ and $j \leq j^*$

$$P(X_{(r)} \in I_{j-1}) \leq P(X_{(r)} \in I_j) \tag{A.3}$$

From Equation (2.2), Equation (A.3) is true if and only if

$$\binom{r+j-2}{j-1}\binom{n-j+1+m-r}{n-j+1} \geq \binom{r+(j-1)-2}{(j-1)-1}\binom{n-(j-1)+1+m-r}{n-(j-1)+1}$$

$$\Leftrightarrow \frac{(r+j-2)!(n-j+m-r+1)!(j-2)!(n-j+2)!}{(r+j-3)!(n-j+m-r+2)!(j-1)!(n-j+1)!} \geq 1$$

$$\Leftrightarrow \frac{(r+j-2)(n-j+2)}{(n-j+2+m-r)(j-1)} \geq 1$$

$$\Leftrightarrow ((r-1)+(j-1))(n-j+2) \geq ((n-j+2)+(m-r))\,(j-1)$$

$$\Leftrightarrow rn - rj + 2r - n + j - 2 \geq mj - m - rj + r$$

$$\Leftrightarrow j(m-1) \leq rn + r - n - 2 + m$$

$$\Leftrightarrow j(m-1) \leq (r-1)(n+1) + (m-1)$$

$$\Leftrightarrow j \leq \frac{(r-1)(n+1)}{(m-1)} + 1 \tag{A.4}$$

from the RHS of inequality in Equation (A.2) it is known that $j^* \leq \frac{(r-1)(n+1)}{(m-1)} + 1$ so together with $j \leq j^*$, Equation(A.4) holds.

Case 2. For $1 \leq r \leq m$ and $j \geq j^*$

$$P(X_{(r)} \in I_j) \geq P(X_{(r)} \in I_{j+1}) \tag{A.5}$$

This is true if and only if

$$\binom{r+j-2}{j-1}\binom{n-j+1+m-r}{n-j+1} \geq \binom{r+(j+1)-2}{(j+1)-1}\binom{n-(j+1)+1+m-r}{n-(j+1)+1}$$

$$\Leftrightarrow \frac{j(n-j+1+m-r)}{(n-j+1)(r+j-1)} \geq 1$$

$$\Leftrightarrow j(m-r) \geq (n-j+1)(r-1)$$

$$\Leftrightarrow jm - jr \geq nr - jr + r - n + j - 1$$

$$\Leftrightarrow j(m-1) \geq (r-1)(n+1)$$

$$\Leftrightarrow j \geq \frac{(r-1)(n+1)}{(m-1)} \tag{A.6}$$

By the first inequality in equation (A.2) it is known that $j^* \geq \frac{(r-1)(n+1)}{(m-1)}$ thus together with $j \geq j^*$ Equation (A.6) holds.

$\square$

Appendix B

Proof of Theorem 3.1

Proof of Theorem 3.1

The NPI lower and upper probabilities (3.3) and (3.4) 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 (3.3) 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 $A_{(.)}$ 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 (3.4) 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 (3.4), which actually involves maximisation of the probability given in Equation (3.5). 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 (3.5) 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 (3.5) 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.

□

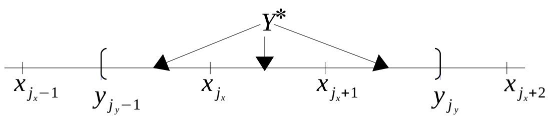

Figure B.1: Upper probabilities, taking the case when $Y_{(r)}, Y_{(s)}$ fall in the same interval I^y_j

For example, if $Y^* \in (y_{j_y-1}, y_{j_y})$ and there are observations from group X that fall in this interval I_{j_y}, as illustrated in Figure B.1, the x observations divided the interval $I^y_{j_y}$ into three intervals (y_{j_y-1}, x_{j_x}),(x_{j_x}, x_{j_x+1}) and (x_{j_x+1}, y_{j_y}), as we have freedom to put $Y^* \in I^y_{j_y}$ at any point in the interval $I^y_{j_y}$. Since we could put Y^* in any point between these intervals, let us take the middle point. Then we take the maximum probabilities of these.

$$\max\left(P_{\frac{y_{j_y-1}+x_{j_x}}{2}}, P_{\frac{x_{j_x}+x_{j_x+1}}{2}}, P_{(\frac{x_{j_x+1}+y_{j_y}}{2})}\right)$$

.

