

# Nonparametric predictive inference in statistical process control

**Citation for published version (APA):**

Arts, G. R. J., Coolen, F. P. A., & Laan, van der, P. (2000). *Nonparametric predictive inference in statistical process control*. (SPOR-Report : reports in statistics, probability and operations research; Vol. 200016). Technische Universiteit Eindhoven.

**Document status and date:**

Published: 01/01/2000

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

SPOR-Report 2000-16

# **Nonparametric predictive inference in statistical process control**

G.R.J. Arts  
F.P.A. Coolen  
P. van der Laan

# Nonparametric predictive inference in statistical process control

G.R.J. Arts, F.P.A. Coolen

University of Durham

Department of Mathematical Sciences

Science Laboratories, South Road

Durham, DH1 3LE, England

P. van der Laan

Eindhoven University of Technology

Department of Mathematics and Computing Science

P.O. Box 513, 5600 MB Eindhoven, The Netherlands

## Abstract

New methods for statistical process control are presented, where the inferences have a nonparametric predictive nature. We consider several problems in process control in terms of uncertainties about future observable random quantities, and we develop inferences for these random quantities based on data available in the form of a reference set. We use Hill's assumption  $A_{(n)}$ , which enables predictive inference while adding only few assumptions to the data observed.

**Keywords:** Control limits; Exchangeability; Extrema chart; Imprecise probabilities; Precedence probability; Run length.

## 1 Introduction

Statistical process control is a topic of major importance, for a good introduction see, for example, Montgomery [17]. Most research contributions and applications have been based on classical frequentist statistical theory, using parametric models such as the normal distribution. However, there has also been ample attention to this topic from nonparametric [4] and from Bayesian perspectives [16, 20].

A wide variety of statistical methods have been suggested and used in this area, some justified by the contextual relevance, others mostly by available statistical theory. The problem at the heart

of statistical process control, namely whether or not to stop a process, has a predictive nature. One would want to interrupt a process if confidence in the quality of the next item, or several next items, is low. This lack of confidence might occur due to reasons in the process environment, such as observing a problem with a machine, or it might be based on product measurements, so that measurements on some recent products might cause you to worry about the quality of the next item. Statistical process control methods address this second reason for possible doubt in the quality of products, so the problem is to draw inferences about a future observation, or several future observations, on the basis of past observations. A Bayesian predictive approach [11] seems attractive, yet this requires use of a parametric model and a prior distribution. If good reasons exist to use a specific parametric model to describe the random nature of the measurements, then such an approach could be useful. However, often such information is not available, and simple parametric models might not fit well with the observations, suggesting the use of nonparametric methods.

Hill [12, 13] introduced a method for statistical inference that is predictive and nonparametric, and therefore seems to provide a natural framework for statistical process control. In this approach, only few assumptions are added to data observed to draw inferences on future observations. We briefly introduce this method in section 2, calling it ‘low structure inference’. The main aim of this paper is to present some analyses of problems in process control using low structure inference, which we do in section 3. Some results are identical to those derived by others, from different perspectives, in which cases we feel that our derivations and interpretations might be attractive. In section 4 we briefly comment on the important question of when and how to use our method.

It should be emphasized that any nonparametric method for statistical inference relies on the availability of data, and throughout the paper we assume that a reference set is present. One would expect this to be a fair representation of observations taken when the process is functioning satisfactorily, which is often called ‘in-control’. Even statistical process control methods that assume parametric probability distributions, or make use of normal approximations, require input of information to reflect the in-control situation. For example, Montgomery [17] suggests to base control charts on estimates of the mean and standard deviation for the characteristic of interest which are based on a sample taken when the process is considered to be in-control. Montgomery mentions the possibility to update the reference set regularly, yet it is difficult to provide useful guidelines for this.

Prior to having the observations of the reference set available, it is important to distinguish between conditional (on these observations) and unconditional inferences. The classical frequentist approach provides inferences in the form of probabilities prior to getting a reference set, which change to confidence statements after the reference set has become available. For those without detailed understanding of foundations of statistics, this might confuse the interpretation of the results, and this is exactly where the  $A_{(n)}$ -based approach makes life less complicated, while remaining a sound and

valid approach to statistical inference, as Hill [12, 13, 14] showed that the inferences based on  $A_{(n)}$  are valid probability statements, also justified from a Bayesian perspective when treated as posterior inferences after an actual reference set is available.

## 2 Low structure inference

Suppose that we have  $n$  real-valued observations forming a reference set, ordered as  $x_1 < x_2 < \dots < x_n$ . Throughout this paper, we call the quantities of interest ‘observations’, but they might for example also be sample statistics computed from subgroups obtained from a process. For ease of notation, define  $x_0 = -\infty$  and  $x_{n+1} = \infty$ , if the quantities of interest are known to be positive one can define  $x_0 = 0$  instead. In this paper, we assume that with probability one ties do not occur. Our approach can easily be generalized to the case that ties can occur [13], but notation would become less clear. The  $n$  observations partition the real line into  $n + 1$  intervals  $I_j = (x_{j-1}, x_j)$  for  $j = 1, \dots, n + 1$ . The assumption  $A_{(n)}$  is that a random quantity  $X_{n+i}$ , for  $i \geq 1$ , falls into any such interval  $I_j$  with equal probability,

$$P(X_{n+i} \in I_j) = \frac{1}{n+1} \quad \text{for } i \geq 1.$$

This assumption implies that the rank of  $X_{n+i}$  amongst the observed  $x_1 < x_2 < \dots < x_n$  has equal probability to be any value in  $\{1, \dots, n + 1\}$ .

The situation becomes more complex when we consider  $m$  random quantities, say  $X_{n+1}, \dots, X_{n+m}$ , with  $m \geq 2$ , under the assumptions  $A_{(n)}, \dots, A_{(n+m-1)}$ , because these random quantities are not conditionally independent given the first  $n$  observations. For example, inferences on two future values,  $X_{n+1}$  and  $X_{n+2}$ , based on the first  $n$  observations and the assumptions  $A_{(n)}$  and  $A_{(n+1)}$ , use joint probabilities (with  $j, k = 1, \dots, n + 1$ )

$$\begin{aligned} P(X_{n+1} \in I_j, X_{n+2} \in I_k) &= P(X_{n+2} \in I_k | X_{n+1} \in I_j)P(X_{n+1} \in I_j) \\ &= \begin{cases} \frac{1}{(n+1)(n+2)} & \text{for } k \neq j \\ \frac{2}{(n+1)(n+2)} & \text{for } k = j. \end{cases} \end{aligned}$$

Now consider the general situation of  $m$  future observations,  $X_{(n+1)}, \dots, X_{(n+m)}$ . Define the random quantity (for  $j = 1, \dots, n + 1$ )

$$S_j = \#\{X_{n+i} \in I_j, i = 1, \dots, m\}.$$

All inferences about the  $m$  future observations, based on the assumptions  $A_{(n)}, \dots, A_{(n+m-1)}$ , use the probabilities for events  $\{S_j = s_j\}$ , with  $s_j$  non-negative integers with  $\sum_{j=1}^{n+1} s_j = m$ . These probabilities are equal to [10, 12]

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

Note that these probabilities do not depend on the values of  $s_j$ , and can be interpreted in terms of equally likely orderings of all  $n + m$  variables, after the first  $n$  have been observed and ordered. One particular case is the event that all  $m$  future observations will fall in the same specified interval  $I_l$ , which has probability  $\binom{n+m}{n}^{-1}$ .

Throughout this paper, one should keep in mind that  $A_{(n)}$ -based results actually use probabilities on future observations, for one particular given reference set, and that indeed these are valid probability statements. The main difference with classical frequentist methods is that in such methods similar results are confidence statements with the randomness of selecting one particular reference set taken into account. Indeed, the work of Hill has shown that our results are valid and coherent probabilities that can be used in case of complete absence of information of underlying probability distributions or processes beyond the  $n$  available observations in the reference set. We therefore indeed have predictive probabilities, and manipulating these is often more straightforward than dealing with confidence statements when following the classical approach. We also believe that interpretation of probabilities is easier than interpretation of confidence results. Nevertheless, as combinatorial computations for our results often coincide with those for classical frequentist methods, as indicated for some results in section 3, our results could also be interpreted from classical frequentist perspectives, in the same way as  $A_{(n)}$  itself can be interpreted purely frequentistically [11].

