



# Robustness of Nonparametric Predictive Inference for Future Order Statistics

Hana N. Alqifari<sup>1</sup> · Frank P. A. Coolen<sup>2</sup>

© The Author(s) 2018

## Abstract

This paper considers robustness of Nonparametric Predictive Inference (NPI), in particular considering inference involving future order statistics. 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 paper we use it in a slightly narrower sense. For our aims, robustness indicates insensitivity to small change 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 to the NPI framework, namely sensitivity curve and breakdown point; these classical concepts require some adoption for application in NPI. Most of our nonparametric inferences have a reasonably good robustness with regard to small changes in the data.

**Keywords** Future order statistics · Lower and upper probabilities · Nonparametric Predictive Inference · Robust · Sensitivity curve · Breakdown point

**Mathematics Subject Classification** 62G30 · 62G35

---

The authors would like to dedicate this paper to Frank Hampel, who passed away on 2nd October 2018. Frank was a great inspiration to the authors, both with regard to foundations of statistics and theory of robustness. More importantly, he was an exceptionally kind person and a personal friend of the second-named author, who feels very privileged to have experienced this friendship.

---

✉ Frank P. A. Coolen  
frank.coolen@durham.ac.uk

Hana N. Alqifari  
hana.alqifari@gmail.com

<sup>1</sup> Department of Mathematics Sciences, Qassim University, Buraidah, Saudi Arabia

<sup>2</sup> Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

## 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 [22]. Since the middle of the twentieth century, much theoretical effort has been dedicated to develop statistical procedures that are resistant with regard to outliers and robust with regard to small deviations from an assumed parametric model [6]. Huber [21] provides the basic theory of robust statistics. Hampel et al. [17] 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 [23]. It is well known that some classical procedures are not robust to slight contamination of the strict model assumptions [6]. From this perspective robustness against small deviations from the assumed model and existence of outliers or contamination, have all been identified as principal issues [23]. In classical robust statistics, there are several tools used to describe robustness, e.g. the influence function, the sensitivity curve and the breakdown point.

This paper introduces robustness of NPI. 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 errors occurring during the measurement and repeating process [11]. 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 paper, we use it in a slightly narrower sense, as for our aims robustness indicates insensitivity to a small change in the data or to outliers.

This paper is organized as follows. Section 2 provides a brief introduction to NPI, including key results on NPI for future order statistics as used in this paper. Section 3 provides a brief overview of some concepts used in robust statistics, namely influence function, sensitivity curve and breakdown point. In Sect. 4 we introduce the sensitivity curve and breakdown point in the NPI framework. Section 5 presents the use of these tools for NPI for events involving the  $r$ -th future observation. In Sect. 6 we use these tools to explore the robustness of the inferences involving the median and the mean of the  $m$  future observations. In Sect. 7, we briefly present NPI robustness of further inferences, namely pairwise comparisons and reproducibility of statistical tests. The paper ends with some concluding remarks in Sect. 8.

## 2 Nonparametric Predictive Inference

Nonparametric Predictive Inference (NPI) [5, 7] is a statistical framework which uses few modelling assumptions, with inferences explicitly in terms of future observations. For real-valued random quantities attention has thus far been mostly restricted to a single future observation, although multiple future observations have been considered for some NPI methods, e.g. in statistical process control [2, 3].

Assume that we have real-valued ordered data  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ , with  $n \geq 1$ . For ease of notation, define  $x_{(0)} = -\infty$  and  $x_{(n+1)} = \infty$ , or define these equal to other known lower and upper bounds of the range of possible values for these random quantities. The  $n$  observations create a partition of the real-line into  $n + 1$  intervals  $I_j = (x_{(j-1)}, x_{(j)})$  for  $j = 1, \dots, n + 1$ . We assume throughout this paper that ties do not occur. If we wish to allow ties, also between past and future observations, we could use closed intervals  $[x_{(j-1)}, x_{(j)}]$  instead of these open intervals  $I_j$ , the difference is rather minimal and to keep presentation easy we have opted not to do this here. We are interested in  $m \geq 1$  future observations,  $X_{n+i}$  for  $i = 1, \dots, m$ . We link the data and future observations via Hill's assumption  $A_{(n)}$  [19], or, more precisely, via  $A_{(n+m-1)}$  (which implies  $A_{(n+k)}$  for all  $k = 0, 1, \dots, m - 2$ ; we will refer to this generically as 'the  $A_{(n)}$  assumptions'), which can be considered as a post-data version of a finite exchangeability assumption for  $n + m$  random quantities. The  $A_{(n)}$  assumptions imply that all possible orderings of the  $n$  data observations and the  $m$  future observations are equally likely, where the  $n$  data observations are not distinguished among each other and neither are the  $m$  future observations. Let 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  $i = 1, \dots, \binom{n+m}{n}$ , so that  $S_j^i = \#\{X_{n+l} \in I_j, l = 1, \dots, m | O_i\}$ . Then the  $A_{(n)}$  assumptions lead to [10]

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

where  $s_j^i$  are non-negative integers with  $\sum_{j=1}^{n+1} s_j^i = m$ . Equation (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  $A_{(n)}$  assumptions with  $n$  data observations and  $m$  future observations is to think that  $n$  randomly chosen observations out of all  $n + m$  real-valued observations are revealed, following which you wish to make inferences about the  $m$  unrevealed observations. The  $A_{(n)}$  assumptions then imply that one has no information about whether specific values of neighbouring revealed observations make it less or more likely that a future observation falls in between them. For any event involving the  $m$  future observations, Eq. (1) implies that we can count the number of such orderings for which this event holds. Generally in NPI, a lower probability for the event of interest is derived by counting all orderings for which this event has to hold, while the corresponding upper probability is derived by counting all orderings for which this event can hold [5, 7].

In NPI, the  $A_{(n)}$  assumptions justify the use of resulting inferences directly as predictive probabilities. Using only precise probabilities, 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 [13], that corresponding optimal bounds can be derived for all events of interest [5]. These bounds are lower and upper probabilities

in the theory of imprecise probability [4]. NPI provides frequentist inferences which are exactly calibrated in the sense of [24], and it has strong consistency properties in theory of interval probability [5]. NPI is always in line with inferences based on empirical distributions, which is an attractive property when aiming at objectivity [7]. In NPI the  $n$  observations are explicitly used through the  $A_{(n)}$  assumptions, yet as there is no use of conditioning as in the Bayesian framework, we do not use an explicit notation to indicate this use of the data. The  $m$  future observations must be assumed to result from the same sampling method as the  $n$  data observations in order to have full exchangeability. NPI is totally based on the  $A_{(n)}$  assumptions, which however should be considered with care as they imply, e.g. that the specific ordering in which the data appeared is irrelevant, so accepting  $A_{(n)}$  implies an exchangeability judgement for the  $n$  observations. It is attractive that the appropriateness of this approach can be decided upon after the  $n$  observations have become available.

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

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

For this event NPI provides a precise probability, as each of the  $\binom{n+m}{n}$  equally

likely orderings of  $n$  past and  $m$  future observations has the  $r$ -th ordered future observation in precisely one interval  $I_j$ . As Eq. (2) only specifies the probabilities for the events that  $X_{(r)}$  belongs to intervals  $I_j$ , it can be considered to provide a partial specification of a probability distribution for  $X_{(r)}$ ; no assumptions are made about the distribution of the probability masses within such intervals  $I_j$ .

Analysis of the probability in Eq. (2) leads to some interesting results, including the logical symmetry  $P(X_{(r)} \in I_j) = P(X_{(m+1-r)} \in I_{n+2-j})$ . For all  $r$ , the probability for  $X_{(r)} \in I_j$  is unimodal in  $j$ , with the maximum probability assigned to interval  $I_{j^*}$  with  $\left(\frac{r-1}{m-1}\right)(n+1) \leq j^* \leq \left(\frac{r-1}{m-1}\right)(n+1) + 1$ . A further interesting property occurs for the special case where the number of future observations is equal to the number of data observations, so  $m = n$ . In this case,  $P(X_{(r)} < x_r) = P(X_{(r)} > x_r) = 0.5$  holds for all  $r = 1, \dots, m$ . This fact can be proven by considering all  $\binom{2n}{n}$  equally likely

orderings, where clearly in precisely half of these orderings the  $r$ -th future observation occurs before the  $r$ -th data observation due to the overall exchangeability assumption.