For the case where interval I_j^y does not contain any observed value from group X then $P(Y_{(r)} \in I_{j_y}^y, Y_{(s)} \in I_{j_y}^y)$ can be assigned to any end point in the interval (y_{j_y-1}, y_{j_y}), as the number of x observations to the right of y_{j_y-1} is the same as the number of the x observations at the right of y_{j_y}. As we count the number of x in the right and left y^*, it does not matter where the joint probability mass is assigned in which end point. To be consistent, we also take the medial point $y^* = \frac{y_{j_y-1}+y_{j_y}}{2}$.

Appendix C

$P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*})$

The probability $P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*})$ based on $n+1$ observations and an observation added to the left of interval I_j, so $y \in I_{j^*}$ for $j^* < j$, is straightforward from Equation (2.2) by replacing n by $n+1$ and j by $j+1$,

$$P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) = \binom{j+r-1}{j}\binom{n-j+1+m-r}{n-j+1}\binom{n+m+1}{n+1}^{-1} \tag{C.1}$$

Similarly, for $j^* > j$, n is replaced in Equation (2.2) by $n+1$ but j is unchanged,

$$P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*}) = \binom{j+r-2}{j-1}\binom{n-j+2+m-r}{n-j+2}\binom{n+m+1}{n+1}^{-1} \tag{C.2}$$

For $j^* = j$, we get

$$\begin{aligned} P(X_{(r)} \in I_j | y \in I_j) &= P(X_{(r)} \in I_j \cup \tilde{I}_{j+1} | y \in I_j) \\ &= P(X_{(r)} \in \tilde{I}_j | y \in I_{j*}) + P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) \end{aligned}$$

The probability in Equation (C.1) is the same for all $j^* = 1, \ldots, j-1$. Similarly, Equation (C.2), is the same for all $j_2^* = j+1, \ldots, n+1$. Analysis of the probabilities in Equations (C.1) and (C.2) leads to the symmetry property $P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*}) = P(X_{(m+1-r)} \in \tilde{I}_{n-j+2} | y \in I_{j^*})$ for $j^* < j$ and $j^* > j$ respectively, which can be proved straightforwardly by replacing r by $m+1-r$ and j by $n-j+2$ in

Equation (C.2), which gives Equation (C.1). Similarly for $j^* = j$

$$P(X_{(r)} \in \tilde{I}_j \cup \tilde{I}_{j+1} | y \in I_j) = P(X_{(m+1-r)} \in \tilde{I}_{n-j+2} \cup \tilde{I}_{n-j+3} | y \in I_{n-j+2})$$

Appendix D

Proof of Theorem 5.1

Proof. **Proof of Theorem 5.1**

We show that for $m > 1$,

$$P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) \geq P\left(X_{(r)} \in I_j\right) \quad \text{for} \quad j^* < j \quad \text{if and only if} \, j \leq \frac{(r-1)(n+1)}{m} \tag{D.1}$$

and

$$P(X_{(r)} \in \tilde{I}_{j} | y \in I_{j^*}) \geq P\left(X_{(r)} \in I_j\right) \quad \text{for} \quad j^* > j \quad \text{if and only if} \quad j \geq \frac{r(n+1)}{m} \tag{D.2}$$

Using Equations (2.2) and (C.1),

$$\begin{aligned}
& P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) \geq P\left(X_{(r)} \in I_j\right) \\
\Leftrightarrow & \frac{\binom{j+r-1}{j}\binom{n-j+1+m-r}{n-j+1}\binom{n+m+1}{n+1}^{-1}}{\binom{j+r-2}{j-1}\binom{n-j+1+m-r}{n-j+1}\binom{n+m}{n}^{-1}} \geq 1 \\
\Leftrightarrow & \frac{(n+1)(j+r-1)}{j(n+m+1)} \geq 1 \\
\Leftrightarrow & \, j \leq \frac{(r-1)(n+1)}{m}
\end{aligned} \tag{D.3}$$

and

$$P(X_{(r)} \in \tilde{I}_{j} | y \in I_{j^*}) \geq P\left(X_{(r)} \in I_j\right)$$

$$
\begin{aligned}
&\Leftrightarrow \frac{\binom{j+r-2}{j-1}\binom{n-j+2+m-r}{n-j+2}\binom{n+m+1}{n+1}^{-1}}{\binom{j+r-2}{j-1}\binom{n-j+1+m-r}{n-j+1}\binom{n+m}{n}^{-1}} \geq 1 \\
&\Leftrightarrow \frac{(n+1)(n-j+2+m-r)}{(n+m+1)(n-j+2)} \geq 1 \\
&\Leftrightarrow (n+1)\left[(n+1)-(j-1)+(m-r)\right] \geq ((n+1)+m)((n+1)-(j+1)) \\
&\Leftrightarrow j \geq \frac{r(n+1)}{m}+1 \qquad \text{(D.4)}
\end{aligned}
$$

$\square$