Based on only these assumptions  $A_{(n)}, \dots, A_{(n+m-1)}$ , and observations  $x_1 < x_2 < \dots < x_n$ , inferences for any  $X_{n+i}$  ( $i = 1, \dots, m$ ) are the same, no matter which future observation is actually considered. Although not strictly fully implied by these assumptions, this is close in nature to finite *exchangeability* of the  $X_{n+i}$  ( $i = 1, \dots, m$ ), which we will actually also assume throughout this paper. Exchangeability [9] will explicitly be used to derive some results on run length in section 3.

Throughout this paper, whenever we are interested in inferences regarding  $m$  future observations, based on a reference set of size  $n$ , we will make the assumptions  $A_{(n)}, \dots, A_{(n+m-1)}$ , as required for our approach, and exchangeability of the  $m$  random quantities  $X_{n+i}$  ( $i = 1, \dots, m$ ), without mentioning this repeatedly.

The  $A_{(n)}$ -based probabilities can be interpreted both in a frequentist and a Bayesian subjective manner [14]. The strength of these inferences can best be indicated by citing the final paragraph of Hill [13]: ‘Let me conclude by observing that  $A_{(n)}$  is supported by all of the serious approaches to statistical inference. It is Bayesian, fiducial, and even a confidence/tolerance procedure. It is simple, coherent and plausible. It can even be argued, I believe, that  $A_{(n)}$  constitutes the fundamental solution to the problem of induction’.

### 3 Inference for statistical process control

In this section we suggest some possibilities for use of  $A_{(n)}$ -based inference for statistical process control. First, we focus on inference related to extremes of future observations, discussing possible use in control charts. Secondly, we consider the run length corresponding to control limits, where a nice feature is that our results for samples of size one are identical to those derived by Willemain and Runger [22] via a more classical approach. We also present a generalisation of these results by using control charts for extrema of samples of size more than one. Thereafter, we present possibilities for inferences related to changes in a process, and consider the suggested extrema charts a bit further. Finally, we briefly mention two more results that could be of use in process control situations, namely the possibility of comparing two independent samples and the use of precedence probabilities.

#### 3.1 Extremes of future observations

It is clear that under the assumption  $A_{(n)}$  a single future observation will be smaller than all  $n$  observations so far with probability  $1/(n+1)$ , with the same probability for it to exceed all  $n$  observed values. Consider two future observations, let us make the appropriate  $A_{(n)}$  assumptions as presented in section 2. For ease of notation write  $p_{j,k} = P(X_{n+1} \in I_j, X_{n+2} \in I_k)$ . If interest is in the probability of the event that the minimum of the two future observations will fall in an interval  $I_l$ , this is easily calculated by summing all joint probabilities of the form above for combinations that lead to this event, so (for  $l = 1, \dots, n+1$ )

$$\begin{aligned} P(\min(X_{n+1}, X_{n+2}) \in I_l) &= \sum_{k>l} p_{l,k} + \sum_{j>l} p_{j,l} + p_{l,l} \\ &= \frac{2(n+1-l)}{(n+1)(n+2)} + \frac{2}{(n+1)(n+2)} \\ &= \frac{2}{n+1} - \frac{2l}{(n+1)(n+2)}. \end{aligned}$$

Probabilities for maxima are derived similarly, leading to

$$P(\max(X_{n+1}, X_{n+2}) \in I_l) = \frac{2l}{(n+1)(n+2)}.$$

For  $m \geq 2$  future observations we have

$$P(\min_{1 \leq i \leq m} X_{n+i} \in I_l) = P(\min_{1 \leq i \leq m} X_{n+i} > x_{l-1}) - P(\min_{1 \leq i \leq m} X_{n+i} > x_l).$$

From the ordering interpretation of all  $n+m$  variables, it is clear that (for  $j = 1, \dots, n$ )

$$P(\min_{1 \leq i \leq m} X_{n+i} > x_j) = \frac{\binom{m+n-j}{m}}{\binom{m+n}{m}}.$$

This follows from a direct counting argument: the  $m$  future observations should be to the right of  $x_j$ , in irrelevant order with the  $n-j$  observed values to the right of  $x_j$ . If  $m$  is fixed, and  $j, n$  tend to infinity

with  $\frac{j}{n} \rightarrow \theta$ , then the limit of  $P(\min_{1 \leq i \leq m} X_{n+i} > x_j)$  equals  $(1 - \theta)^m$ , which is intuitively logical, and in agreement with results in other concepts of statistical inference. Indeed, one could consider this  $\theta$  as a parameter related to a very large or infinite hypothetical population, but we do not need such an assumption as this  $\theta$  does not play any fundamental role in our inferences, and is unlikely to be of much practical interest. Remark here, that if we were actually interested in a number of future observations that becomes extremely large, so tends to infinity, then our necessary  $A_{(n)}$  assumptions and exchangeability become similar to the assumption of infinite exchangeability, which indeed allows representation as conditionally independent and identically distributed (*ciid*) random quantities with a probability distribution with parameter  $\theta$  [9].

The above result leads to (for  $l = 1, \dots, n + 1$ , with the convention  $\binom{m-1}{m} = 0$ )

$$\begin{aligned} P(\min_{1 \leq i \leq m} X_{n+i} \in I_l) &= \frac{\binom{m+n-l+1}{m} - \binom{m+n-l}{m}}{\binom{m+n}{m}} \\ &= \frac{n!(m+n-l)!m}{(m+n)!(n-l+1)!}. \end{aligned}$$

Similar results for the maximum value of  $m$  future observations are based on (for  $j = 1, \dots, n$ )

$$P(\max_{1 \leq i \leq m} X_{n+i} < x_j) = \frac{\binom{m+j-1}{m}}{\binom{m+n}{m}}.$$

This equality is again justified by a straightforward counting argument, and a similar limit result as above is achieved if  $m$  is fixed, and  $j, n$  tend to infinity with  $\frac{j}{n} \rightarrow \theta$ , in which case the limit of  $P(\max_{1 \leq i \leq m} X_{n+i} < x_j)$  equals  $\theta^m$ . We now derive (for  $l = 1, \dots, n + 1$ )

$$\begin{aligned} P(\max_{1 \leq i \leq m} X_{n+i} \in I_l) &= P(\max_{1 \leq i \leq m} X_{n+i} < x_l) - P(\max_{1 \leq i \leq m} X_{n+i} < x_{l-1}) \\ &= \frac{\binom{m+l-1}{m} - \binom{m+l-2}{m}}{\binom{m+n}{m}} \\ &= \frac{n!(m+l-2)!m}{(m+n)!(l-1)!}. \end{aligned}$$

It is important to consider how these probability distributions for the extremes of  $m$  future observations, over the intervals created by the observations  $x_1, \dots, x_n$  in the reference set, can be used in contexts relevant to statistical process control. A minimum (maximum) of  $m$  future observations which is large (small) relative to the reference set indicates that there may have been a change in the process. For example, a shift in the process or increased process variation may result in a larger (smaller) minimum (maximum). The probabilities above can be used to define control limits based on the reference set, to be used in relation to the minimum or maximum value of  $m$  observations. We consider this, focussing on implications for run lengths, in the next section, after we have looked at run lengths when considering control limits for single observations (the case  $m = 1$ ).



### 3.2 Run length

A topic of interest in statistical process control that clearly has a predictive nature is study of the random run length of future observations related to a control chart. As before, we assume to have a reference set of  $n$  observations, and use  $A_{(n)}$ -based inferences. Willemain and Runger [22] study run lengths from a similar perspective, calling the reference set an ‘empirical reference distribution’. In this section, we derive probabilistic results for the run length if the control limits are set at values identical to two earlier observations within the reference set.

Let us assume that we set control limits at the values of two observations in the reference set  $\{x_1, \dots, x_n\}$ , namely such that an interval is formed within the control limits as the union of  $b$  of the  $n + 1$  intervals  $I_j$  created by the data. Denote the area between these control limits as  $B$ , which is therefore the union of  $b$  intervals  $I_j$ . Remark that  $b = n - 1$  would provide the widest possible finite difference between the control limits, with  $B = (x_1, x_n)$ . When considering one-sided control charts, we could also consider the case  $b = n$ , leading to  $B$  being unbounded from either below or above. Also, we could imagine control charts made up of  $b$  intervals  $I_j$  that are not all consecutive intervals, in which case our results would also hold, but this is unlikely to be of practical interest.