For an event  $X_{(r)} \in I_j$ , the  $A_{(n)}$  assumptions provide precise probabilities. More generally, interest may be in an event  $X_{(r)} \in Z$ , with  $Z$  any subset of the real values, for example an interval not equal to one of the intervals  $I_j$  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 [4]. The NPI lower and upper probabilities are

$$\underline{P}(X_{(r)} \in Z) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \subseteq Z\}P(X_{(r)} \in I_j) \tag{3}$$

$$\bar{P}(X_{(r)} \in Z) = \sum_{j=1}^{n+1} \mathbf{1}\{I_j \cap Z \neq \emptyset\}P(X_{(r)} \in I_j) \tag{4}$$

The lower probability (3) is obtained by summing up only the probability masses that must be in  $Z$ . The upper probability (4) is obtained by summing up all probability that can be in  $Z$ . The NPI lower and upper probabilities for the event that  $X_{(r)} > z$ , where  $z$  is not equal to one of the data observations, are

$$\underline{P}(X_{(r)} > z) = \sum_{j=1}^{n+1} \mathbf{1}\{x_{j-1} > z\}P(X_{(r)} \in I_j) \tag{5}$$

$$\bar{P}(X_{(r)} > z) = \sum_{j=1}^{n+1} \mathbf{1}\{x_j > z\}P(X_{(r)} \in I_j) \tag{6}$$

We denote the median of  $m$  future observations by  $M_m$ . For  $m$  odd, so  $M_m = X_{(\frac{m+1}{2})}$ , the NPI probability for the event  $M_m \in I_j = (x_{j-1}, x_j)$  can be derived straightforwardly from Eq. (2). NPI for the median of  $m$  future observations is relatively more complicated if  $m$  is even, in which case  $M_m = (X_{(\frac{m}{2})} + X_{(\frac{m}{2}+1)})/2$ . In this case NPI does not provide precise probabilities for the event  $M_m \in I_j$  but lower and upper probabilities, which are presented in the PhD thesis of [1].

We denote the mean of  $m$  future observations by  $\mu_m$ , and the mean corresponding to a specific ordering  $O_i$  of the future observations among  $n$  observations by  $\mu_m^i$ . When we consider  $\mu_m$  and  $\mu_m^i$ , we must avoid possible probability mass in  $-\infty$  or  $\infty$ , 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  $L < x_1 < \dots < x_n < R$ , and we define  $x_0 = L$  and  $x_{n+1} = R$  for the  $A_{(n)}$  assumptions. The maximum lower bound and the minimum upper bound for the mean  $\mu_m^i$  of the  $m$  future observations, for given ordering  $O_i$ , are

$$\underline{\mu}_m^i = \frac{1}{m} \sum_{j=1}^{n+1} s_j^i x_{j-1} \tag{7}$$

$$\bar{\mu}_m^i = \frac{1}{m} \sum_{j=1}^{n+1} s_j^i x_j \quad (8)$$

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

$$\underline{P}(\mu_m \geq z) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{\underline{\mu}_m^i \geq z\} \quad (9)$$

$$\bar{P}(\mu_m \geq z) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{\bar{\mu}_m^i \geq z\} \quad (10)$$

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

$$\underline{P}(\mu_m \in (z_1, z_2)) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{z_1 \leq \underline{\mu}_m^i \leq \bar{\mu}_m^i \leq z_2\} \quad (11)$$

$$\bar{P}(\mu_m \in (z_1, z_2)) = \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{\binom{n+m}{n}} 1\{(\underline{\mu}_m^i, \bar{\mu}_m^i) \cap (z_1, z_2) \neq \emptyset\} \quad (12)$$

### 3 Classical Concepts for Evaluating Robustness

In the literature of robustness, many measures of robustness of an estimator have been introduced [16, 17]. 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 [16]. Let the CDF  $F$  denote the true underlying distribution function, and CDF  $G_\xi$  a contaminating distribution which puts all its mass in  $\xi$ . For an estimator  $T$  based on data from a population with CDF  $F$ , the influence function of  $T$  at basic distribution  $F$  is

$$IF_{T,F}(\xi) = \lim_{\epsilon \rightarrow 0} \frac{T((1 - \epsilon)F + \epsilon G_\xi) - T(F)}{\epsilon} \tag{13}$$

Here  $(1 - \epsilon)F + \epsilon G_\xi$  with  $0 < \epsilon < 1$  is a mixture distribution of  $F$  and  $G_\xi$ . 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 (13), the most important being the sensitivity curve [28] and the empirical influence function [17]. Let  $T_n(X) = T_n(x_1, \dots, x_n)$  denote a statistic of the sample  $X = (x_1, \dots, x_n)$  and let  $T_{n+1}(X, \xi)$  denote the corresponding statistic of the sample  $x_1, \dots, x_n, \xi$ . The simplest idea is the empirical influence function [17].

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

This  $EIF_i$  is defined by replacing the  $i$ -th value in the sample  $X$  by an arbitrary value  $\xi$  and looking at the output of the estimator [17]. Alternatively, one can define it by adding an observation, i.e. when the original sample consists of  $n$  observations one can add an arbitrary value  $\xi$  [17, p. 93]. The second tool is the sensitivity curve [28]. Again there are two versions, one with addition and one with replacement [17]. In case of additional an observation, the sensitivity curve (SC) is defined as [22]

$$SC_n(\xi, T_n, X) = (n + 1)(T_{n+1}(X, \xi) - T_n(X))$$

$SC_n(\xi, T_n, X)$  measures the sensitivity of  $T_n$  to the addition of one observation with value  $\xi$  [22]. The sensitivity curve measures sensitivity of an estimator to a change in the sample. In case of replacing an observation  $x_i$  by  $\xi$ , let  $T_n(X, \xi, i)$  denote a statistic of the sample  $(x_1, \dots, x_{i-1}, \xi, x_i, \dots, x_n)$ , then the SC is defined as [22]  $SC_i(\xi, T_n, X) = n(T_n(X, \xi, i) - T_n(X))$ . This version of SC measures the sensitivity of  $T_n$  to replacing the  $i$ -th value in the sample by an arbitrary value.

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 [23]. Let  $X = (x_1, \dots, x_n)$  be a fixed sample of size  $n$ . We can contaminate this sample in many ways [22]. We consider the following two;  $\lambda_a$  replacement and  $\lambda_b$  contamination. These will also be considered in the NPI setting in Sect. 4. First,  $\lambda_a$  replacement: we replace an arbitrary subset of size  $l$  of the sample by arbitrary values  $y_1, \dots, y_l$ , so  $1 \leq l \leq n$  [22]. Let  $X'$  denote the contaminated sample. The fraction of contaminated values in the contaminated sample  $X' = (x_1, \dots, x_{l-1}, y_l, \dots, y_n)$ , is  $\lambda_a = \frac{l}{n}$ . Secondly,  $\lambda_b$  contamination: we add  $l$  arbitrary additional values  $Y = (y_1, \dots, y_l)$  to the sample  $X$  [22]. Let  $X''$  denote the sample contaminated by adding  $l$  arbitrary additional values. Thus, the

fraction of contaminated values in the contaminated sample  $X'' = X \cup Y$  is  $\lambda_b = \frac{l}{l+n}$ . 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  $\lambda$ , which is either  $\lambda_a$  or  $\lambda_b$ , is [22]

$$b(\lambda; X, T) = \sup \{|T(X, Y) - T(X)|\} \quad (14)$$

where the supremum is taken over the set of all  $\lambda$ -contaminated samples, which is either  $X'$  or  $X''$ . The definition of the breakdown point is

$$\lambda^*(X, T) = \inf\{\lambda | b(\lambda; X, T) = \infty\} \quad (15)$$

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

## 4 Robustness Concepts in NPI

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 substituted for an original observation. We do not consider IF for NPI, as IF depends on the assumed distribution and in the NPI approach we do not assume any underlying distribution. In our study of the robustness of NPI, we will focus on the sensitivity curve (SC) and breakdown point (BP) as they typically rely on the actual data at hand rather than on a hypothetical underlying population. We can also adopt EIF, but we prefer to only focus on SC as local measurement of our predictive inferences.

Let  $\underline{x} = \{x_1, \dots, x_n\}$  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, as discussed in Sect. 3, and we consider two of them; substituting a contaminant for one of the  $n$  observations or adding an additional observation to the past data. We denote these contaminated data by  $\underline{x}(j, \delta)$  and  $(\underline{x}, y)$ , respectively. These two ways of contaminating the sample will be studied separately in the NPI framework. We first focus on the effect of adding  $\delta$  to one of the observations in the past data, as it is convenient and logical to do this in the NPI method. Let  $I(\underline{x}(j, \delta))$  denote the inference of interest based on the contaminated data  $\underline{x}(j, \delta)$ , where the data are contaminated by replacing  $x_j$  by  $x_j + \delta$  in  $\underline{x}$ . 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

$$SC_I(\underline{x}(j, \delta)) = I(\underline{x}(j, \delta)) - I(\underline{x}) \quad (16)$$

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 Sect. 3. We may multiply  $SC_I(\underline{x}(j, \delta))$  by  $n$ , but in our case Eq. (16) 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 the sample



size  $n$  to make the evaluation less sensitive to  $n$ . Let  $I(\underline{x}, y)$  denote the inference of interest based on the contaminated data, where the data are contaminated by adding  $y$  to  $\underline{x}$ . The NPI-SC, in the case of adding an additional observation  $y$  to the data, is

$$SC_I(\underline{x}, y) = I(\underline{x}, y) - I(\underline{x}) \tag{17}$$

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.