We define the *run length*  $R$  as the number of consecutive observations until an observation falls outside this area  $B$ , this includes the observation outside  $B$ . According to this definition,  $R$  can take any positive integer value. A run length  $r \geq 1$  occurs if  $X_{n+i} \in B$  for all  $i = 1, \dots, r - 1$ , and  $X_{n+r} \notin B$ . The assumptions  $A_{(n)}, \dots, A_{(n+r-1)}$  lead to (for  $r = 1, 2, \dots$ )

$$\begin{aligned} P(R = r) &= \frac{b}{n+1} \times \frac{b+1}{n+2} \times \dots \times \frac{b+r-2}{n+r-1} \times \frac{n+1-b}{n+r} \\ &= \frac{(b+r-2)!}{(b-1)!} \times \frac{n!}{(n+r)!} \times (n-b+1). \end{aligned}$$

It is interesting to notice that Willemain and Runger [22] present a frequentist method which, although fundamentally different to our method, leads to the same probability distribution for  $R$ . We should remark that, although Willemain and Runger define the run length as the number of observations between alarms, they actually use the same definition as ours, including the observation outside  $B$  in the run length.

The frequentist argument by Willemain and Runger [22] which also leads to the above probability distribution for  $R$  is as follows. For specified  $B$ , assume that future observations are independent and identically distributed, conditional on the parameter  $p$  which is the probability for each of these observations to fall in  $B$ . Clearly, the run length distribution, for given  $p$ , is geometric, so  $P(R = r|p) = p^{r-1}(1-p)$ , for  $r = 1, 2, \dots$ . Due to the definition of  $p$ , if we consider  $p$  before the reference set has actually become available, then a standard result from the theory of order statistics [18] is that  $p$  is a random variable with a beta distribution, with probability density function proportional to  $p^{b-1}(1-p)^{n-b}$ . To derive the unconditional distribution for  $R$  one integrates  $p$  out, so

effectively taking the expectation of  $P(R = r|p)$  with regard to the beta distribution for  $p$ . It should be remarked that, in nature, this is close to a Bayesian argument if the beta distribution for  $p$  were the posterior distribution for  $p$  after observing the reference set, which would be feasible when using an improper (non-informative) prior distribution for  $p$ .

Although our result here coincides with Willemain and Runger's result, there are a few important differences. Effectively, the link to the Bayesian argument indicated above can be understood from infinite exchangeability, necessary to derive a probability distribution for  $R$  over all positive integers, using De Finetti's representation theorem [9] which then allows representation in terms of *ciid* random quantities, with a distribution function for the model parameter,  $p$  in this case. Willemain and Runger's result effectively uses the same concepts, although it is not explicitly stated that a probability distribution for a parameter is used. However, to get meaningful probabilities, they must necessarily make the assumption of independent and identically distributed random quantities. The fact that we get the same numerical values for run length probabilities is interesting, and perhaps a slight coincidence, yet pleasing at the same time as there are now three fundamentally different approaches leading to the same probability distribution for run length in such cases. We strongly feel that our approach is intuitively attractive, as it appears to be more direct than both the frequentist approach of Willemain and Runger, and the Bayesian approach described above, both of which use a parametric distribution and integrate out the parameter.

It might be useful, for example to decide on choices of the size  $n$  of the reference set and the value of  $b$ , to consider the expected run length and probabilities for minimal run length. Of course, these are again the same as those derived by Willemain and Runger [22], namely  $E(R) = \frac{n}{n-b}$  and  $P(R > r) = \frac{(b+r-1)!n!}{(b-1)!(n+r)!}$  for  $1 \leq b \leq n - 1$  and  $r \geq 1$ . For example, if we would want to set  $n$  and  $b$  such that  $E(R) \geq 50$ , these must satisfy  $b \geq \frac{49n}{50}$ , so  $n = 100$  and  $b = 98$  would be a possible choice.

We briefly illustrate our probability distribution for the run length in two examples.

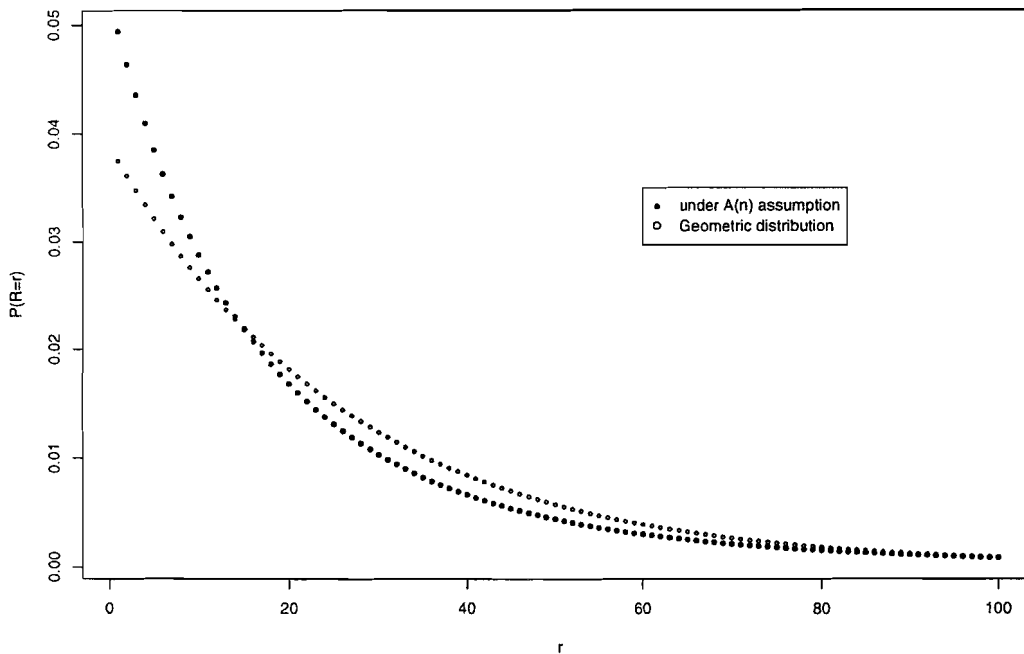
### Examples run length distribution

Suppose we have  $n = 4$  and  $b = 2$ , then the probabilities for  $R$  being equal to, say, 1,2,3,4,5, are, in that order, 0.6, 0.2, 0.0857, 0.0429, 0.0238, with probability 0.0476 for  $R$  exceeding 5. The expected run length is equal to 2, which in practice is probably considered to be too small. If the process is actually in control, one would not want to get a false alarm too often. Even if one sets  $b = 3$ , the expected run length of 4, is very small. To prevent false alarms happening too often, one needs more information on the production process, which implies that the reference set has to be larger.

For a more interesting example assume that the reference set contains 80 observations, and that we are interested in  $B$  created by taking the control limits equal to the 2nd and 79th ranked observation, so  $b = 77$ . The distribution of the run length is plotted in Figure 3.1, together with the geometric

distribution with  $p = 77/80$ . This geometric distribution corresponds to the situation that the control limits are set such that the probability that a future observation falls between the control limits is equal to  $p = 77/80$ , assuming that future observations are independent. The distribution of the run length under the  $A_{(n)}$ -assumption has a longer tail than the geometric distribution. Under the  $A_{(n)}$ -assumption,  $P(R > 100) = 0.037$ , while for the geometric distribution,  $P(R > 100|p = 77/80) = 0.022$ . The expected run length in both cases is equal to  $80/3$ , implying that if the process is in-control, one would expect a false alarm for 3 out of 80 future observations. The expected run length could be increased to 80 by taking the control limits equal to the 1st and 80th observation.

Figure 3.1: Distribution of the run length



On studying the random run length, we do not necessarily have to restrict attention to control limits set at values of observations in the reference set. However, if such limits were set at different values, precise probabilistic results are not possible following our approach, as we do not add any assumptions on the location of the probability masses  $1/(n + 1)$  for the next observation within each interval  $I_j$ . We could, however, easily derive bounds for probabilities  $P(R = r)$  for any given control limits, by optimising this probability with regard to the probability masses per interval  $I_j$ , for the intervals that contain the control limits. We do not pursue this further as we regard the above precise result for  $P(R = r)$  to be of main interest.

It is clear that to get a large expected run length, one needs a large reference set. This is probably not very realistic, especially if one wants to control a low volume production process. Therefore, next we consider the possibility of using a control chart based on the extreme(s) of  $m$  future observations, as introduced in section 3.1. Even if one uses the extreme(s) of only two future observations, one can reduce the size of the reference set considerably, still obtaining a reasonably large expected run length.

In section 3.1 we derived the probability that the minimum (maximum) of  $m$  future observations falls in an interval created by two subsequent observations from the reference set. Suppose that we want to detect a change in the production process, using a one-sided control chart which gives an out of control signal if the minimum of a sample of size  $m$  is larger than  $x_b$  ( $b = 1, \dots, n$ ), the  $b^{\text{th}}$  ordered observation from the reference set. From section 3.1 we can find that the probability that the first sample gives an out-of-control signal is equal to

$$P(\min(X_{n+1}, \dots, X_{n+m}) > x_b) = \frac{\binom{m+n-b}{m}}{\binom{m+n}{m}}.$$

For this scenario, this is the probability that the run length  $R$  is equal to 1. Let  $M_l$  be the minimum of the  $l^{\text{th}}$  sample, i.e.  $M_l = \min(X_{n+(l-1)m+1}, \dots, X_{n+lm})$ , for  $l \geq 1$ . Then the distribution of the run length  $R$  is given by (for  $r = 1, 2, \dots$ )

$$\begin{aligned} P(R = r) &= P(M_1, \dots, M_{r-1} \leq x_b, M_r > x_b) \\ &= P(M_1, \dots, M_{r-2} \leq x_b, M_r > x_b) - P(M_1, \dots, M_{r-2} \leq x_b, M_{r-1}, M_r > x_b) \\ &= P(M_1, \dots, M_{r-2} \leq x_b, M_{r-1} > x_b) - P(M_1, \dots, M_{r-2} \leq x_b, M_{r-1}, M_r > x_b) \\ &= P(M_1, \dots, M_{r-3} \leq x_b, M_{r-2} > x_b) - 2P(M_1, \dots, M_{r-3} \leq x_b, M_{r-2}, M_{r-1} > x_b) \\ &\quad + P(M_1, \dots, M_{r-3} \leq x_b, M_{r-2}, M_{r-1}, M_r > x_b) \\ &\quad \vdots \\ &= \sum_{l=1}^r \binom{r-1}{l-1} (-1)^l P(M_1, \dots, M_l > x_b) \\ &= \sum_{l=1}^r \binom{r-1}{l-1} (-1)^l P(X_{n+1}, X_{n+2}, \dots, X_{n+lm} > x_b) \\ &= \sum_{l=1}^r \binom{r-1}{l-1} (-1)^l \frac{n!}{(n-b)!} \frac{(n-b+(l+1)m)!}{(n+(l+1)m)!}. \end{aligned}$$

It should be noted that some of the equalities are justified by the exchangeability assumption for the future observations, which implies for example that

$$P(M_1, \dots, M_{r-2} \leq x_b, M_r > x_b) = P(M_1, \dots, M_{r-2} \leq x_b, M_{r-1} > x_b).$$

We can also again look at the run length from the perspective of classical frequentist theory, as was used by Willemain and Runger [22] for the case  $m = 1$ , by first conditioning the run length on  $p$ , the parameter representing the probability that a future observation exceeds  $x_b$ . Then the distribution of

the run length is given by

$$P(R = r|p) = (1 - (1 - p)^m)^{r-1} (1 - p)^m = \sum_{l=1}^r \binom{r-1}{l-1} (-1)^l (1 - p)^{m+l(r-1)}.$$

To derive the distribution of  $R$ , we note once more that  $p$  is here regarded as a random variable with a beta distribution, and we can integrate  $p$  out, which gives us the same result as above. In this situation it is fairly easy to calculate the average run length by first taking the expectation conditional on  $p$ , which is equal to  $1/(1 - p)^m$ , and then taking the expectation with regard to  $p$ , resulting in

$$ARL = ER = \frac{(n - b - m)!}{(n - m)!} \frac{n!}{(n - b)!},$$

where  $ARL$  is the abbreviation of *Average Run Length*. This is, of course, also the expected run length under the  $A_{(n)}$ -assumption. Note that  $ARL$  reduces to  $n/(n - b)$  if  $m = 1$ , the expected run length from the previous section.

For example, to obtain an  $ARL$  of about 250 if  $m = 2$ , one needs a reference set of size  $n = 23$ , setting the upper control limit at  $x_{21}$ . It is clear that using the minimum of two observations instead of just one observation, the size of the reference set can be reduced considerably, still obtaining a reasonably large  $ARL$  if a process is in-control.

### 3.3 Changes in a process

In this section we analyse the control chart based on extremes, which we call the *extrema chart*, in case the process changes. We look at the distribution of the run length and the expected run length, corresponding to the results of sections 3.1 and 3.2 based on all the required  $A_{(n)}$ -assumptions and finite exchangeability. Note that the extrema chart is especially useful to detect shifts in the process. It can also detect changes in the process variation, but charts that use statistics representing the process variation in some way (for example the range of a sample) might be more efficient. We will compare the extrema chart with the  $\bar{X}$ -chart (which uses the normality assumption), both for a process which is actually known to be normal and for a process with a non-normal distribution.

First we consider a process which is controlled by looking at a single observation at regular points in time ( $m = 1$ ). Suppose that after observing the reference set of size  $n$ , a positive shift of size  $\theta$  occurs in the production process. In our predictive setting, we can denote observations after this shift has occurred by  $Y_{n+i}$ , which are assumed to be represented as future observations under the  $A_{(n)}$  assumption plus the shift  $\theta$ , so  $Y_{n+i} = X_{n+i} + \theta$ , where probabilities for  $X_{n+i}$  on the basis of the given reference set are as presented in section 2.

After the shift has occurred, an out-of-control signal is given if  $Y_{n+i} > x_b$ , or equivalently, if  $X_{n+i} > x_b - \theta$ . If  $x_b - \theta$  is equal to  $x_k$ , for a particular  $0 < k < b$ , than the  $ARL$  after the process has shifted is equal to  $n/(n - k)$ , rather than the in-control  $ARL$  of  $n/(n - b)$ . If we can only determine a

value  $k$  such that  $x_{k-1} < x_b - \theta < x_k$ , then we cannot determine the  $ARL$  exactly, but we know that it will be between  $n/(n-k+1)$  and  $n/(n-k)$ .

Willemain and Runger [22], whose results for  $m = 1$  are equal to ours as discussed in section 3.2, present the  $ARL$ s for different shifts based on a particular reference set from a standard normal distribution. The size of the reference set is 500, and the upper control limit is set at  $x_{498}$ , resulting in an  $ARL$  of 250 if the process is in-control. We denote this upper control limit by  $UCL_e$ . If one would actually know that, if the process is in-control, the observations are from the standard normal distribution, then the  $ARL$  would also be equal to 250 when using an upper control limit equal to 2.652, the 99.6-percentile of that distribution. We shall denote this latter upper control limit by  $UCL_{\bar{X}} = 2.652$ , as it is the upper control limit of an  $\bar{X}$ -chart with the size of the samples equal to  $m = 1$ .

The  $ARL$  using  $UCL_e = x_{498}$  can be compared to the  $ARL$  using  $UCL_{\bar{X}} = 2.652$ . Of course, for a particular reference set the value  $x_{498}$  will not be equal to 2.652, so the  $ARL$ s using  $UCL_e = x_{498}$  and  $UCL_{\bar{X}} = 2.652$  will differ. However, as Willemain and Runger [22] remark, provided that the control limits are set sufficiently far from the extremes, the sample percentiles converge to the population percentiles, and hence the  $ARL$  using the  $b^{\text{th}}$  ordered observation from the reference set as upper control limit will converge to the  $ARL$  using the upper control limit based on the normality assumption as the size of the reference set increases, as long as the in-control process is indeed perfectly modelled by the standard normal distribution. However, if this is not the case, then the control chart with the upper control limit based on the normality assumption will no longer assure a prescribed  $ARL$  if the process is in-control, in contrast to the control chart with the upper control limit equal to the  $b^{\text{th}}$  ordered observation from the reference set, which would still assure a prescribed  $ARL$ .

Next we consider the situation where samples of size  $m \geq 1$  are used to control the process. Using samples of size 2 or more enables us to set control limits using a smaller reference set, and results in control charts that detect a shift in the process quicker than a control chart where only one observation at a time is used. As before, we can obtain the  $ARL$  exactly if the process shifts upwards with a shift  $\theta = x_b - x_k$ . Let  $M'_l$  be the minimum of the  $l^{\text{th}}$  sample from the shifted process, so  $M'_l = \min(Y_{n+(l-1)m+1}, \dots, Y_{n+lm})$ . Representing observations from the shifted process again as  $Y_{n+i} = X_{n+i} + \theta$ , we obtain an out-of-control signal at the  $l^{\text{th}}$  sample if this is the first sample with  $M'_l > x_b$ , or equivalently, if sample  $l$  is the first with  $M_l > x_k$ , where  $M_l = \min(X_{n+(l-1)m+1}, \dots, X_{n+lm})$ . This implies an  $ARL$  of

$$\frac{(n-k-m)!}{(n-m)!} \frac{n!}{(n-k)!}.$$

If we can only determine  $k$  such that  $x_b - x_k < \theta < x_b - x_{k-1}$ , then we have

$$\frac{(n-k-m)!}{(n-m)!} \frac{n!}{(n-k)!} < ARL < \frac{(n-k-m+1)!}{(n-m)!} \frac{n!}{(n-k+1)!}.$$

For a given shift  $\theta$ , the value of  $k$  for which  $x_b - x_k < \theta < x_b - x_{k-1}$  will of course depend on the actual reference set. Nevertheless, for a given reference set and shift  $\theta$ , so with  $n$  and  $k$  known, these bounds clearly indicate how the *ARL* depends on the chosen sample size  $m$ .

### Example average run length

To illustrate how the *ARL* depends on  $m$ , suppose that we have a production process which is controlled by using a one-sided extrema chart to detect a positive shift in a characteristic of interest. We consider sample sizes  $m = 1, \dots, 5$ , and we use a reference set of size  $n = 100$ . We actually simulated the reference set from the  $t(4)$ -distribution, but do not make any further use of this fact, which fits nicely with practical situations where no further knowledge of underlying distributions is available. For each value of  $m$ , a value of  $b$  is chosen such that the *ARL* for the in-control process, denoted by  $ARL(0)$ , is as close to 100 as possible, with  $ARL(0)$  as derived in section 3.2. Table 3.1 gives, for  $m = 1, \dots, 5$ , the values of  $b$ ,  $x_b$ , and the corresponding  $ARL(0)$ . In addition, for shifts  $\theta = 0.7$  and  $\theta = 1.4$  (these are shifts of about half and one standard deviation),  $x_{k-1}$  and  $x_k$  are given such that  $x_{k-1} < x_b - \theta < x_k$ , together with the corresponding value of  $k$ , and finally the lower and upper bounds for the *ARL* are given for both these shifts.

Table 3.1: Bounds for *ARLs* for different sample sizes (example).

| $m$ | $b$ | $x_b$ | $ARL(0)$ | $x_{k-1}$ | $x_k$ | $k(0.7)$ | $x_{k-1}$ | $x_k$ | $k(1.4)$ | $ARL(0.7)$ | $ARL(1.4)$ |      |      |
|-----|-----|-------|----------|-----------|-------|----------|-----------|-------|----------|------------|------------|------|------|
| 1   | 99  | 4.35  | 100.0    | 3.26      | 3.80  | 96       | 2.91      | 3.26  | 95       | 20.0       | 25.0       | 16.7 | 20.0 |
| 2   | 90  | 2.05  | 110.0    | 1.33      | 1.35  | 79       | 0.64      | 0.68  | 67       | 21.4       | 23.6       | 8.8  | 9.4  |
| 3   | 78  | 1.33  | 105.0    | 0.59      | 0.64  | 66       | -0.08     | -0.06 | 42       | 24.7       | 27.0       | 5.0  | 5.2  |
| 4   | 67  | 0.68  | 95.8     | -0.04     | -0.00 | 46       | -0.81     | -0.60 | 23       | 11.5       | 12.4       | 2.7  | 2.9  |
| 5   | 59  | 0.42  | 110.5    | -0.33     | -0.26 | 35       | -1.01     | -0.98 | 20       | 8.4        | 9.1        | 2.9  | 3.1  |

Although the in-control *ARLs* are slightly different, this example gives reasonable insight into the decrease of the *ARLs* according to our method with the extrema chart when the sample size  $m$  increases, in case of shifts in the process.

This example is relevant for situations where, for example, once per hour a sample is taken from a production process and compared to the reference set. To make a choice for the sample size, one should balance the costs of sampling and the costs related to delayed detection of a shift. For example, if the costs related to sampling are hardly influenced by the sample size, so sampling 5 units at the same time is not much more expensive than sampling a single unit, then it is probably better to use a sample size of 5 instead of just 1, as it will detect a shift earlier. It is important, per value of  $m$ , to compare  $ARL(0)$  with  $ARL(0.7)$  and  $ARL(1.4)$ . For  $m = 5$ , if there is no shift the *ARL* is 110.5, indicating that on average one in 110.5 samples would lead to interruption of the process caused by

a false alarm. However, a shift of size  $\theta = 0.7$  would tend to be noticed after a much shorter run length, with expected value between 8.4 and 9.1. So, if each hour one sample would be taken of size 5, one would expect such a shift to be noticed about 9 hours after it occurred. If  $m = 1$  were used, it would take on average between 20 and 25 hours to notice a similar shift. The improvement achieved by larger  $m$ -values for a larger shift  $\theta = 1.4$  are even better in this example.

In practice, one could first get a reference set while the process is in-control, then determine lower and upper bounds for  $ARL$  for different shifts as we did in this example, and then decide how large  $m$  should be for the actual control chart to be used.

### 3.4 Further consideration of the extrema chart

In this section, we pay a bit more attention to the extrema chart by comparing it to the traditional  $\bar{X}$ -chart. Again, we consider a situation where we want to detect a positive shift using a one-sided control chart. In case of the extrema chart, an out-of-control signal is given when the minimum of a sample of size  $m$  is larger than the upper control limit. In case of the  $\bar{X}$ -chart, an out-of-control signal is given when the average of a sample of size  $m$  is larger than the upper control limit. The control limit in the extrema chart is set at  $x_b$ , the  $b^{\text{th}}$  ordered observation from the reference set. In the  $\bar{X}$ -chart, the upper control limit is set such that the  $ARL$  for an in-control process is the same as for the extrema chart, under the assumption that the process is normally distributed with perfect knowledge of the mean and variance of the in-control process.

First, we compare the two charts theoretically in the case that the actual production process is normally distributed, with standard deviation equal to 1. The mean is equal to 0 if the process is in control. For the extrema chart, the  $ARL$  if the process is in control does not depend on the actual distribution, but on the values of  $n$ ,  $b$  and  $m$ , and can be derived as in section 3.2. This  $ARL$  is the basis to set the control limit for the  $\bar{X}$ -chart. The control limit for the  $\bar{X}$ -chart is determined such that if the process is in control, and indeed is normally distributed, the  $ARL$  is the same as for the extrema chart. Now, given the values of  $n$ ,  $b$  and  $m$  for the extrema chart and the upper control limit for the  $\bar{X}$ -chart, we compare the  $ARL$  of the two charts if a shift occurs. As we actually know the distribution of the process, we can calculate the  $ARL$  for both charts exactly, in the frequentist context that  $ARL$  is the average run length over infinitely many possible reference sets, using the chart over an infinite time period in which the process is shifted upwards with  $\theta$ , but otherwise remains the same. Computation for the extrema chart was performed using that  $E(R|p) = 1/(1-p)^m$ , where  $p$  is the probability that an observation from the shifted process falls outside the control limit  $x_b$ , and then take the expectation over  $p$ , having a Beta( $b, n-b+1$ )-distribution. Note that the results for this normally distributed process are merely for comparison purposes, and the assumption of normality is not needed to actually use the extrema chart.



In Table 3.2 we give the *ARL* of the extrema chart for different values of  $n$ ,  $b$  and  $m$ , and for different shifts. We compare these with the *ARL* using the  $\bar{X}$ -chart, also using samples of size  $m$ .

Table 3.2: *ARLs* for a normally distributed process with shift  $\theta$ .

| $(n, b, m)$ | $\theta:$        | 0.0   | 0.1   | 0.2   | 0.5  | 1.0  | 1.5 | 2.0 |
|-------------|------------------|-------|-------|-------|------|------|-----|-----|
| (250,249,1) | ARL(extr)        | 250.0 | 184.0 | 136.9 | 59.9 | 18.5 | 7.2 | 3.5 |
|             | ARL( $\bar{X}$ ) | 250.0 | 186.8 | 140.8 | 63.7 | 20.3 | 8.0 | 3.9 |
| (25,23,2)   | ARL(extr)        | 300.0 | 191.1 | 124.8 | 40.1 | 9.4  | 3.5 | 1.9 |
|             | ARL( $\bar{X}$ ) | 300.0 | 197.6 | 132.5 | 44.6 | 10.3 | 3.6 | 1.8 |
| (25,20,3)   | ARL(extr)        | 230.0 | 141.0 | 89.1  | 26.9 | 6.2  | 2.5 | 1.5 |
|             | ARL( $\bar{X}$ ) | 230.0 | 140.2 | 87.9  | 25.4 | 5.4  | 2.0 | 1.3 |
| (25,15,5)   | ARL(extr)        | 210.8 | 117.7 | 68.9  | 18.0 | 4.0  | 1.8 | 1.2 |
|             | ARL( $\bar{X}$ ) | 210.8 | 112.6 | 62.9  | 14.3 | 2.8  | 1.3 | 1.0 |
| (20,12,5)   | ARL(extr)        | 276.9 | 147.8 | 83.2  | 19.9 | 4.2  | 1.8 | 1.2 |
|             | ARL( $\bar{X}$ ) | 276.9 | 145.1 | 79.5  | 17.1 | 3.1  | 1.3 | 1.0 |

From Table 3.2 we can observe that in most cases the  $\bar{X}$ -chart detects a positive shift earlier than the extrema chart, but the *ARLs* of both charts are not far apart. If the sample size  $m$  is small, the extrema chart detects a small positive shift a bit quicker than the  $\bar{X}$ -chart does, but again, the difference in *ARL* is not large. So, even in the ‘ideal’ situation that the normality assumption is true, the performance of the extrema chart is similar to the performance of the  $\bar{X}$ -chart.

Next we compare the extrema chart with the  $\bar{X}$ -chart in case the process follows a  $t(10)$ -distribution, which has heavier tails than the normal distribution. For the extrema chart we use the same value for  $b$  as before, resulting in the same *ARL* as before in case the process is in-control, as in that case the *ARL* does not depend on the actual distribution of the process. Again, as the distribution of the process is assumed to be known, we can calculate the *ARL* for the extrema chart exactly. For the  $\bar{X}$ -chart we used simulation, as the distribution of the mean of  $t$ -distributed variables is not available analytically. So the *ARL* results for the  $\bar{X}$ -chart are actually estimates based on simulation, we give the standard error for each estimate between brackets.

For the  $\bar{X}$ -chart we use two different upper control limits. The first one is the control limit that results in an estimated *ARL* which is the same as the *ARL* for the extrema chart if the process is in-control. The results using this control limit can be found in Table 3.3. Remark that in reality we would not be able to determine such a control limit for the  $\bar{X}$ -chart because the distribution is unknown. Therefore, we also used the control limit based on the assumption that the process is normally distributed. As the actual distribution is not normal, this will result in an *ARL* for an in-

control process different from the anticipated value (i.e. the  $ARL$  for the extrema chart). If the upper control limit for the  $\bar{X}$ -chart is based on the normality assumption, the  $ARL$  for an in-control process is much smaller than anticipated. For the different combinations of  $n$ ,  $b$  and  $m$  that we considered, the  $ARL$ s for the in-control process using the  $\bar{X}$ -chart are 158.7, 199.2, 173.5, 170.6 and 218.6. So the false alarm rate is much larger than anticipated if the actual distribution of the process is  $t(10)$  rather than normal.

Table 3.3:  $ARL$ s for a  $t(10)$ -distributed process with shift  $\theta$ .

| $(n, b, m)$ | $\theta:$        | 0.0    | 0.1    | 0.2    | 0.5    | 1.0    | 1.5     | 2.0     |
|-------------|------------------|--------|--------|--------|--------|--------|---------|---------|
| (250,249,1) | ARL(extr)        | 250.0  | 212.2  | 179.9  | 109.4  | 47.5   | 20.9    | 9.5     |
|             | ARL( $\bar{X}$ ) | 250.0  | 211.0  | 178.1  | 106.6  | 45.3   | 19.6    | 9.0     |
| (25,23,2)   | ARL(extr)        | 300.0  | 216.6  | 156.9  | 61.5   | 15.0   | 4.9     | 2.3     |
|             | ARL( $\bar{X}$ ) | 299.9  | 223.6  | 164.9  | 73.2   | 20.1   | 6.7     | 2.9     |
|             |                  | (7.33) | (4.72) | (2.98) | (0.88) | (0.12) | (0.02)  | (0.01)  |
| (25,20,3)   | ARL(extr)        | 230.0  | 150.1  | 99.4   | 32.0   | 7.2    | 2.8     | 1.6     |
|             | ARL( $\bar{X}$ ) | 230.0  | 163.2  | 113.2  | 40.0   | 9.0    | 3.0     | 1.6     |
|             |                  | (4.92) | (2.94) | (1.70) | (0.35) | (0.04) | (<0.01) | (<0.01) |
| (25,15,5)   | ARL(extr)        | 210.8  | 121.4  | 72.5   | 19.4   | 4.4    | 2.0     | 1.3     |
|             | ARL( $\bar{X}$ ) | 210.8  | 129.4  | 80.2   | 21.6   | 4.1    | 1.6     | 1.1     |
|             |                  | (4.32) | (2.07) | (1.70) | (0.14) | (0.01) | (<0.01) | (<0.01) |
| (20,12,5)   | ARL(extr)        | 276.9  | 154.1  | 89.1   | 21.8   | 4.6    | 2.0     | 1.3     |
|             | ARL( $\bar{X}$ ) | 276.9  | 166.4  | 101.9  | 25.9   | 4.6    | 1.7     | 1.1     |
|             |                  | (6.50) | (3.03) | (1.45) | (0.18) | (0.01) | (<0.01) | (<0.01) |

If we use control limits such that both charts have the same  $ARL$  if the process is in-control, then from Table 3.3 we see that there is not much difference in  $ARL$  if a shift occurs. The extrema chart works slightly better in detecting small shifts. Larger shifts are detected earlier by the extrema chart in case of small sample sizes, but by the  $\bar{X}$ -chart for larger sample sizes, although the difference in the  $ARL$  is not large.

In addition to the  $t(10)$ -distribution, we also considered the Gamma(2, 1)-distribution. Results have been derived in a similar way as for the  $t(10)$ -distribution, which can be found in the Table 3.4.

Table 3.4: *ARLs* for a Gamma(2, 1)-distributed process with shift  $\theta$ .

| $(n, b, m)$ | $\theta$ :       | 0.0    | 0.1    | 0.2    | 0.5    | 1.0    | 1.5    | 2.0     |
|-------------|------------------|--------|--------|--------|--------|--------|--------|---------|
| (250,249,1) | ARL(extr)        | 250.0  | 228.7  | 209.3  | 160.4  | 103.3  | 66.8   | 43.4    |
|             | ARL( $\bar{X}$ ) | 250.0  | 228.8  | 209.5  | 160.9  | 103.9  | 67.4   | 44.0    |
| (25,23,2)   | ARL(extr)        | 300.0  | 253.6  | 214.5  | 130.3  | 57.6   | 26.1   | 12.2    |
|             | ARL( $\bar{X}$ ) | 299.9  | 252.9  | 216.5  | 140.7  | 66.2   | 32.2   | 16.5    |
|             |                  | (7.33) | (5.64) | (4.49) | (2.35) | (0.76) | (0.25) | (0.09)  |
| (25,20,3)   | ARL(extr)        | 230.0  | 182.0  | 144.2  | 72.5   | 24.1   | 8.6    | 3.4     |
|             | ARL( $\bar{X}$ ) | 230.0  | 188.2  | 155.8  | 86.5   | 34.2   | 14.2   | 6.4     |
|             |                  | (4.92) | (3.64) | (2.74) | (1.13) | (0.28) | (0.07) | (0.02)  |
| (25,15,5)   | ARL(extr)        | 210.8  | 147.8  | 104.1  | 37.6   | 8.0    | 2.3    | 1.2     |
|             | ARL( $\bar{X}$ ) | 210.8  | 158.8  | 119.0  | 51.6   | 14.8   | 5.1    | 2.2     |
|             |                  | (4.32) | (2.82) | (1.83) | (0.52) | (0.08) | (0.01) | (<0.01) |
| (20,12,5)   | ARL(extr)        | 276.9  | 192.6  | 134.6  | 47.4   | 9.4    | 2.6    | 1.2     |
|             | ARL( $\bar{X}$ ) | 276.9  | 207.6  | 156.2  | 68.0   | 18.9   | 6.2    | 2.6     |
|             |                  | (6.50) | (4.22) | (2.75) | (0.78) | (0.11) | (0.02) | (<0.01) |

If instead we would use the upper control limit for the  $\bar{X}$ -chart based on the normality assumption, then the *ARL* for an in-control process is much smaller than anticipated, even lower than for the  $t(10)$ -distribution. The *ARLs* for the values of  $n$ ,  $b$  and  $m$  in Table 3.4 would be 9.0, 4.7, 2.6, 1.4 and 1.5, resulting in very frequent false alarms. Using control limits that result in the same *ARL* for an in-control process, we see that the extrema chart works better than the  $\bar{X}$ -chart, especially for the sample size  $m = 5$ , where the *ARL* for the  $\bar{X}$ -chart is about twice as large as for the extrema chart for shifts of one standard deviation and more.

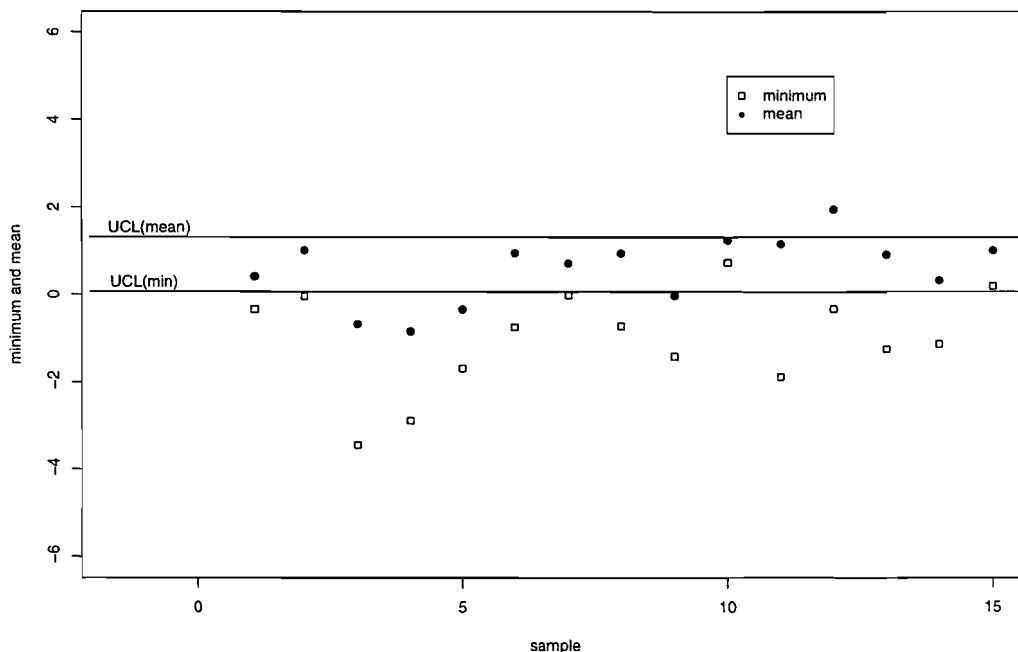
From the results presented here, we see that the extrema chart performs well compared to the  $\bar{X}$ -chart. The big advantage of the extrema chart over the  $\bar{X}$ -chart is that the *ARL* for an in-control process is known. Furthermore, even in the unrealistic situation that the  $\bar{X}$ -chart is used with the *ARL* equal to the anticipated value, the extrema chart works just as well in case of small sample sizes. The *ARL* for the extrema chart is in many situations smaller than for the  $\bar{X}$ -chart, and only slightly larger for some large positive shifts.

### Example extrema chart

To illustrate the use of the extrema chart, consider a process that has a  $t(4)$ -distribution, and while the process is in-control, 5 samples of size 5 are obtained to set the control limits for both the extrema chart and the  $\bar{X}$ -chart. After observing the reference set, the process is shifted upwards with a shift of

size  $\theta = 1$ . Another 10 samples of size 5 are obtained from the shifted process. The control limit for the extrema chart is set at the 15th observation from the reference set, corresponding to an  $ARL$  of 210.8. For the  $\bar{X}$ -chart the control limit is set such that if the process would be normally distributed, with a mean and standard deviation as estimated from the reference set, then the corresponding  $ARL$  would also be equal to 210.8. In Figure 3.2 we have plotted the minimum and the mean of the samples in the reference set and of the samples after the process has changed.

Figure 3.2: Control charts based on minimum and mean



We see that for the samples in the reference set, no out-of-control signal is given by either minimum or mean. After the shift has occurred, an out-of-control signal is given at sample 10 if the extrema chart is used, while with the  $\bar{X}$ -chart, we have to wait until the twelfth sample to get an out-of-control signal. It is interesting to note that in sample 12 the extrema chart would not give an out-of-control signal. Overall in this example, the extrema chart appears to be a bit better than the  $\bar{X}$ -chart in detecting the shift.

### 3.5 Some other inferential methods

In this section, we briefly mention a few other inferential methods based on the  $A_{(n)}$ -assumption that might be of use in some process control situations.

Coolen [6] shows how two independent samples can be compared using the method of section 2. Suppose that the ordered observations for the two samples are  $x_1, \dots, x_n$  (we will refer to this as the

$X$ -sample) and  $y_1, \dots, y_m$  ( $Y$ -sample), where we assume that  $n, m \geq 2$  and that there are no ties at all, for convenience of notation. We can compare the two sources, from which these samples were taken, predictively by comparing future observations  $X_{n+1}$  and  $Y_{m+1}$ , based on the two samples and the assumptions  $A_{(n)}$  for the  $X$ -sample and  $A_{(m)}$  for the  $Y$ -sample. We assume that these two sources are independent, in the sense that any information we learn about one source does not affect our beliefs about the other source. Denoting, as before,  $x_0 = -\infty$  and  $x_{n+1} = \infty$ , we introduce the notation (for  $j = 1, \dots, n$ )

$$t_j = \#\{y_i \mid x_{j-1} < y_i < x_j, i = 1, \dots, m\},$$

so  $t_j$  is the number of observations from the  $Y$ -sample in interval  $I_j$ , created by the observations from the  $X$ -sample. To compare  $X_{n+1}$  and  $Y_{m+1}$  we can consider the probability  $P(X_{n+1} > Y_{m+1})$ . On the basis of the  $n + m$  observations, and the assumptions  $A_{(n)}$  and  $A_{(m)}$ , we cannot derive this probability precisely. Coolen [6] derives the maximum lower bound ( $\underline{P}$ ) and minimum upper bound ( $\overline{P}$ ) for this probability under these assumptions as

$$\begin{aligned} \underline{P}(X_{n+1} > Y_{m+1}) &= \frac{1}{(n+1)(m+1)} \sum_{j=1}^n (n-j+1)t_j, \\ \overline{P}(X_{n+1} > Y_{m+1}) &= \frac{1}{(n+1)(m+1)} \left( n+m+1 + \sum_{j=1}^n (n-j+1)t_j \right). \end{aligned}$$

These bounds are *imprecise probabilities* in the sense of Walley [21]. Coolen and van der Laan [8] generalize these results to comparison of more than two samples, which could also play a role in statistical process control, but we do not consider this any further.

It is interesting to see how such predictive comparison based on two independent samples might be used for statistical process control. One could, for example, take one sample of products per day, and compare two days predictively. If there is considerable difference, e.g. in an extreme case where all  $X$ -sample observations are smaller than all  $Y$ -sample observations, these bounds will clearly indicate such a difference via the probabilities for the next observations, e.g. that  $Y_{m+1}$  is very likely to exceed the  $X_{n+1}$ . One could also use a reference set, for example a sample taken under favourable conditions ('process in-control'), and compare this with a sample taken at a different time, and even do so repeatedly. One could conclude that the process is out-of-control if the lower or upper bound is too small or too large, i.e. exceeds some specified lower or upper control limit. Unfortunately, however, it is difficult to establish control limits for either this lower or upper bound, corresponding to a prescribed average run length for a process which is in-control. Note that the probability bounds above are basically rank-based inferences, and they are equivalent to the Wilcoxon rank sum statistic [15]. It is easy to see that

$$\overline{P}(X_{n+1} > Y_{m+1}) = \underline{P}(X_{n+1} > Y_{m+1}) + \frac{n+m+1}{(n+1)(m+1)},$$

and the Mann-Whitney form of the Wilcoxon statistic is

$$\begin{aligned}
 U &= \sum_{j=1}^n \sum_{i=1}^m 1_{\{x_j > y_i\}} = \sum_{j=1}^n \#\{y_i \mid y_i < x_j, i = 1, \dots, m\} \\
 &= \sum_{j=1}^n \sum_{k=1}^j t_k = \sum_{j=1}^n (n - j + 1)t_j = (n + 1)(m + 1)\underline{P}(X_{n+1} > Y_{m+1}).
 \end{aligned}$$

Alloway and Raghavachari [1] and Pappanastos and Adams [19] discuss control charts based on the Hodges-Lehmann estimator, which is related to Wilcoxon's rank sum statistic. However, Pappanastos and Adams conclude that the way in which the control limits are set results in *ARLs* that are quite different from the anticipated values.

As these probability bounds are rank-based inferences, they do not use further information on location from the data. Coolen [6] also briefly presents an alternative method, based on the same assumptions, but providing bounds on expected value for the next observation. Such inferences could also be useful for statistical process control, but we do not consider this further.

A different way of predictively comparing two independent samples might also be of interest, it could be used, for example, to compare aspects of variation. Using the  $Y$ -sample and  $A_{(m)}$  for predictive probabilities for  $Y_{m+1}$ , we can consider bounds of probabilities for the position of  $Y_{m+1}$  with regard to the  $X$ -sample. For example, lower and upper bounds for  $P(Y_{m+1} < x_1)$  and  $P(Y_{m+1} > x_n)$  can easily be derived, and if these are particularly large it would reflect that the two samples are quite different. Of course, if  $m$  is about equal to  $n$ , one would expect these probabilities to be close to  $1/(n + 1)$  if the two samples are fairly similar, as this is the value which would have occurred if we had considered  $X_{n+1}$  instead of  $Y_{m+1}$ .

Another concept that can be used for statistical process control [5] makes use of so-called *precedence probabilities*. Let  $X_{(n+k)}$ , for  $k = 1, \dots, m$ , denote the  $k$ -th ordered value of the future observations  $X_{n+1}, \dots, X_{n+m}$ . The probability that  $X_{(n+k)}$  does not exceed a reference set observation  $x_b$  ( $b = 1, \dots, n$ ) is of interest, and is called a precedence probability [2, 3]. Using our method, this probability is again easily derived using a combinatorial argument and the random quantities  $S_j$  as introduced in section 2:

$$\begin{aligned}
 P(X_{(n+k)} \leq x_b) &= P\left(\sum_{j=1}^b S_j \geq k\right) = \sum_{l=k}^m P\left(\sum_{j=1}^b S_j = l\right) \\
 &= \sum_{l=k}^m \frac{\binom{l+b-1}{l} \binom{m-l+n-b}{m-l}}{\binom{n+m}{n}}.
 \end{aligned}$$

The probability  $P(\sum_{j=1}^b S_j = l)$  in this derivation plays a central role, and is effectively identical to a corresponding result at the heart of a classical frequentist approach to precedence tests (equation (4) in

Chakraborti and van der Laan [2]), except that in our  $A_{(n)}$ -based framework this is simply a probability on the future observations for a given reference set, whereas in the approach by Chakraborti and van der Laan the probability statement holds before the reference set has actually been observed, and is considered to be justified in the common frequentist sense of embedding the particular reference set of interest in an infinite sequence of reference sets that could occur. For both approaches, of course, this probability is only correct under the assumption that the process is in-control. As the actual results of these two approaches are identical, our approach can be used for similar inferences using such precedence probabilities as discussed in more detail by Chakraborti and van der Laan [2, 3, 5]. Again, the difference in these two approaches is in the ease of dealing throughout with predictive probabilities in our approach, and the different interpretation of the results.

So far, we have suggested a variety of possible inferences for process control. One obvious possibility not yet mentioned, and perhaps more suitable for problems in quality control (although the difference between these terms is sometimes rather vague), is to consider discrete random quantities counting numbers of successes and failures in batches. Coolen [7] presents a nonparametric predictive approach to such problems, closely related to the work in this paper, where on the basis of observed numbers of successes and failures, bounds for probabilities of such numbers in future observations are derived. We do not consider possible application of such an approach to problems in process control any further at this stage.

## 4 Discussion

The methods we have suggested and outlined in this paper can be used in several ways. One could use them on their own, which especially for inferences which are identical with those based on more classical nonparametric approaches would appear to be logical and acceptable. However, as only few assumptions have been made, one might also use our inferences as a sort of baseline method, next to, for example, inferences based on parametric models. If the overall guidelines to actually control the processes do not differ too much between the methods, that should give confidence in the appropriateness of both approaches. If, however, such guidelines are quite different, one may need to consider the assumptions underlying both approaches more carefully, where the assumptions underlying parametric models and related inferences might give cause for reconsideration. So, our method could actually be used to provide further insight into the effect of modelling assumptions in other approaches.

In section 3 it has been shown that the extrema chart is a useful tool to control a production process. It has been suggested [17] that practical use of control charts, and related characteristics

such as the *ARL*, benefits from regularly updating the reference set and the corresponding control limits. A topic of interest is study of theoretically justified guidelines for such updating, which could quite well suggest different frequencies of updating for different types of control charts. We are not aware of any existing work that explicitly addresses this topic.

### Acknowledgement

Part of this research was carried out while the third named author visited the University of Durham, with financial support from The British Council and NWO.

### References

- [1] Alloway, J.A., Jr. and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *Journal of Quality Technology*, **23**, 336-347.
- [2] Chakraborti, S. and van der Laan, P. (1996). Precedence tests and confidence bounds for complete data: an overview and some results. *The Statistician*, **45**, 351-369.
- [3] Chakraborti, S. and van der Laan, P. (2000). Precedence probability and prediction intervals. *The Statistician*, **49**, 219-228.
- [4] Chakraborti, S., van der Laan, P. and Bakir, S.T. (2000). Nonparametric statistical process control: an overview and some results. *Journal of Quality Technology*, to appear.
- [5] Chakraborti, S., van der Laan, P. and van de Wiel, M.A. (2000). A class of distribution-free control charts. *SPOR-Report 2000-11*, Eindhoven University of Technology.
- [6] Coolen, F.P.A. (1996) Comparing two populations based on low stochastic structure assumptions. *Statistics & Probability Letters*, **29**, 297-305.
- [7] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, **36**, 349-357.
- [8] Coolen, F.P.A. and van der Laan, P. (2001). Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference*, to appear.
- [9] De Finetti, B. (1974). *Theory of Probability* (2 volumes). Wiley, London.
- [10] Dempster, A.P. (1963). On direct probabilities. *Journal of the Royal Statistical Society, Series B*, **25**, 100-110.



- [11] Geisser, S. (1993). *Predictive Inference: an Introduction*. Chapman and Hall, London.
- [12] Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, **63**, 677-691.
- [13] Hill, B.M. (1988). De Finetti's Theorem, Induction, and  $A_{(n)}$  or Bayesian nonparametric predictive inference. In: *Bayesian Statistics 3* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.). Oxford University Press, 211-241 (with discussion).
- [14] Hill, B.M. (1993). Parametric models for  $A_n$ : splitting processes and mixtures. *Journal of the Royal Statistical Society, Series B*, **55**, 423-433.
- [15] Hollander, M. and Wolfe, D.A. (1999). *Nonparametric Statistical Methods*, 2nd Edition. Wiley, New York.
- [16] Irony, T.Z., Pereira, C.A. de B., Barlow, R.E. (1992). Bayesian models for quality assurance. In: *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.). Oxford University Press, 675-688.
- [17] Montgomery, D.C. (1996). *Introduction to Statistical Quality Control*, 3rd Edition. Wiley, New York.
- [18] Mosteller, F. and Rourke, R.E.K. (1973). *Sturdy Statistics: Nonparametric and Order Statistics*. Addison-Wesley, Reading, MA.
- [19] Pappanastos, E.A. and Adams, B.M. (1996). Alternative designs of the Hodges-Lehmann control chart. *Journal of Quality Technology*, **28**, 213-223.
- [20] Raghunathan, T.E. (2000). Bayesian analysis of quality level using simulation methods. *Journal of Quality Technology*, **32**, 172-182.
- [21] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.
- [22] Willemain, T.R. and Runger, G.C. (1996). Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, **28**, 31-38.