A finite sample breakdown point (BP) was first proposed by [20], as ‘‘tolerance of extreme values’’ in the situation of location parameter problems, and it was generalized for a variety of cases by [15]. As far as we know, 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  $\lambda_c^*(I, \underline{x}(j_1, \dots, j_l, \delta))$ .

To introduce the  $c$ -breakdown point concept, we first need to introduce some notation related to the way of contamination of the data  $\underline{x}$ , as discussed in Sect. 3. First, ‘replacement’: we replace a subset of size  $l$  of the data  $\underline{x}$  by  $x_{j_1} + \delta, \dots, x_{j_l} + \delta$ , where  $1 \leq l \leq n$ . We denote these contaminated data by  $\underline{x}(j_1, \dots, j_l, \delta)$ . Let  $I(\underline{x}(j_1, \dots, j_l, \delta))$  denote the inference of interest based on the contaminated data. Note that  $\delta$  can be vary for each value, i.e.  $\delta_{j_i}$  for  $i = 1, \dots, l$ , and we denote these contaminated data by  $\underline{x}(j_1, \dots, j_l, \delta_{j_1}, \dots, \delta_{j_l})$  for different  $\delta$ . The fraction of contaminant values in the contaminated sample  $\underline{x}(j_1, \dots, j_l, \delta)$  is  $\lambda_a = \frac{l}{n}$ . Secondly, ‘addition’: we add  $l$  arbitrary additional observations  $y_1, \dots, y_l$  to the past data  $\underline{x}$ . We denote these contaminated data by  $(\underline{x}, y_1, \dots, y_l)$ . The inference is denoted by  $I(\underline{x}, y_1, \dots, y_l)$ . The fraction of contaminant values in the contaminated sample  $(\underline{x}, y_1, \dots, y_l)$ , is  $\lambda_b = \frac{l}{l+n}$ . The maximum bias which might be caused by  $\lambda_a$ -replac-

ment, is

$$\begin{aligned} b(\lambda_a; \underline{x}, I) &= \sup |I(\underline{x}(j_1, \dots, j_l, \delta)) - I(\underline{x})| \\ &= \sup |(SC_I(\underline{x}(j_1, \dots, j_l, \delta)))| \end{aligned} \tag{18}$$

where the supremum is taken over the set of all  $\lambda_a$ -replacement samples  $\underline{x}(j_1, \dots, j_l, \delta)$ , with  $\{j_1, \dots, j_l\} \subset \{1, \dots, n\}$  for fixed  $\delta$  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  $\lambda_b$ -contamination is

$$\begin{aligned} b(\lambda_b; \underline{x}, I) &= \sup |I(\underline{x}, y_1, \dots, y_l) - I(\underline{x})| \\ &= \sup |(SC_I(\underline{x}, y_1, \dots, y_l))| \end{aligned} \tag{19}$$

where the supremum is taken over the set of all  $\lambda_b$ -contaminated samples  $(\underline{x}, y_1, \dots, y_l)$ , with  $y_1, \dots, y_l \in \mathbb{R}$  of given data  $\underline{x}$ . The  $c$ -breakdown point, where  $c \in [0, 1]$ , for the case of  $\lambda_a$ -replacement, is defined as

$$\lambda_c^*(I, \underline{x}(j_1, \dots, j_l, \delta)) = \inf \{ \lambda_a | b(\lambda_a; \underline{x}, I) > c \} \tag{20}$$

Alternatively, the *c-breakdown point* for the case of adding  $l$  observations to the original sample ( $\lambda_b$ -contamination), is

$$\lambda_c^*(I, (\underline{x}, y_{j_1}, \dots, y_{j_l})) = \inf\{\lambda_b | b(\lambda_b; \underline{x}, I) > c\} \tag{21}$$

The *c-breakdown point* is the smallest fraction of contamination in the past data that could cause a predictive inference 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 of the inference of interest. The value  $c$  determines how much we allow the inference to change before its breakdown.

### 5 Robustness of NPI for the $r$ th Future Order Statistic

To illustrate the use of the robustness concepts for NPI, namely NPI-SC and NPI-BP as defined in Sect. 4, we first consider the probabilities for events involving the  $r$ -th ordered future observation. We illustrate both ways that the sample can be contaminated.

#### 5.1 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 Eq. (2). The probability (2) is only affected by replacing contamination if the indices,  $k = 1, \dots, n + 1$ , 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

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

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

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

This NPI-SC depends on the value of  $r$  and which interval it falls in, and will be illustrated in Example 1 in Sect. 5.4.

### 5.2 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  $r$ th ordered future observation falls in interval  $I_j$ , by considering

$$SC_{P(X_{(r)} \in I_j)}(\underline{x}, y) = P_{(\underline{x}, y)}(X_{(r)} \in I_j) - P_{\underline{x}}(X_{(r)} \in I_j) \tag{22}$$

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 Eq. (2). For  $j^* < j$ ,

$$P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) = \binom{j+r-1}{j} \binom{n-j+1+m-r}{n-j+1} \binom{n+m+1}{n+1}^{-1} \tag{23}$$

Similarly, for  $j^* > j$ ,  $n$  is replaced in Eq. (2) by  $n + 1$  but  $j$  is unchanged,

$$P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*}) = \binom{j+r-2}{j-1} \binom{n-j+2+m-r}{n-j+2} \binom{n+m+1}{n+1}^{-1} \tag{24}$$

For  $j^* = j$ , we get

$$\begin{aligned}
 P(X_{(r)} \in I_j | y \in I_j) &= P(X_{(r)} \in I_j \cup \tilde{I}_{j+1} | y \in I_j) \\
 &= P(X_{(r)} \in \tilde{I}_j | y \in I_{j^*}) + P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*})
 \end{aligned}$$

It is quite easy to proof [1] that  $SC_{P(X_{(r)} \in I_j)}(\underline{x}, y) > 0$  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 SC for the event that  $X_{(r)}^m \in I_j$ , 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

$$\begin{aligned} SC_{P(X_{(r)} \in I_j)}(\underline{x}, y) &= P(X_{(r)} \in \tilde{I}_{j+1} | y \in I_{j^*}) - P(X_{(r)} \in I_j) \\ &= P(X_{(r)} \in I_j) \left[ \frac{(r-1)(n+1) - jm}{j(n+1+m)} \right] \end{aligned} \quad (25)$$

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

$$\begin{aligned} SC_{P(X_{(r)} \in I_j)}(\underline{x}, y) &= P(X_{(r)} \in I_j | y \in I_{j^*}) - P(X_{(r)} \in I_j) \\ &= P(X_{(r)} \in I_j) \left[ \frac{m(j-1) - r(n+1)}{(n-j+2)(n+m+1)} \right] \end{aligned} \quad (26)$$

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 will be illustrated in Example 1 in Sect. 5.4. 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

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

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.

### 5.3 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  $x_1, \dots, x_{k-1}$  fixed and let  $x_k, \dots, x_n$  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  $x_1, \dots, x_{k-2}$  fixed and let  $x_{k-1}, \dots, x_n$  go to infinity then  $[P, \bar{P}](X_{(r)} > z)$  will increase. For  $c = 0$  the minimum fraction of the contaminated values in the contaminated sample that can cause  $b(\lambda_a; \underline{x}, [P, \bar{P}](X_{(r)} > z)) > 0$ , is

$$\lambda_0^*([P, \bar{P}](X_{(r)} > z), \underline{x}(\delta, n, \dots, j_{n-k+2})) = \frac{n-k+2}{n} \quad (27)$$

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  $X_{(r)} \in I_j$ , for  $c = 0$ , is

$$\begin{aligned} \lambda_0^*(P(X_{(r)} \in I_i), (\underline{x}, y_{j_1}, \dots, y_{j_l})) &= \inf\{\lambda_b | b(\lambda_b; \underline{x}, P(X_{(r)} \in I_j)) > 0\} \\ \lambda_0^*(P(X_{(r)} \in I_i), (\underline{x}, y_{j_1})) &= \frac{1}{n+1} \end{aligned} \tag{28}$$

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 1 in Sect. 5.4. 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  $\frac{1}{n+1}$ . We have only considered the NPI-

BP for  $c = 0$  here. In Example 1, we will also illustrate NPI-BP for  $c > 0$ .

### 5.4 Example

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

**Example 1** 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 1 presents the NPI-SC for the lower and upper probabilities for the event  $X_{(r)} \geq 1$ , for  $m = 5$  and  $r = 1, \dots, 5$ . These inferences are not affected at all 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  $r = 5$ .

To illustrate the NPI-BP, we consider the data set  $\underline{x}$  and the case with  $m = 5$  and interest in event  $X_{(r)} \geq 1$ . Table 2 presents the NPI-SC for the NPI lower and upper probabilities for  $X_{(r)} \geq 1$  for the values  $r = 1, \dots, 5$ , in the case where we keep  $x_1, \dots, x_{8-l}$  and we added  $\delta = 100$  to  $x_{9-l}, \dots, x_8$  for  $l = 1, \dots, 8$ . 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

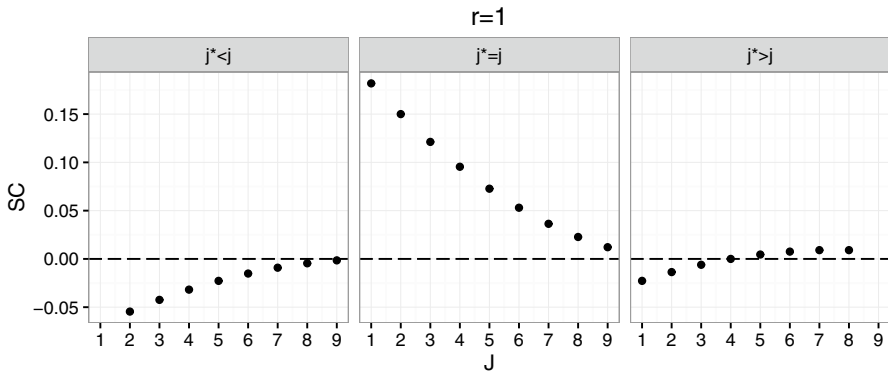
**Table 1** SC $_{P(X_{(r)} \geq 1)}(\underline{x}(j, \delta))$  for  $m = 5$  and  $\delta \geq 8$

	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
SC $_P$	0.09790	0.17405	0.16317	0.09324	0.02720
SC $_{\bar{P}}$	0.16317	0.19580	0.13054	0.05439	0.01166

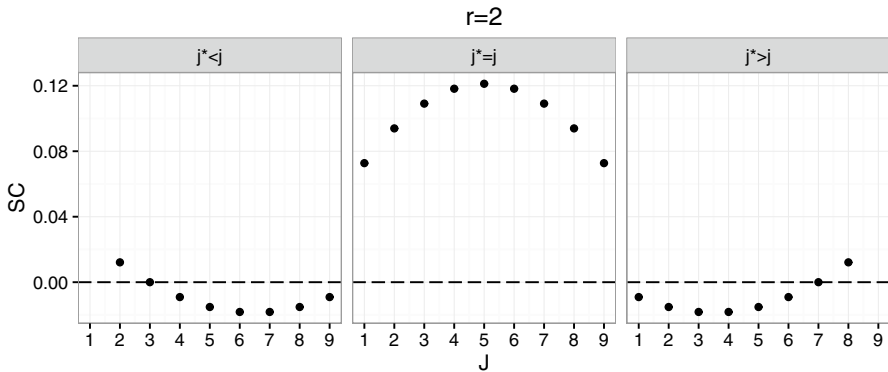
**Table 2**  $SC_{P(X_{(r)} \geq 1)}(\underline{x}(9-l, \dots, 8, 100))$  for  $m = 5$

$l$	$r = 1$		$r = 2$		$r = 3$		$r = 4$		$r = 5$	
	$SC_{\underline{p}}$	$SC_{\overline{p}}$	$SC_{\underline{p}}$	$SC_{\overline{p}}$	$SC_{\underline{p}}$	$SC_{\overline{p}}$	$SC_{\underline{p}}$	$SC_{\overline{p}}$	$SC_{\underline{p}}$	$SC_{\overline{p}}$
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

probabilities for  $X_{(r)} \geq 1$  is decreasing. If we chose  $c = 0.15$ , then the maximum NPI-BP for the event  $X_{(r)} \geq 1$  occurs for  $r = 5$ , whereas the minimum NPI-BP occurs for  $r = 2$ . The higher the breakdown point of an inference, the more robust it is.  $\lambda_0^*(P(X_{(1)} \geq 1), \underline{x}(2, \dots, 8, 100)) = \lambda_0^*(P(X_{(3)} \geq 1), \underline{x}(2, \dots, 8, 100)) = \frac{7}{8}$  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 1, whereas the upper probability for  $X_{(4)} \geq 1$  does not breakdown. For  $r = 5$  the inferences did not breakdown.



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



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

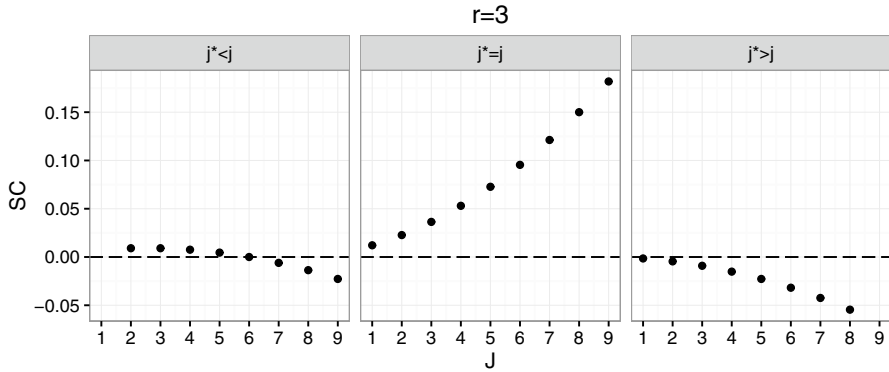


Fig. 3  $SC_{P(X_{(3)} \in I_j)}(\underline{x}, y)$  for  $n = 8$  and  $m = 3$

Figures 1, 2 and 3 illustrate the NPI-SC for the event  $X_{(r)} \in I_j$ , for  $r = 1, 2, 3$ ,  $j = 1, \dots, 9$  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 such as  $SC_{P(X_{(1)} \in I_9)}(\underline{x}, y) = SC_{P(X_{(3)} \in I_1)}(\underline{x}, y)$ . For all  $r$ , the NPI-SC for  $X_{(r)} \in I_j$  is unimodal in  $j$ .

To illustrate the  $c$ -breakdown point  $\lambda_c^*$  for the event  $X_{(r)} \in I_j$ , we choose  $c = 0.05$  and plot the absolute value of  $SC_{P(X_{(r)} \in I_j)}(\underline{x}, y_1, \dots, y_l)$  as function of  $l$ , where  $l$  is the

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

$r$	$l$	$j = 1$	$l$	$j = 2$	$l$	$j = 3$	$l$	$j = 4$	$l$	$j = 5$
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 4  $SC_{P(X_{(r)} \in I_j)}(\underline{x}, y_1, \dots, y_l)$  for  $m = 3$

$r$	$l$	$j = 6$	$l$	$j = 7$	$l$	$j = 8$	$l$	$j = 9$
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

number of the contaminated values that have been added to the data set of size  $n = 8$ . These are given in Tables 3 and 4 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  $j = 1$ ,  $\lambda_{0.05}^*(P(X_{(1)} \in I_1), (\underline{x}, y_9, y_{10}, y_{11})) = \frac{3}{11} = 0.2727$  and for  $j = 2$ ,  $\lambda_{0.05}^*(P(X_{(1)} \in I_2), (\underline{x}, y_9, \dots, y_{13})) = \frac{5}{13} = 0.3846$ . These tables present 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  $j = 2$  it is  $\frac{5}{13} = 0.3846$  and for  $j = 6$  it 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  $SC_{P(X_{(2)} \in I_j)}(\underline{x}, y_1, \dots, y_8) < 0.05$ . For  $r = 3$ , and  $j = 4$ ,  $\lambda_{0.05}^*(P(X_{(3)} \in I_j), (\underline{x}, y_9, \dots, y_{16})) = \frac{8}{16} = 0.5$ , whereas as  $j$  increases the NPI-BP decreases, such that for  $j = 8, 9$ ,  $\lambda_{0.05}^*(P(X_{(3)} \in I_j), (\underline{x}, y_9)) = 1/9$ .

### 6 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.

#### 6.1 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 odd-valued  $m$ . We consider the NPI-SC for the lower and upper probabilities for the event  $M_m < z$ . We wish to examine the effect on  $[P, \bar{P}](M_m < z)$  of adding a contaminant  $\delta$  to one of the observations  $x_j$  with  $j = 1, \dots, n$ . Let  $z \in I_k = (x_{k-1}, x_k)$ , if we add  $\delta$  to  $x_j$  this becomes  $\tilde{x}_j = x_j + \delta$ , where  $\delta \in \mathbb{R}$ . The NPI-SC for event  $M_m < z$  is

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

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



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  $\delta$  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  $\delta$  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  $x_j < z_1$

$$SC_{\underline{P}(M_m \in (z_1, z_2))}(\underline{x}(j, \delta)) = \underline{P}_{\underline{x}(j, \delta)}(M_m \in (z_1, z_2)) - \underline{P}_{\underline{x}}(M_m \in (z_1, z_2)) = \begin{cases} 0 & \text{if } \tilde{x}_l < z_1 \\ P(M_m \in I_k) & \text{if } \tilde{x}_l \in (z_1, z_2) \\ P(M_m \in I_k) - P(M_m \in I_{d-1}) & \text{if } \tilde{x}_l > z_2 \end{cases} \quad (29)$$

$$SC_{\overline{P}(M_m \in (z_1, z_2))}(\underline{x}(j, \delta)) = \overline{P}_{\underline{x}(j, \delta)}(M_m \in (z_1, z_2)) - \overline{P}_{\underline{x}}(M_m \in (z_1, z_2)) = \begin{cases} 0 & \text{if } \tilde{x}_l < z_1 \\ P(M_m \in I_{k-1}) & \text{if } \tilde{x}_l \in (z_1, z_2) \\ P(M_m \in I_{k-1}) - P(M_m \in I_d) & \text{if } \tilde{x}_l > z_2 \end{cases} \quad (30)$$

Secondly, if  $x_j > z_2$

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

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

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

$$SC_{\underline{P}(M_m \in (z_1, z_2))}(\underline{x}(j, \delta)) = \begin{cases} 0 & \text{if } \tilde{x}_l \in (z_1, z_2) \\ -P(M_m \in I_{d-1}) & \text{if } \tilde{x}_l > z_2 \\ -P(M_m \in I_{k+1}) & \text{if } \tilde{x}_l < z_1 \end{cases} \quad (31)$$

$$\text{SC}_{\bar{P}(M_m \in (z_1, z_2))}(\underline{x}(j, \delta)) = \begin{cases} 0 & \text{if } \tilde{x}_l \in (z_1, z_2) \\ -P(M_m \in I_d) & \text{if } \tilde{x}_l > z_2 \\ -P(M_m \in I_k) & \text{if } \tilde{x}_l < z_1 \end{cases} \quad (32)$$

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, which leads to more complicated analysis due to the definition of  $M_m$  as the overage of two observations. For study of the robustness of  $M_m$  for even-valued  $m$  we refer to the PhD thesis of [1].

The  $c$ -breakdown point for the NPI lower and upper probabilities for the event  $M_m > z$  and  $M_m > (z_1, z_2)$ , where  $z, z_2 \in I_k$  and  $m$  is odd, are similar as presented in Sect. 5, if we replace  $X_{(r)}$  by  $M_m$  in Eq. (27). The NPI lower and upper probabilities for such an event depend only on the number of observations that are greater than  $z$  or within  $(z_1, z_2)$ , so in the sample of  $n$  observations, only  $n - k + 2$  or more outliers can cause these probabilities to change.

## 6.2 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 a single contamination in the data [22]. We investigate the robustness of inferences involving the mean of the  $m$  future observations. The lower and upper bounds for the mean of the  $m$  future observations given the ordering  $O_i$ , as given in Eqs. (7) and (8), depend on the value of  $s_j^i$ . The NPI-SC for the lower and upper

bounds of the  $\mu_m^i$ , if  $x_j$  becomes  $x_j + \delta = \tilde{x}_l$ , for  $\delta > 0$  and  $l > j$  or for  $\delta < 0$  and  $l < j$ , are

$$\text{SC}_{\underline{\mu}_m^i}(\underline{x}(j, \delta)) = \frac{1}{m} \left[ \sum_{k=j}^l s_{k+1}^i [\tilde{x}_k - x_k] \right] \quad (33)$$

$$\text{SC}_{\overline{\mu}_m^i}(\underline{x}(j, \delta)) = \frac{1}{m} \left[ \sum_{k=j}^l s_k^i [\tilde{x}_k - x_k] \right] \quad (34)$$

As a special case, if  $l = j$ , i.e.  $x_j + \delta$  did not shift from its rank among the observations, so  $x_{j-1} < x_j + \delta < x_{j+1}$ , then the NPI-SC for  $\underline{\mu}_m^i$  and  $\overline{\mu}_m^i$  are  $\frac{1}{m} s_{j+1}^i \delta$  and  $\frac{1}{m} s_j^i \delta$ , respectively. If the value of  $s_j^i = s_{j+1}^i = 0$  then there is no influence at all on the lower and upper  $\mu_m^i$ , whereas if  $s_j^i = m$  and  $s_{j+1}^i = m$  then NPI-SC of the lower or the upper

bound for  $\mu_m^i$ , will exceed any bound for  $\delta$  large or small enough. The NPI-SC for  $\mu_m \geq z$ , if  $x_j$  becomes  $x_j + \delta = \tilde{x}_j$  and  $\delta \in \mathbb{R}$ , is

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

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

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

and

$$\begin{aligned}
 SC_{\overline{P}(\mu_m \in (z_1, z_2))}(\underline{x}(j, \delta)) &= \overline{P}_{\underline{x}(j, \delta)}(\mu_m \in (z_1, z_2)) - \overline{P}_{\underline{x}}(\mu_m \in (z_1, z_2)) \\
 &= \binom{n+m}{n} \sum_{i=1} P(O_i) \left[ 1 \left\{ (\underline{\mu}_m^i(\underline{x}(j, \delta)), \overline{\mu}_m^i(\underline{x}(j, \delta))) \cap (z_1, z_2) \neq \emptyset \right\} \right. \\
 &\quad \left. - 1 \left\{ (\underline{\mu}_m^i(\underline{x}), \overline{\mu}_m^i(\underline{x})) \cap (z_1, z_2) \neq \emptyset \right\} \right]
 \end{aligned} \tag{35}$$

These NPI-SC will be illustrated in Example 2 in Sect. 6.3.

The  $c$ -breakdown points of the lower and upper bounds of  $\mu_m^i$ , are  $\frac{1}{n}$  for  $s_{l+1}^i \neq 0$  and  $s_l^i \neq 0$  respectively. This is because if we hold  $x_1, \dots, x_{n-1}$  fixed and let  $x_n$  go to infinity then  $\mu_m^i$  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 assumed bounds for the data observations

$L < x_1 < \dots < x_n < R$ . So  $\lambda_c^*(\mu_m^i, \underline{x}(\delta, j_1, \dots, j_l))$  may not be equal to  $\frac{1}{n}$ . This will be illustrated in Example 2 in Sect. 6.3.

### 6.3 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 2. The 0-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  $\tilde{x}_j$ , 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  $O_l$ , is not equal to zero. Example 2 illustrates the NPI-SC and NPI-BP for inferences involving the mean and the median of the  $m$  future observations.

**Example 2** 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  $\delta$  to  $x_2 = -7$  and  $\delta \in \mathbb{R}$ . When we consider the mean of the 3 future observations, we set  $x_0 = L = -17$  and  $x_9 = R = 18$  as bounds for the observations.

Figure 4 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  $LP$ , and the NPI upper probability by  $UP$ . 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 contaminated value, as  $\delta$  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 shows that for  $\delta < 7$  and  $\delta > 19$ , the inferences involving the mean are more sensitive than the inference involving

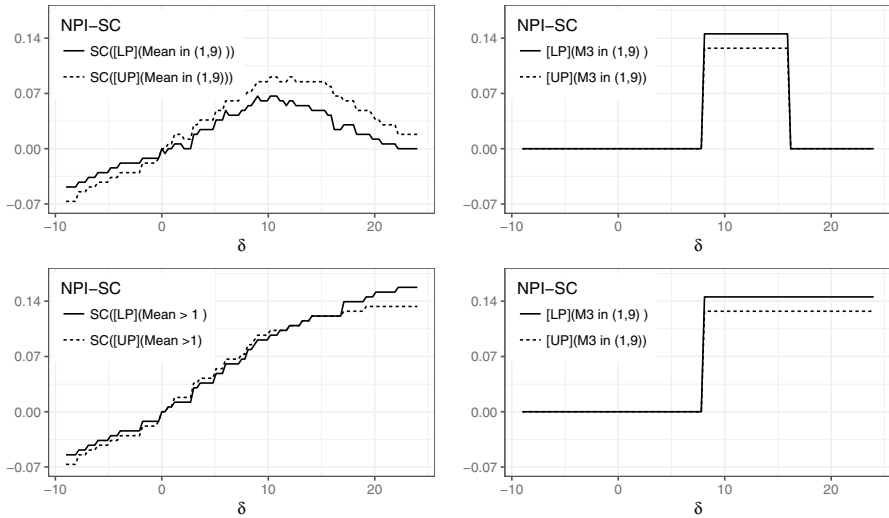


Fig. 4  $SC_j(x(2, \delta))$  for the events  $\mu_3 \geq 1, \mu_3 \in (1, 9), M_3 \geq 1$  and  $M_3 \in (1, 9)$

the median. In contrast, for  $8 < \delta \leq 15.3$  the inferences involving the mean are more robust.

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  $x_8 + \delta_8$ , then  $x_8$  and  $x_7$  by  $x_8 + \delta_8$  and  $x_7 + \delta_7$ , and so on, until all observations have been contaminated by  $\{\delta_1, \dots, \delta_8\} = \{18.5, 17.5, 11, 9.5, 7, 5.5, 3, 1\}$ . Figure 5 shows the NPI-SC for the lower and upper probabilities for  $\mu_3 \geq 1$  and  $M_3 \geq 1$ , as functions of the number of the observations that have been contaminated by adding different values of  $\delta$  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 observations 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.

### 7 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 by [8, 9].

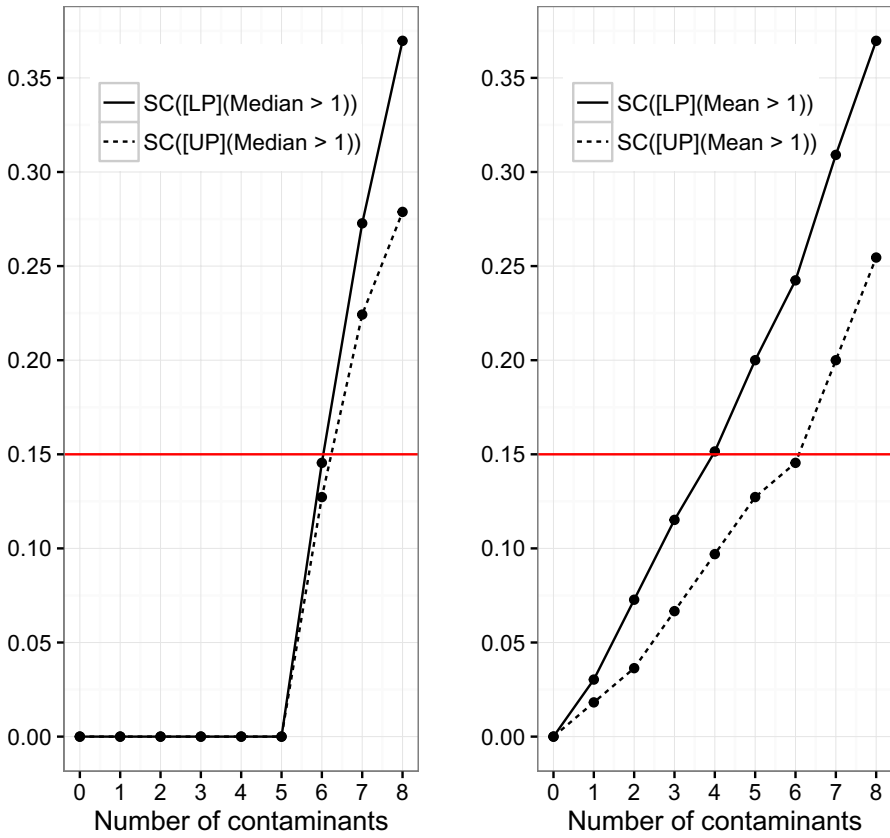


Fig. 5  $SC_j(\alpha(j_1, \dots, j_r, \delta_{j_1}, \dots, \delta_{j_r}))$  for the events  $\mu_3 \geq 1$  and  $M_3 \geq 1$

### 7.1 Pairwise Comparisons

We investigate the robustness of one of the applications of NPI for future order statistics for statistical inference problems, as presented by [9]. Suppose that we have two independent groups of real-valued observations,  $X$  and  $Y$  and their ordered observed values are  $x_1 < x_2 < \dots < x_{n_x}$  and  $y_1 < y_2 < \dots < y_{n_y}$ . For ease of notation, let  $x_0 = y_0 = -\infty$  and  $x_{n_x+1} = y_{n_y+1} = \infty$ . Let  $I_{j_x}^x = (x_{j_x-1}, x_{j_x})$  and  $I_{j_y}^y = (y_{j_y-1}, y_{j_y})$ . We focus attention on  $m \geq 1$  future observations from each group,  $X_{n_x+i}$  and  $Y_{n_y+i}$  for  $i = 1, \dots, m$ . We wish to compare the  $r$ -th future order statistics from these two groups by considering the event  $X_{(r)} < Y_{(r)}$ , for which the NPI lower and upper probabilities, based on the  $A_{(n_x)}$  and  $A_{(n_y)}$  assumptions per group, are derived by

**Table 5**  $SC_r(x(2, \delta))$  for  $m = 3$

$\delta$	$\mu_3 \geq 1$		$M_3 \geq 1$		$\mu_3 \in (1, 9)$		$M_3 \in (1, 9)$	
	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
-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

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

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  $\tilde{y}_j$ , are

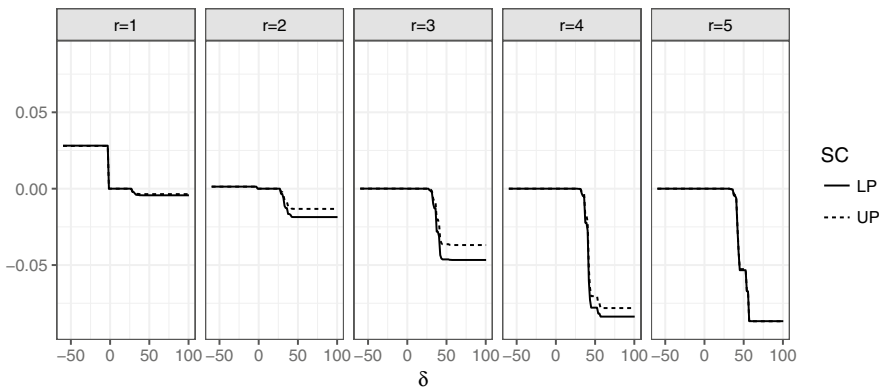
**Table 6** Rats weight gain data

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	

$$\begin{aligned}
 & SC_{\underline{P}(X_{(r)} < Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ P(Y_{(r)} \in I_{l+1}^y) \times P(X_{(r)} \in I_d^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ P(Y_{(r)} \in I_{l+1}^y) \times [P(X_{(r)} \in I_d^x)P(X_{(r)} \in I_{d+1}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases} \\
 & SC_{\bar{P}(X_{(r)} < Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ P(Y_{(r)} \in I_l^y) \times P(X_{(r)} \in I_{d+1}^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ P(Y_{(r)} \in I_l^y) \times [P(X_{(r)} \in I_{d+1}^x)P(X_{(r)} \in I_{d+2}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases}
 \end{aligned}$$

The NPI-BP for such NPI pairwise comparisons, for  $c = 0$ , is

$$\lambda_c^*([\underline{P}, \bar{P}](X_{(r)} < Y_{(r)}), \underline{x}(j_1, \dots, j_l, \delta)) = \frac{1}{n} \text{ if } x_n < y_j \text{ and } x_n + \delta > y_j$$



**Fig. 6**  $SC_{\underline{P}(X_{(r)} < Y_{(r)})}(\underline{x}(2, \delta))$  for  $m = 5$



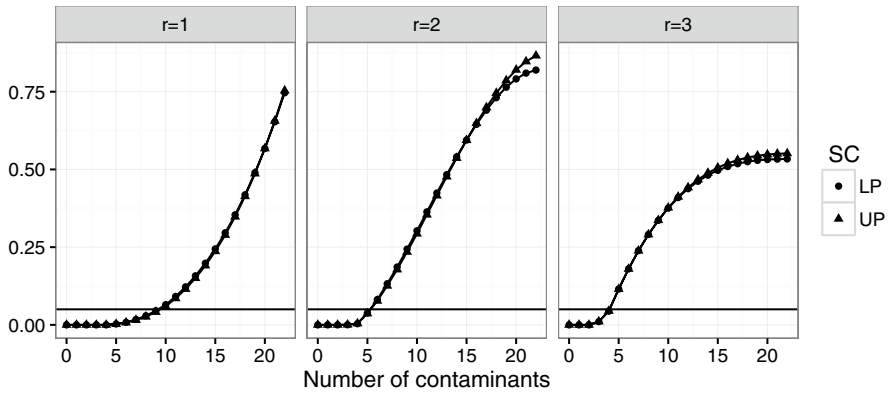


Fig. 7  $|SC_{P(X_{(r)} < Y_{(r)})}(\underline{x}(23 - l, \dots, 22, 100))|$  for  $m = 3$

**Table 7** The absolute value of  $SC_{P(X_{(r)} < Y_{(r)})}(\underline{x}(j_1, \dots, j_l, 100))$  for  $m = 3$  and  $n = 22$

$l$	$r = 1$		$r = 2$		$r = 3$			
	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$l$	$SC_{\underline{P}}$	$SC_{\overline{P}}$	$l$	$SC_{\underline{P}}$	$SC_{\overline{P}}$
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

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 Example 3 we illustrate the NPI-SC and NPI-BP for such NPI-pairwise comparisons.

**Example 3** To illustrate the NPI-SC and NPI-BP for pairwise comparisons, we consider the data set of a study of the effect of ozone environment on rats growth [12, 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 6. We use this dataset to illustrate the effect of replacing  $x_2 = -14.7$  by  $-14.7 + \delta$ , for  $\delta$  from  $-50$  to  $100$ , on the pairwise comparisons based on the events  $X_{(r)} < Y_{(r)}$ ,  $r = 1, \dots, m$ , and  $m = 5$ .

Figure 6 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  $\delta$  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  $r = 2, \dots, 5$ , are close to zero. For  $-1.5 \leq \delta \leq 27$  the  $SC_{P(X_{(r)} < Y_{(r)})}(x(2, \delta)) = 0$  and  $SC_{\bar{P}(X_{(r)} < Y_{(r)})}(x(2, \delta)) = 0$  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  $r$  increases. The inferences involving  $r = 4$  and  $5$  have large NPI-SC when the value  $x_2 + \delta$  exceeds all the  $Y$  observations.

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  $X$ . This is shown in Fig. 7 and Table 7. Figure 7 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.

## 7.2 NPI for Test Reproducibility

Reproducibility of statistical hypothesis tests is an issue of major importance in applied statistics: if the test were repeated, would the same conclusion, rejection or non-rejection of the null hypothesis, be reached? 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 section, the robustness of the NPI method for reproducibility of statistical tests is presented for two basic tests using order statistics, namely a one sample quantile test and a two sample precedence test. For these inferences, NPI for future order statistics [9] is used, as briefly reviewed in Sect. 2. We assume that the first, actual experiment led to ordered real-valued observations  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ . As we consider an imaginary repeat of this experiment, we use NPI for  $n = m$  future ordered observations [8].

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. Most of the literature on robustness [18, 25] 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  $\delta$  to one of the observations could change the original test decision from rejecting to not rejecting the null hypotheses, or the other way around.

### 7.2.1 Quantile Test

The quantile test is a basic nonparametric test for the value of a population quantile [14]. Let  $\kappa_p$  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  $X_i, i = 1, \dots, n$ , 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  $\kappa_p^0$ . Under  $H_0$ ,  $\kappa_p^0$  is the  $100 \times p$ -th quantile of the distribution function of the  $X_i$ , so  $P(X_i \leq \kappa_p^0 | H_0) = p$ . For a given data set  $x_1, \dots, x_n$ , define the test statistic of the one-sided quantile test as the number of observations  $X_i$  in the sample that are less than or equal to  $\kappa_p^0$ , denoted by  $k = \sum_{i=1}^n \mathbf{1}\{x_i \leq \kappa_p^0\}$ . 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 the value  $r$  derived using the Binomial distribution [14].

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  $H_0$ . This will occur if  $X_{(r)} > \kappa_p^0$ . The NPI lower and upper reproducibility probabilities for this event, as function of  $k \leq r - 1$ , are

$$\begin{aligned} \underline{\text{RP}}(k) &= \underline{P}\left(X_{(r)} > \kappa_p^0 | k\right) = \sum_{j=1}^{n+1} \mathbf{1}\left\{x_{j-1} > \kappa_p^0\right\} P(X_{(r)} \in I_j) \\ \overline{\text{RP}}(k) &= \overline{P}\left(X_{(r)} > \kappa_p^0 | k\right) = \sum_{j=1}^{n+1} \mathbf{1}\left\{x_j > \kappa_p^0\right\} P(X_{(r)} \in I_j) \end{aligned}$$

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  $\kappa_p^0$ . 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 function of  $k \geq r$ , are

$$\begin{aligned} \underline{\text{RP}}(k) &= \underline{P}\left(X_{(r)} \leq \kappa_p^0 | k\right) = \sum_{j=1}^{n+1} \mathbf{1}\left\{x_j \leq \kappa_p^0\right\} P(X_{(r)} \in I_j) \\ \overline{\text{RP}}(k) &= \overline{P}\left(X_{(r)} \leq \kappa_p^0 | k\right) = \sum_{j=1}^{n+1} \mathbf{1}\left\{x_{j-1} \leq \kappa_p^0\right\} P(X_{(r)} \in I_j) \end{aligned}$$

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  $\kappa_p^0$ . Reproducibility of result of rejection of  $H_0$  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  $\delta$  to any of the data observations, say  $x_j$  which becomes  $\tilde{x}_j$ , on the reproducibility of the quantile test for that event is

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

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

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

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

The NPI-SC for the NPI upper probability

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

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  $\kappa_p^0$ .

**Table 8**  $SC_{RP(k)}(\underline{x}(j, \delta))$  for  $n = 15$

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

**Example 4** Suppose that the original 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  $\kappa_{0.75}^0$  against the alternative hypothesis that this third quartile is greater than  $\kappa_{0.75}^0$ , 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  $x_{(8)} > \kappa_{0.75}^0$  and  $H_0$  is not rejected if  $x_{(8)} < \kappa_{0.75}^0$ . 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  $X_{(8)} > \kappa_{0.75}^0$  and  $X_{(8)} < \kappa_{0.75}^0$ , respectively. Let  $\bar{k}$  denote to the number of observations that are less than  $\kappa_p^0$  based on the contaminated sample  $\underline{x}(j, \delta)$ .

Table 8 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  $SC_{[\text{RP}(k), \overline{\text{RP}}(k)]}(\underline{x}(j, \delta)) = 0$ . The size of the effect for such an inference increases as the value of  $k$  increases.

Table 8 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  $X_{(8)} < \kappa_{0.75}^0$ . The RP for  $X_{(8)} < \kappa_{0.75}^0$  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.

## 7.2.2 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 [26], 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.

We consider the classical scenario with two independent samples. Let  $X_{(1)} < X_{(2)} < \dots < X_{(n_x)}$  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  $\lambda_x$ . Similarly, let  $Y_{(1)} < Y_{(2)} < \dots < Y_{(n_y)}$  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  $\lambda_y$ . 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 significance  $\alpha$ , one determines the value  $k$  (which therefore is a function of  $\alpha$  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 discussed by [8].

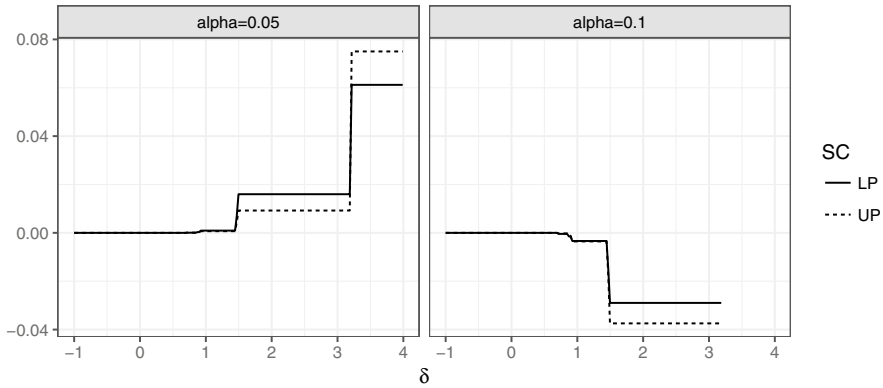
The NPI approach for reproducibility of this two-sample precedence test considers again the same test scenario applied to future order statistics, and derives the NPI lower and upper probabilities for the event that the same overall test conclusion will be derived, given the data from the original test. This involves the NPI approach for inference on the  $r$ -th future order statistic  $Y_{(r)}$  out of  $n_y$  future observations based on the data from the  $Y$  population, and similarly for the  $k$ -th future order statistic  $X_{(k)}$  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. 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  $T$  [8].

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 Sect. 2, 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

$$\begin{aligned} \underline{\text{RP}} &= \underline{P}(X_{(k)} < Y_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x)} < y_{(j_y-1)}\} P(X_{(k)} \in I_{j_x}^x) P(Y_{(r)} \in I_{j_y}^y) \\ \overline{\text{RP}} &= \overline{P}(X_{(k)} < Y_{(r)}) = \sum_{j_x=1}^{n_x+1} \sum_{j_y=1}^{n_y+1} \mathbf{1}\{x_{(j_x-1)} < y_{(j_y)}\} P(X_{(k)} \in I_{j_x}^x) P(Y_{(r)} \in I_{j_y}^y) \end{aligned}$$

**Table 9** Times to insulating fluid breakdown

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	*	*	*	*



**Fig. 8**  $SC_{RP}(\alpha(2, \delta))$  for  $X_{(10)} > Y_{(6)}$  and  $X_{(9)} < Y_{(6)}$

If  $H_0$  is not rejected in the original test, so  $x_{(k)} > y_{(r)}$ , then

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

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

We consider NPI-SC for the NPI-RP of the precedence test. 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  $\delta$  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

$$\begin{aligned}
 & SC_{\underline{P}(X_{(k)} < Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ P(Y_{(r)} \in I_{l+1}^y) \times P(X_{(k)} \in I_d^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ P(Y_{(r)} \in I_{l+1}^y) \times [P(X_{(k)} \in I_d^x) + P(X_{(k)} \in I_{d+1}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases} \\
 & SC_{\overline{P}(X_{(k)} < Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ P(Y_{(r)} \in I_l^y) \times P(X_{(k)} \in I_{d+1}^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ P(Y_{(r)} \in I_l^y) \times [P(X_{(k)} \in I_{d+1}^x) + P(X_{(k)} \in I_{d+2}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases}
 \end{aligned}$$

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  $\delta$  to  $y_j$  in group  $Y$ , so  $y_j$  becomes  $\tilde{y}_l$ , on  $\underline{RP}$  and  $\overline{RP}$  are

$$\begin{aligned}
 & SC_{\underline{P}(X_{(k)} > Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ -P(Y_{(r)} \in I_l^y) \times P(X_{(k)} \in I_{d+1}^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ -P(Y_{(r)} \in I_l^y) \times [P(X_{(k)} \in I_{d+1}^x) + P(X_{(k)} \in I_{d+2}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases} \\
 & SC_{\overline{P}(X_{(k)} > Y_{(r)})}(y(j, \delta)) \\
 &= \begin{cases} 0 & \text{if } y_j < x_d \text{ and } \tilde{y}_l < x_d \\ -P(Y_{(r)} \in I_l^y) \times P(X_{(k)} \in I_d^x) & \text{if } y_j < x_d \text{ and } x_d < \tilde{y}_l \\ -P(Y_{(r)} \in I_{l+1}^y) \times [P(X_{(k)} \in I_d^x) + P(X_{(k)} \in I_{d+1}^x)] & \text{if } y_j < x_d < x_{d+1} \\ & \text{and } x_d < x_{d+1} < \tilde{y}_l \end{cases}
 \end{aligned}$$

**Example 5** To illustrate the NPI-SC for the NPI-RP for the precedence test, we consider a data set presented by [27] consisting of six groups of times (in minutes) to breakdown of an insulating fluid subjected to different levels of voltage. These times are presented in Table 9. These data were also used by [8]. 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 5 are only known to exceed 3.83.

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, 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  $\overline{\text{RP}} = \overline{P}(X_{(10)} > Y_{(6)}) = 0.8669$  for  $\alpha = 0.05$ . While for  $\alpha = 0.1$ ,  $\underline{\text{RP}} = \underline{P}(X_{(9)} < Y_{(6)}) = 0.3029$  and  $\overline{\text{RP}} = \overline{P}(X_{(9)} < Y_{(6)}) = 0.7079$ . Let us now assume that we added an increasing value of  $\delta$  to  $x_2 = 0.64$ , then we examine its effect on the NPI lower and upper reproducibility probabilities.

The left plot of Fig. 8 presents the NPI-SC for the NPI-RP for the event that  $X_{(10)} > Y_{(6)}$ , as a function of  $\delta$ . The results clearly illustrate 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 taken 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  $y_6$ .

The right plot of Fig. 8 presents the NPI-SC for the lower and upper reproducibility probabilities for the event  $X_{(9)} < Y_{(6)}$ , as a function of  $\delta$ . Increasing the values of  $\delta$  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  $\delta$ , as if  $x_2 + \delta$  exceeds  $y_6$  that will change the original test conclusion and also the reproducibility probability.

## 8 Concluding Remarks

This paper 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 predictive 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 sections of this paper. We investigated robustness of the mean and the median for the  $m$  future observations. The robustness of the inference involving the median of the  $m$  future observations is a step function, whereas for the mean is continuously changing function, 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 for NPI, and also, of course, robustness of other NPI methods. Further details, examples and discussion of the tests presented in this paper are given in the PhD thesis of [1].

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Alqifari HN (2017) Nonparametric predictive inference for future order statistics. Ph.D. thesis, Durham University, Durham. [www.npi-statistics.com](http://www.npi-statistics.com)
2. Arts GRJ, Coolen FPA (2008) Two nonparametric predictive control charts. *J Stat Theory Pract* 2:499–512
3. Arts GRJ, Coolen FPA, van der Laan P (2004) Nonparametric predictive inference in statistical process control. *Qual Technol Quant Manag* 1:201–216
4. Augustin T, Coolen FPA, de Cooman G, Troffaes MCM (eds) (2014) Introduction to imprecise probabilities. Wiley, Chichester
5. Augustin T, Coolen FPA (2004) Nonparametric predictive inference and interval probability. *J Stat Plan Inference* 124:251–272
6. Bellio R, Ventura L (2005) An introduction to robust estimation with R functions. In: Proceedings of 1st international work. University of Udine and University of Padova
7. Coolen FPA (2006) On nonparametric predictive inference and objective Bayesianism. *J Log Lang Inf* 15:21–47
8. Coolen FPA, Alqifari HN (2017) Nonparametric predictive inference for reproducibility of two basic tests based on order statistics. *REVSTAT: Stat J* 16:167–185
9. Coolen FPA, Coolen-Maturi T, Alqifari HN (2017) Nonparametric predictive inference for future order statistics. *J Commun Stat Theory Methods* 47:2527–2548
10. Coolen FPA, Maturi T (2010) Nonparametric predictive inference for order statistics of future observations. In: Borgelt C (ed) Combining soft computing and statistical methods in data analysis. Advances in intelligent and soft computing, vol 77. Springer, Berlin, pp 97–104
11. De Waal T, Pannekoek J, Scholtus S (2011) Handbook of statistical data editing and imputation. Wiley, New Jersey
12. Desu M, Raghavarao D (2004) Nonparametric statistical methods for complete and censored data. Chapman and Hall, Boca Raton
13. de Finetti B (1974) Theory of probability. Wiley, London
14. Gibbons JD, Chakraborti S (2011) Nonparametric statistical inference, 5th edn. Springer, Berlin
15. Hampel FR (1968) Contribution to the theory of robust estimation. Ph.D. thesis, University of California, Berkeley
16. Hampel FR (1974) The influence curve and its role in robust estimation. *J Am Stat Assoc* 69:383–393
17. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics: the approach based on influence functions. Wiley, New York
18. He X, Simpson DG, Portnoy SL (1990) Breakdown robustness of tests. *J Am Stat Assoc* 85:446–452
19. Hill BM (1968) Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *J Am Stat Assoc* 63:677–691
20. Hodges JL (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. University of California Press, Berkeley, pp 163–186
21. Huber PJ (1996) Robust statistical procedures. SIAM, Philadelphia
22. Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley, Hoboken
23. Jurečková J, Sen PK (1996) Robust statistical procedures: asymptotics and interrelations. Wiley, New York
24. Lawless J, Fredette M (2005) Frequentist prediction intervals and predictive distributions. *Biometrika* 92:529–542
25. Markatou M, He X (1994) Bounded influence and high breakdown point testing procedures in linear models. *J Am Stat Assoc* 89:543–549
26. Nelson LS (1963) Tables for a precedence life test. *Technometrics* 124:491–499
27. Nelson W (1982) Applied life data analysis. Wiley, Hoboken
28. Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading