Nonparametric Predictive Inference for Acceptance Decisions

ELSAEITI, MOHAMED

Nonparametric Predictive Inference for Acceptance Decisions

Mohamed A. Elsaeiti

A thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
University of Durham
England

December 2011

Dedicated

To my parents, children and wife

Nonparametric Predictive Inference for Acceptance Decisions

Mohamed A. Elsaeiti

Submitted for the degree of Doctor of Philosophy
December 2011

Abstract

This thesis presents new solutions for two acceptance decisions problems. First, we present methods for basic acceptance sampling for attributes, based on the nonparametric predictive inferential approach for Bernoulli data, which is extended for this application. We consider acceptance sampling based on destructive tests and on non-destructive tests. Attention is mostly restricted to single stage sampling, but extension to two-stage sampling is also considered and discussed.

Secondly, sequential acceptance decision problems are considered with the aim to select one or more candidates from a group, with the candidates observed sequentially, either per individual or in subgroups, and with the ordering of an individual compared to previous candidates and those in the same subgroup available. While, for given total group size, this problem can in principle be solved by dynamic programming, the computational effort required makes this not feasible for problems once the number of candidates to be selected, and the total group size are not small. We present a new heuristic approach to such problems, based on the principles of nonparametric predictive inference, and we study its performance via simulations. The approach is very flexible and computationally straightforward, and has advantages over alternative heuristic rules that have been suggested in the literature.

Declaration

The work in this thesis is based on research carried out at the Department of Mathematical Sciences, Durham University, UK. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all the author's original work unless referenced to the contrary in the text.

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor Prof Frank Coolen, who has supported me throughout my thesis with his patience and kindness, knowledge, as well as his academic experience. Actually, his enthusiasm, untiring support and calm advice have been invaluable to me.

I would also like to thank all the members of the Durham Statistics and Probability Group for being there and helping when asked.

I am immensely grateful to my mother, my brothers and my father who inspired and encouraged me to undertake my PhD and I hope he is proud of this thesis.

My special thanks go to my wife for her unwavering belief in me and unlimited support, also I cannot forget my lovely and cute twins Faress and Remas who have changed my life in such an incredibly wonderful way. Thanks Faress and Remas for lighting my life with your smile.

Contents

Chapter 1

Introduction

1.1 Overview

Logical and optimal decisions have been considered in a number of disciplines. Acceptance decisions are an important topic in scientific and industrial circles. It is important for achieving the required objectives and standards in production processes. Acceptance decisions is a framework of statistical tools and concepts, used to help decision makers in making an optimal, or at least a good, decision from a set of alternatives. In this thesis we present Nonparametric Predictive Inference (NPI) for two acceptance decisions problems in Chapters 2 and 3.

Chapter 2 is focused on acceptance sampling, an important scenario in quality and reliability applications, where items from a production process are tested in order to decide whether or not to accept a batch of items. A specific form of such testing appears when the test result is simply whether or not a tested unit functions, which is known as 'attribute acceptance sampling'.

We present methods for basic acceptance sampling for attributes, based on the nonparametric predictive inferential approach for Bernoulli data presented by Coolen [7], which is extended for this application. We consider both acceptance sampling based on destructive tests and on non-destructive tests. Attention is mostly restricted to single-stage sampling, but extension to two-stage sampling is also considered and discussed.

In Chapter 3, sequential acceptance decision problems are considered with the

aim to select one or more candidates from a group, with the candidates observed sequentially, either per individual or in subgroups, and with the rank of an individual candidate compared to previous candidates and those in the same subgroup available. While, for given total group size, this problem can in principle be solved by dynamic programming, the computational effort required makes this not feasible for problems once the number of candidates to be selected, and the total group size, are not trivially small, as the number of paths to be used in the backward solution algorithm increases exponentially. We present a heuristic approach to such problems, based on the principles of nonparametric predictive inference, and we study its performance via simulations. We compare our approach to alternative methods. The approach is very flexible and computationally straightforward, and has advantages over alternative heuristic rules that have been suggested in the literature.

This study investigates the use of nonparametric predictive inference (NPI) for acceptance decisions. Section 1.2 is a brief introduction to NPI. Section 1.3 gives a detailed outline of the thesis.

1.2 Nonparametric Predictive Inference (NPI)

Nonparametric Predictive Inference (NPI) is a statistical inference technique based on Hill's assumption $A_{(n)}$ [28]. Hill proposed the assumption $A_{(n)}$ for prediction about future observations. This assumption was proposed particularly for situations in which there is no strong prior information about the probability distribution for a random quantity of interest. Suppose that we have n exchangeable real-valued observations $x_1, x_2, ..., x_n$, which can be ordered as $x_{(1)} < x_{(2)} < ... < x_{(n)}$. Let $x_{(0)} = -\infty$, $x_{(n+1)} = \infty$ and assume that the probability of ties is zero. These n observations divide the real-line into $n+1$ intervals $I_j = (x_{j-1}, x_j)$ where $j = 1, 2, ..., n+1$. The assumption $A_{(n)}$ is that the probability that the future random value X_{n+i}, for $i \geq 1$, will be in interval I_j is $\frac{1}{n+1}$, for each $j = 1, 2, ..., n+1$. Thus

$$P(X_{n+1} \in (x_{(j-1)}, x_{(j)})) = \frac{1}{n+1} \quad \text{for} \quad j = 1, 2, ..., n+1 \tag{1.1}$$

Augustin and Coolen [3] refered to the statistical inference approach based on

$A_{(n)}$ as 'Nonparametric Predictive Inference' (NPI), as inferences based on $A_{(n)}$ are predictive and nonparametric. Despite that $A_{(n)}$ is not an adequate assumption by itself to provide precise probabilities for many events of interest, it can give predictive probability bounds for one or more future observations, without requiring any further assumptions about the probability distribution. NPI is based on only few assumptions, one of which is that the data are exchangeable, thus inferences do not depend on the ordering of the data.

Lower and upper probabilities generalize the classical ('precise') concept of probability and allow several interpretations. From subjective perspective [43], the lower (upper) probability for an event can be interpreted as the supremum buying (infimum selling) price for the gamble that pays one if the event occurs and zero else, so for all prices less than the lower probability one would want to buy the gamble. From classical perspective, lower and upper probabilities can be interpreted as bounds on precise probabilities, due to limited information available or the wish not to add further assumptions. In this study, as in NPI more generally, the latter interpretation is used, without further claims in terms of subjective betting behaviour [8]. $\underline{P}(.)$ and $\overline{P}(.)$ denote lower and upper probabilities, respectively. The NPI lower and upper probabilities for the event $X_{n+i} \in B$, for $i \geq 1$, given the intervals $I_1, I_2, ..., I_{n+1}$, where $B \subset \mathbb{R}$, are

$$\underline{P}(X_{n+1} \in B) = \frac{1}{n+1} \left|\{j : I_j \subseteq B\}\right| \tag{1.2}$$

$$\overline{P}(X_{n+1} \in B) = \frac{1}{n+1} \left|\{j : I_j \cap B \neq \emptyset\}\right| \tag{1.3}$$

The lower probability (1.2) is obtained by summing up only the probability masses that must be in B, which happens only for the probability mass $\frac{1}{n+1}$ per interval I_j as long as the interval I_j is totally contained within B. The upper probability (1.3) is obtained by summing up all probability mass that can be in B, so any probability mass $\frac{1}{n+1}$ for interval I_j as long as the intersection of I_j and B is non-empty. The lower and upper probabilities presented by Coolen [7] fit in the framework of Nonparametric Predictive Inference (NPI) [3], hence we also call them

'NPI lower and upper probabilities'.

As we mentioned, in case there is not much information, one may not wish to make further assumptions to derive precise probability, so incomplete knowledge is reflected by imprecision in the lower and upper probabilities. One of the first developers of theory of imprecise probability was Boole [5]. It has received increasing attention by several researchers including the development of interval probability theory. Actually, for any event A, the precise 'classical' probability is just a special case of imprecise probability, when $\overline{P}(A) = \underline{P}(A)$, and the case with $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$ can be used to represent complete absence of information about event A.

The basic and fundamental concepts of interval probability theory have been presented by Walley [43] and by Weichselberger [45, 46], who generalizes Kolmogorov's axioms. Some elements of theory of interval probability as relevant to $A_{(n)}$-based inference are presented below.

Interval probability on a space $\mathcal{A}$ can be defined as $P(A) = [\underline{P}(A), \overline{P}(A)]$, where $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ for all $A \in \mathcal{A}$. The structure of this model as defined by Weichselberger [45] is

$$\mathcal{M} = \{p | \underline{P}(A) \leq p(A) \leq \overline{P}(A), \forall A \in \mathcal{A}\} \neq \emptyset$$

A function $P(.)$ on $\mathcal{A}$ is called an F-probability with structure $\mathcal{M}$, if $P(.)$ satisfies

$$\inf_{p(.) \in \mathcal{M}} \quad p(A) = \underline{P}(A)$$

and

$$\sup_{p(.) \in \mathcal{M}} \quad p(A) = \overline{P}(A)$$

for all $A \in \mathcal{A}$. According to [3] NPI lower and upper probabilities are F-probabilities. F-probability leads to some significant consequences, for example, $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$, $\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B)$ and also for simplyfing the calculation of the imprecise probabilities, for every F-probability the lower and upper probabilities are conjugate $\underline{P}(A) = 1 - \overline{P}(A^c)$, where A^c is the complementary event to A.

Recently, NPI has been developed for several problems in statistics, risk and reliability, and operations research. For example, NPI methods have been applied

for multiple comparisons of groups of real-valued data, which are of interest for situations where such comparisons are naturally formulated in terms of comparison of future observations from the different groups [17], and for precedence testing, where saving time or costs is considered when comparing different groups [16]. NPI for Bernoulli random quantities [7] has been applied to several other inferential problems, including high reliability demonstration [12]. NPI has also been used to compare m future numbers of successes in Bernoulli trials for different groups, where the inferences tend to become more imprecise when m increases, while increasing the number of observations in the data set tends to decrease the imprecision [14]. NPI for Bernoulli quantities is introduced in detail in Section 2.3. Furthermore, NPI has been applied to problems with multinomial data, using a latent variable model in the form of a probability wheel. This provides inference even when the number of outcome categories is not known, including the problem of prediction of occurrence of new failure modes [10]. Further examples of NPI applications are in the area of probabilistic safety assessment [9], and comparison and subset selection for proportions data [13,14]. For more details about NPI see $www.npi-statistics.com$.

1.3 Outline of thesis

Work described in this thesis covers the use of nonparametric predictive inference for acceptance sampling and sequential acceptance decision problems. This thesis is structured as follows. Chapter 2 presents nonparametric predictive methods for acceptance sampling, starting with an introduction in Section 2.1 and a brief overview about statistical process control (SPC) in Section 2.2. Section 2.3 introduces NPI for Bernoulli quantities including some numerical examples. In Sections 2.4 and 2.5 destructive tests and non-destructive tests are considered and discussed, respectively. Section 2.6 considers the application of NPI to two-stage acceptance sampling. We finish this chapter with some conclusions and remarks in Section 2.7. A paper based on this work has been published in Journal of Statistical Theory and Practice [15]. Parts of this chapter were presented at the 18th Advances in Risk and Reliability Technology Symposium: AR^2TS [22], and at the 6th International Symposium on

Imprecise Probability: Theories and Applications: ISIPTA'09.

Chapter 3 addresses the nonparametric predictive approach to the sequential acceptance problem. We begin this chapter with an introduction in Section 3.1, then we introduce the NPI method for solving the sequential acceptance problem in Section 3.2. Section 3.3 gives the application of the NPI method to select only one candidate. A generalisation is provided in Section 3.4, with more than one candidate to be selected. Section 3.5 considers the case of selecting multiple candidates observed in groups. This is followed by comparative studies between NPI and some other methods including a fixed-percentage method in Section 3.6, methods that improve the subset of accepted candidates in Section 3.7 and a randomised selection procedure in Section 3.8. Section 3.9 gives some concluding and remarks. A paper based on Chapter 3 is currently in submission and it has been presented at a seminar at Durham University. Finally, Chapter 4 presents the conclusions of this thesis, further remarks, and some suggestions for future work.

Chapter 2

Acceptance Sampling

2.1 Introduction

Acceptance sampling is a statistical technique which is widely used to decide whether or not to accept a batch of products that has already been produced. So, it focuses on the quality assurance and ignores the production process. Acceptance sampling for attributes or 'go, no-go inspection' is one of the key topics in quality control. It is a sampling-based inspection method where an item is classified as defective ('bad') or non-defective ('good') with respect to a specified quality requirement [34]. In this chapter, we present the application of Nonparametric Predictive Inference (NPI) to acceptance sampling scenarios, enabled via an extension of the lower and upper probabilities for Bernoulli random quantities presented by Coolen [7]. The problem considered is the decision whether or not a batch of products, also called items, can be accepted, in the sense that it satisfies a quality criterion, on the basis of tests of some items. We mostly restrict attention to testing of a single sample, but we also investigate the development of NPI for sequential sampling by considering two-stage sampling.

Throughout this chapter we assume that testing is perfect, and that it simply reveals whether a tested item functions or not. We distinguish between destructive testing, where tested items cannot be used again, and non-destructive testing, where tested items can be used again after testing and are then assumed to have the same condition as was revealed during testing. In the latter situation, we will assume that

the items tested are sampled from within the batch considered, but in the case of destructive testing it is more natural to assume that the items tested are not sampled from the batch for which the acceptance decision is required. Throughout we assume exchangeability of all items with regard to their functioning, so exchangeability of all the tested items and of those items in the batch under consideration which have not been tested.

We begin this chapter by briefly reviewing statistical process control (SPC) in Section 2.2. In Section 2.3 we review NPI for Bernoulli quantities [7], and we present an extension that is relevant for acceptance sampling. The justification of this new result is given in the Subsection 2.3.1. In Section 2.4 we apply NPI to acceptance sampling with destructive tests and in Section 2.5 to acceptance sampling with non-destructive tests. Section 2.6 provides an exploration of NPI for acceptance sampling over two stages, where first some items are tested, followed by the decision to either accept the whole batch, to reject it, or to continue sampling and only take the final acceptance decision after the items of the second sample have been tested. Some concluding remarks are given in Section 2.7.

2.2 Statistical Process Control (SPC)

The term 'quality' is always associated with an excellent service or product that meets the minimum requirement of the consumer. The person who is responsible for quality starts with determining the customer requirements and then the production process goes on until the product is satisfactory to the customer. From a technical point of view, 'quality' is a group of properties or specifications of a product, which together guarantee to satisfy the needs of customers [36]. This can be achieved only with certain requirements, standards or specifications.

Quality control is a practice which has been known since the early days of manufacture. However, given the current industrial development worldwide and the complexity of production procedures, the examination and testing of many products and services has become an inevitable process. For example, deviations from a production plan are very common and have negative effects on the product. However,

the main problem is that no one can predict these deviations, as things sometimes may get out of control all of a sudden. Consequently, it has become a necessity that production procedures should be very closely monitored. Hence in order to achieve the required competence of the system, technical deviations need to be kept at a minimum. This can be achieved through the use of Statistical Process Control (SPC), which involves taking the necessary measurements to help decision makers decide whether or not the production process is going as planned.

In 1924 W.A. Stewart, who worked for Bell Telephone Laboratories in the US, established a statistical chart to control the different variables in the production process. This chart constituted the basis of Statistical Quality Control (SQC). Then later on H. F. Dodge and H. G. Roming developed the process of Acceptance Sampling as an alternative to full examination procedures. SQC constitutes one of the branches of quality control procedure, which involves the acquisition, analysis and interpretation of data to be used in the process of quality control [34, 36]. This chapter however, focuses on the statistical approach to acceptance sampling for attributes as a means for quality control, and this approach constitutes only a small part of the quality control spectrum.

In most production processes, it is impossible to produce two items which are exactly identical. This is consistent with the natural fact that elements of the same group are not exactly identical. Yet, the differences between elements may be unnoticeable. Variation in production is always unavoidable due to factors such as variation of equipment, materials, production environment and manpower. As long as these variations exist in a natural and consistent manner, one should expect a number of chance causes or variations to happen. When the causes for variation are significantly large these can be identified as assignable causes. When all causes, apart from chance causes, are eliminated the process is said to be under statistical control, as it will be stable and predictable. By contrast when the causes for variation are specific, this may lead to excessive variation in which case the process is described as being out of control, going beyond the expected limit of variation [34, 36].

SPC is applicable to many production processes. It can help us to make a reliable assessment of the process, to determine the statistical process control limits,

to judge whether or not the process is out of control and to provide a process for early warning systems. Some researchers defined SPC in different ways. For example, Woodall and Montgomery [48] and Dale [18] defined it as a part of the field of statistical quality control consisting of several statistical methods and tools used to understand, adjust and monitor the performance of operations. Gaafar and Keats [24] defined SPC as a method that is used to monitor, evaluate and analyse the process to the continuous improvement of quality, reliability and service, by reducing the variations in operations. Furthermore, according to the definition by Goetsch and Davis [26], SPC is a statistical tool used to separate the variations due to special reasons of natural differences, or in general to eliminate the causes of the variations, in order to achieve and establish consistency and compatibility of the output of processes. Dale [18] summarised the fundamental aims for SPC as follows: to stabilize the process, to improve and reduce the variations in the output processes, to evaluate the performance of processes and to provide information about the processes to assist management in decision-making.

One issue to be considered in SPC is when the process should be stopped according to some quality specifications. In some cases one needs to determine whether the observed variation in quality exceeds the expected and acceptable level of variation. This is where control charts fit in to help with the analysis and presentation. Such charts represent an important tool in SPC, as they tend to assist decision makers to judge whether or not the production process is stable. This judgement is usually made in relation to two lines representing the upper and lower control limits (UCL and LCL, respectively), which help in detecting any significant variation in the quality of the product within the limits allowed for every single product.

In many cases qualitative or quantitative data may be involved to investigate the quality of a certain product. In this regard the term attribute may be used in relation to quality control. This term usually refers to product characteristics and specifications, and whether these specifications meet the potential requirements of consumers. Such quality characteristics are normally not quantitative, but they are attributes which can be described as either conforming or non-conforming to some requirements. Control charts for quantitative variables are not used in case where

the use of attribute is important for quality control. Hence, the control charts usually used in this type of quality characteristics are called control charts for attributes [34, 42].

One of the most important areas of quality control and improvement is inspection sampling or acceptance sampling for attributes. Ensuring the quality of deliveries (orders) which arrive at a company is important and necessary for the quality of the final product. This is achieved by examining the quality of the supply (so a process of assessing quality). It is also used in the last stages of production for inspection on the quality of the final product before marketing to the consumer. In brief, it can be said that acceptance sampling is the process of examination and inspection for quality of the materials entering a company, as well as the final product.

This field of quality control has two different classifications: acceptance sampling plans for variables and plans for attributes, depending on whether the quality is measured numerically or not. This study will focus on inspection sampling for attributes. In order to examine the batch of attributes there are several acceptance sampling plans. First, a single sampling plan can be used. This considers results from a single test. Secondly, double sampling plans were invented to give a batch another chance, where if the results of the first sample are not conclusive with regard to accepting or rejecting the batch, a second sample is selected. Thirdly, multiple sampling plans can be used, generally both double and multiple plans are just an extension of the single-sampling plan. In addition, double-sampling and multiple sampling programs can be repeated, and also sequential sampling plans can be used [34, 36, 42].

Inspection sampling or acceptance sampling for attributes implies the initial determination of the units of the sample from each batch, which will then be examined according to the required specifications. If the number of nonconforming units is less than a certain limit the sample will be accepted, otherwise the sample will be rejected on the grounds that it deviates from the ideal sample.

However, acceptance sampling for attributes, also called lot by lot sampling or 'go, no-go' method, is usually used in the following two situations. In cases where the test targets defective samples for purposes of destruction (such as testing the

earth wire in electricity or safety airbags in vehicles), the importance of sampling increases from the fact that defective samples will be destroyed straight away. Also, when comprehensive examination of the units in question is expensive, or if the process is laborious or time consuming, lot by lot sampling is used.

From the foregoing, it can be seen that the sampling process may be economically feasible as it saves time and effort, also it tends to reduce the defective units in the accepted batch. Ihe process of sampling may involve some risks such as accepting defective batches or rejecting valid batches. Moreover, the sampling process does not guarantee that all samples in the batch meet the required specifications. Acceptance sampling always implies a conflict of interest between the consumer and producer. The Operating Characteristic curve, or the OC curve, is often used in quality control to determine the probability of accepting production lots when using different sampling schemes. The ideal sampling plan can be obtained when the OC curve develops to a vertical straight line, which satisfies the requirements of both the producer and consumer. This ideal curve however, could be obtained only when all the units have been comprehensively tested. Hence given the above mentioned problems associated with comprehensive testing, it could be maintained that obtaining the ideal OC curve could be a very difficult, if not impossible, task in most cases [34].

It is obvious that the sampling process is always associated with some hazards such as the hazard of rejecting good batches. The producer's risk is always associated with an index defining the qualified batch referred to as Acceptance Quality Level (AQL). The AQL actually indicates the deviation from the ideal sample in terms of percentage, and this percentage could be reasonable for the purpose of acceptance sampling. The consumer's risk is also associated with an index defining the disqualified batch, better known as the Limiting Quality Level (LQL). The LQL is defined as the percentage deviation from the specification as required by the consumer, which reduces the chances of acceptance of the product by consumers. Finally it is worth mentioning that both AQL and LQL can be determined by the OC curve. Generally, the Operating Characteristic curve describes the probability of accepting a batch as a function of the batch's quality. This chapter focuses on

acceptance sampling for attributes, which is a special problem and does not use all these concepts. It will be an interesting challenge to link NPI to these concepts.

Recently, nonparametric predictive inference has been used in the field of SPC. Arts, Coolen and van der Laan [2] introduced a control chart called 'extrema chart' that uses NPI to decide whether the process is in control or not. These extrema charts are a generalisation of control charts presented by Willemain and Runger [47]. Both methods coincide if the attention is restricted to only one future observation $m = 1$. Arts, Coolen and van der Laan [2] presented one-sided extrema charts which can be created by using the minimum or the maximum of m future observations. Two-sided extrema charts can also be created by using the minimum and maximum m future observations together. Whether or not a process is in a state of statistical control depends on the fact that any shift in the under study process leads to a larger minimum or a smaller maximum observation than expected on the basis of a reference set. Simulation studies showed that the extrema charts gave good performance compared to the well known $\bar{X}$ and CUSUM charts. Due to the fact that NPI is based on $A_{(n)}$ assumption, and inferences based on $A_{(n)}$ are predictive and nonparametric, enabled by the use of lower and upper probabilities to quantify uncertainty, it is worth to use the extrema charts for the advantage that it does not require any prior information about the production process apart from the reference set.

Arts and Coolen [1] introduced two NPI-based control charts, namely control charts using order statistics and subsequent sampling charts as an attractive alternative control charts. These NPI control charts use other order statistics and require a larger reference set, but they are good alternative tools to the extrema charts because they avoid the condition that a false signal occurs only if the whole sample exceeds the control limit. For the new charts using order statistics the signal occurs if a part of the sample exceeds the control limit. They also presented control charts which use subsequent samples, where the authors paid attention only to the case of $m = 1$. This type of control charts reveals the small shifts in the process more quickly than the extrema charts and the subsequent sampling charts, and it does not require a large reference set.

2.2.1 Formation of batches

The formation of batches could affect the sampling plan. For example, the batches should be homogenous, which means that all the products included by the batch under investigation should be produced by the same machine, the same manpower, and all other equipments should be the same and so on. Also, big batches are preferable.

A batch of items can be inspected in several ways. The sampling procedure may consist of selecting a single sample, or it may be done in two or more steps.

2.2.2 Single sampling plans for attributes

The single plan involves a single sample of the batch and the decision of rejection or acceptance of the whole batch is made on the results of testing this sample. A random sample of size n is taken from a homogeneous batch of size N and the sample acceptance number C is the required number of items to reject or accept the batch. For example, the batch is accepted only if the number of defect items in n is not more than C, otherwise we reject the batch.

The Operating Characteristic (OC) Curve

In general, the operating characteristic (OC) curve constitutes a popular method of evaluation for production batches. For judging a specific sampling plan it is preferable to know that a certain batch with a certain percentage of failed or nonconforming items is likely to be accepted or rejected. This curve shows the probability of accepting the provided batch and the batch percentage of the failed items. So, the OC curve illustrates the ability of sampling plans to distinguish between accepting and rejecting a questionable batch and it also gives the relationship between the batch acceptance probability and its quality. Despite of these advantages the OC curve cannot deal with the uncertainty or in the case of extremely vague a priori knowledge about the form of the underlying distribution, but NPI can provide predictive probability bounds without making an assumption about the distribution that the data have come from.

Suppose that we have a large batch of size N items and a random sample of size n items is inspected. It is assumed that the batch has been chosen from a continuous production flow which can be considered infinite. Hence a binomial distribution can be used in the calculations. The number of defective units D is binomially distributed with parameters n and p, where p is the fraction of defective units in the batch. In this case,

$$P(D = d) = \binom{n}{d} p^d (1-p)^{n-d} \tag{2.1}$$

and the probability of acceptance of the batch is

$$P(D \leq C) = \sum_{d=0}^{C} \binom{n}{d} p^d (1-p)^{n-d} \tag{2.2}$$

To explain the idea, suppose that we have a random sample of size $n = 89$, the batch fraction defective is just one percent $p = 0.01$ and we will accept the batch if there are 2 or fewer defectives items, so $C = 2$. According to these information the probability of acceptance of the batch is

$$P(D \leq 2) = \sum_{d=0}^{2} \binom{89}{d} (0.01)^d (0.99)^{89-d} = 0.939$$

Dodge [20] suggested the 'chain sampling plan' or 'ChPS-1' as an alternative plan when inspection is costly or destructive for situations where small sample sizes and a zero-acceptance number $C = 0$ are desired. As when $C = 0$ the OC curve for the single sample plan is inappropriate tool, as it has no point of inflection and hence it drops dramatically even for any small increase in the fraction failure items. Dodge uses prior information of the previous batches as the method uses cumulative results of several samples in making a decision about the current batch. The procedure of the chain sampling plan is, if the sample has 0 failure items, accept the batch; if the sample has 2 or more failure items, do not accept the batch; and if the sample has 1 failure item, it may be accepted provided that there are 0 failure items in the previous i samples of size n.

2.2.3 Double-sampling plans

Double sampling plans are more complex. The decision will be taken on the basis of testing of a preliminary sample to either accept the whole batch, reject the whole

batch or to continue the test with another sample. In other words, if the quality is excellent the first test will be sufficient and the batch will be qualified, but if the results of the test show a very poor quality the whole batch will be rejected and no further tests are made. However, a second test is essential in case of a moderate quality, i.e. neither very poor nor very good. In cases where another sample is necessary the result of the tests of the first and second samples are used simultaneously for making the decision for acceptance or rejection of the batch. So, it is an extension method of single sampling plans. The double-sampling plan is determined by five numbers: n_1 is the sample size of the first sample; C_1 is the acceptance number of the first sample; R_1 is the rejection number of the first sample; n_2 is the sample size of the second sample; C_2 is the acceptance number for the combined samples. To illustrate the plan, suppose a random sample of n_1 units is taken from a specific batch. This sample has D_1 defective items, and if

$$\begin{cases} D_1 \leq C_1 \text{ the batch is accepted on the first sample} \\ D_1 > R_1 \text{ the batch is rejected on the first sample} \\ C_1 < D_1 \leq R_1 \text{ a second sample of size } n_2 \text{ is selected} \end{cases}$$

In the last case, n_2 further items are taken from the batch. Let D_2 be the number of defective items in the second sample. If the total number of defective items in the combined samples, $D_1 + D_2$, is not more than C_2, the whole batch is accepted, while the batch is rejected if $D_1 + D_2$ is greater than C_2.

A double-sampling plan has some advantages and disadvantages. The main advantage of a double-sampling plan is that the total amount of required inspection could be reduced. Moreover, it could give a batch a second chance. This feature has been called 'the psychological advantage' [34].

The multiple sampling plan is a continuation of the dual sampling plan procedure. In case of multiple sampling it is preferable to select three, four, five or even more samples but the size should be small. Since the same procedure is followed as in the dual sampling plan, the multiple sampling plan can be used with any number of stages, with the possibility that the whole batch can be sampled before a decision is made.

It is possible that the same results could be obtained by any of the three types

of sampling plans above. The choice of the type of plan for testing a certain unit can depend on factors other than its effectiveness. These factors include simplicity, costs of management, available data, the number of units to be examined, and psychological impact. However, since simplicity is the most important factor, it follows that individual sampling is the most favourable while multiple sampling is the least favourable.

2.3 NPI for Bernoulli quantities

Coolen [7] presented NPI lower and upper probabilities, also called 'imprecise probabilities' [43] or 'interval probability' [45, 46], for prediction of Bernoulli random quantities. These lower and upper probabilities followed from an assumed underlying latent variable model similar to Bayes' original representation [4, 7], yet without a prior distribution, with future outcomes of random quantities related to data by Hill's assumption $A_{(n)}$ [28, 29].

In this section, we summarize results from [7] on NPI inference for Bernoulli random quantities. We refer to [7] for justifications, which are based on a latent variable representation of Bernoulli quantities as real-valued outcomes of an experiment similar to that used by [4], with Hill's assumption $A_{(n)}$ [28, 29] used to derive direct predictive probabilities [19] for future observations using available data. In short, Hill's assumption $A_{(n)}$ for real-valued random quantities is that a future observation will fall into each interval on the real line, as created by the first n observations, with equal probability $1/(n+1)$. Due to the use of $A_{(n)}$ in deriving the NPI lower and upper probabilities, they fit in a frequentist framework of statistics but can also be interpreted from Bayesian perspective [29, 30]. For further discussion of properties of NPI see [3].

Suppose that we have a sequence of $n+m$ exchangeable Bernoulli trials, each with 'success' and 'failure' as possible outcomes, and data consisting of s successes in n trials. Let Y_1^n denote the random number of successes in trials 1 to n, and Y_{n+1}^{n+m} the random number of successes in trials $n+1$ to $n+m$. Let $R_t = \{r_1, \ldots, r_t\}$, with $1 \leq t \leq m+1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, and, for ease of notation, let

us define $\binom{s+r_0}{s} = 0$. Then the NPI upper probability [7] for the event $Y_{n+1}^{n+m} \in R_t$, given data $Y_1^n = s$, for $s \in \{0, \ldots, n\}$, is

$$\overline{P}(Y_{n+1}^{n+m} \in R_t | Y_1^n = s) = \binom{n+m}{n}^{-1} \sum_{j=1}^{t} \left[\binom{s+r_j}{s} - \binom{s+r_{j-1}}{s} \right] \binom{n-s+m-r_j}{n-s} \tag{2.3}$$

The corresponding NPI lower probability [7] can be derived by

$$\underline{P}(Y_{n+1}^{n+m} \in R_t | Y_1^n = s) = 1 - \overline{P}(Y_{n+1}^{n+m} \in R_t^c | Y_1^n = s) \tag{2.4}$$

where $R_t^c = \{0, 1, \ldots, m\} \backslash R_t$. This relation between these upper and lower probabilities is justified by [7], and agrees with the fact that these are F-probabilities in Weichselberger's theory of interval probability [3, 45, 46].

More general results, for events $(Y_{n+1}^{n+m} \in R | Y_1^n \in S)$ for general sets R and S are being developed, together with study of foundational properties for statistical inference using lower and upper probabilities on the basis of such set-valued data. For the acceptance sampling application in this chapter, we need to extend the results from [7] to the event $(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s)$, as explained in Subsection 2.3.1. Using the information $Y_1^n \geq s$ instead of a precise value for Y_1^n is important at the stage where we define the procedure, which is done by specifying the minimum number of successes in the test for which a batch will be accepted. For such events, the NPI lower and upper probabilities are given below, followed by a brief explanation of the derivation of these results. The NPI lower probability for the event $(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s)$ is, for $0 \leq s \leq n$ and $0 \leq r \leq m$,

$$\underline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s) = \binom{n+m}{m}^{-1} \sum_{j=r}^{m} \binom{s-1+j}{j} \binom{n-s+m-j}{m-j} \tag{2.5}$$

and the corresponding NPI upper probability is always equal to 1, so

$$\overline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s) = 1 \tag{2.6}$$

To put these forms (2.5) and (2.6) more simply, the lower probability (2.5) is derived by counting the number of orderings out of the $\binom{n+m}{m}$ orderings as explained in Subsection 2.3.1, for which $Y_1^n \geq s$ must be followed by at least r successes in m future trials, and the upper probability is the number of orderings for which $(Y_1^n \geq s)$ can be followed by $(Y_{n+1}^{n+m} \geq r)$.

2.3.1 NPI lower probability for acceptance sampling

Let us give a brief explanation of the derivation of the NPI lower and upper probabilities (2.5) and (2.6). The explanation below builds directly on the approach and results for NPI for Bernoulli random quantities presented by [7], where the data are restricted to a known number of successes in n trials, so $Y_1^n = s$, while in this study the generalization to information of the form $Y_1^n \geq s$ is presented and used in Sections 2.5, 2.6 and 2.7.

Throughout, interest is in the number of successes in m future trials, with information on the number of successes in n past trials. In NPI for Bernoulli random quantities [7], past observations are related to future random quantities via an assumed underlying latent variable representation, such that each value is represented by a point on the real line, with a threshold such that all points to one side of the threshold represent 'successes', and all points to the other side of the threshold represent 'failures'. No knowledge about this threshold is assumed. This representation is very similar to that used by Bayes [4], with the exception that Bayes made explicit assumptions on the threshold, which in the later development of Bayesian statistical methodology corresponded to the assumption of a prior distribution.

In NPI, with the latent variable representation, past observations are related to future observations via Hill's assumption $A_{(n)}$ [28, 29]. Suppose that the ordered values of the latent variables corresponding to the n observations are $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$. These n values define a partition of the real line, consisting of $n+1$ intervals. Hill's assumption $A_{(n)}$ states that a future random quantity U_{n+1} has equal probability $1/(n+1)$ to be in each of these intervals. In our NPI setting this U_{n+1} is the latent variable corresponding to the first future observation, which will again be a success or failure, depending on which side of the threshold U_{n+1} is. When interested in m future observations, as we are in this study, the same assumption needs to be made for each future observation consecutively, so one needs to assume $A_{(n)}, \ldots, A_{(n+m-1)}$. In fact, assuming $A_{(n+m-1)}$ is sufficient, as Hill [28] shows that the assumption $A_{(n)}$ implies $A_{(k)}$ for all $k \leq n$. Under these assumptions, the following result holds [7, 28]. Suppose that there is no interest in precisely which of the first n observations are successes or failures, so that one considers the number

of successes as a sufficient statistic, and the same is assumed for the m future observations of interest. Then all $\binom{n+m}{n}$ different orderings of the underlying latent variables on the real line, which represent the first n observations and the m future observations, have the same probability, also after information about the number of successes in the first n observations has become available. Denoting these $\binom{n+m}{n}$ different orderings by O_j, for $j = 1, \ldots, \binom{n+m}{n}$, the lower and upper probabilities (2.5) and (2.6) are derived by counting orderings, in the same manner as in [7]: for the lower probability, only those orderings are included for which $Y_1^n \geq s$ *has* to be followed by $Y_{n+1}^{n+m} \geq r$, while for the upper probability all orderings are included for which $Y_1^n \geq s$ *can* be followed by $Y_{n+1}^{n+m} \geq r$.

Let us first consider the NPI upper probability $\overline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s)$, for which we count all orderings O_j of the $n+m$ latent variables on the real line for which $Y_1^n \geq s$ *can* be followed by $Y_{n+1}^{n+m} \geq r$. The information $Y_1^n \geq s$ does not exclude the possibility that the threshold between latent variable representations of successes and of failures lies, effectively, at infinity (assuming that a success is represented by a latent variable to the left of the threshold), in which case all m future observations would also be successes. Therefore, each ordering O_j is included in the count, and hence this upper probability for the event $Y_{n+1}^{n+m} \geq r$ given $Y_1^n \geq s$ is equal to one. This also agrees with the intuition that, if we only know that there are at least s successes in the first n observations, it is not logically excluded that the process considered will only give successes, never failures. As mentioned in this chapter, this upper probability is of little further interest in applications of NPI to acceptance sampling.

The NPI lower probability $\underline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s)$ is more interesting, it is determined by counting the orderings O_j of the $n+m$ latent variables on the real line for which $Y_1^n \geq s$ *must* be followed by $Y_{n+1}^{n+m} \geq r$. Assume again that a success is represented by a latent variable to the left of the threshold. Any ordering for which this is the case, must be such that r of the m latent variables representing future observations are to the left of the s-th ordered latent variable representing the first n observations ($u_{(s)}$ in the notation used above). A direct counting argument now leads to the lower probability (2.5). Reasoning in an alternative manner provides

important further insight. When considering the representations of the Bernoulli random quantities by latent variables on the real line, with a threshold such that points to the left of the threshold represent 'successes' and to its right represent 'failures', it is clear that for each ordering O_j for which s successes in the first n values must be followed by at least r successes in the future m observations, it is necessarily the case that i successes in the first n values, for every $i \geq s$, also must be followed by at least r successes in the future m observations. This suggests, for NPI for Bernoulli random quantities, that the following relation holds,

$$\underline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s) = \underline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n = s)$$

This is indeed the case, as can be confirmed directly from (2.3), (2.4) and (2.5). Clearly, for the NPI upper probability discussed above, we similarly have

$$\overline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s) = \overline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n = n) = 1 \tag{2.7}$$

Example 2.1

Suppose that we have 2 past trials ($n = 2$) and we are interested in the number of successes in 3 future trials ($m = 3$). In this case we have $\binom{2+3}{3}$ orderings as explained in Figure 2.1. According to the explanation above, let us assume that the left side represents the 'successes' (S), the right side represents the 'failures' (F). We are interested in the events $(Y_3^5 \geq r | Y_1^2 \geq s)$, the NPI lower and upper probabilities for these events are presented in Table 2.1. If we are interested in having at least 1 success in the future 3 items, so $r \geq 1$, given at least 1 success in the first two checked items, so $s \geq 1$, then the threshold can be anywhere in the intervals I_2 or I_3. For the lower probability for this event, only in the first 6 orderings $Y_1^2 \geq 1$ *must* be followed by $Y_3^5 \geq 1$, thus $\underline{P} = \frac{6}{10}$, while all orderings *can* be followed by $Y_3^5 \geq 1$, so the upper probability is equal to one. The other lower and upper probabilities in Table 2.1 are derived by similar arguments.

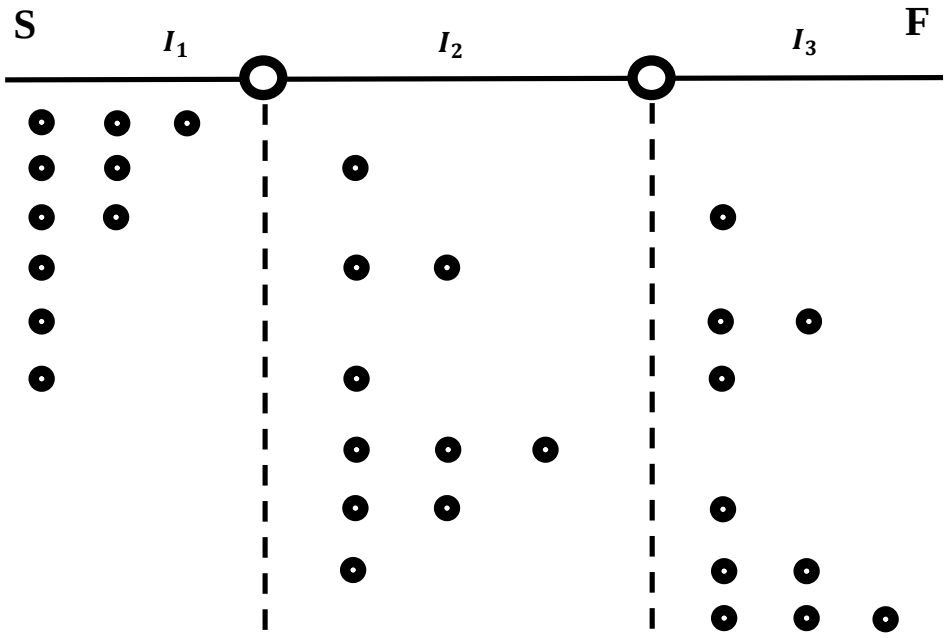

Figure 2.1: Orderings for $n = 2$ and $m = 3$

r	$s = 0$		$s = 1$		$s = 2$	
	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
0	1	1	1	1	1	1
1	0	1	0.6	1	0.9	1
2	0	1	0.3	1	0.7	1
3	0	1	0.1	1	0.4	1

Table 2.1: NPI lower and upper probabilities for $(Y_3^5 \geq r | Y_1^2 \geq s)$

2.4 Acceptance sampling with destructive tests

In this section, the use of NPI for acceptance sampling for quality control for attributes is explored. Assume that items are either good ('success') or bad ('failure') and that they can be tested in a perfect manner, that is a test reveals with certainty whether an item is good or bad. This test might destroy the item, so a tested item cannot be used thereafter, or it may be possible to use a tested item again. In this section, we assume that tested items cannot be used any further, which is also called 'destructive testing'. The alternative of 'non-destrictive testing' is considered in Section 2.5.

Suppose that per unit of time considered, say e.g. per day, one wishes to produce $m > 0$ items for delivery to the market, and in addition one also produces the number required to be tested during that day. Suppose, in line with notation above, that one tests n items on a day (so one would actually produce $n+m$ items that day), of which Y_1^n are good. On the basis of the outcome of this test, one wishes to decide on whether or not to 'accept' the m further items produced that day. Let us suppose that one chooses the following criterion, where Y_{n+1}^{n+m} is the random total number of good items out of the m units that are produced that day but that are not tested (so $Y_{n+1}^{n+m} \in \{0, 1, \ldots, m\}$).

The question considered is how many items should be tested, and how many of these should be good, in order to meet a suitable quality requirement. Assume the following quality requirement is used for a batch of m units

$$\underline{P}(Y_{n+1}^{n+m} \geq r | Y_1^n \geq s) \geq p \tag{2.8}$$

where r and p are chosen in line with the required quality. So the number of items to be tested is n, and the minimum number of these which must function is s.

The main question of acceptance sampling is to determine pairs (n, s) for which this criterion is satisfied, given m, r and p. Typically, the quality criterion must be set prior to testing, and this might involve cost considerations and negotiation between producer and consumer. It is important to emphasize that, at the stage prior to testing when the quality criterion must be set, it is logical to consider the

lower probability (2.8), that is with conditioning on $Y_1^n \geq s$. This is because if one sets a minimum number s of good items in n tests, in order to decide to accept the batch, then it implies logically that one would also accept the batch if the actual number of good items was larger than s.

The criterion (2.8), with the NPI lower probability (2.5), does not provide a general closed-form expression for solution pairs (n, s), so the discussion in this section is mostly based on numerical examples. Of course, pairs (n, s) for which this criterion is satisfied are easily computed as numerical calculation of the lower probability (2.5) is straightforward. Explicit solutions can only be derived for some special cases. For example, suppose that one would demand very high quality, in the sense that tests revealing any failures would lead to rejection of the batch, which leads to the requirement that $s = n$. If, simultaneously, one sets $r = m$ in the criterion (2.8), then the required sample size n has to satisfy $n \geq \frac{pm}{1-p}$, while for $r = m-1$ this criterion is satisfied if $n \geq \frac{1}{2} - m + \sqrt{\frac{m^2-m}{1-p} + \frac{1}{4}}$. For the situation with $s = n-1$ and $r = m$, so where one would accept the sample if testing revealed zero or one failures, but with the explicit quality criterion on the lower probability for zero failures in the m further items, criterion (2.8) is satisfied if $n \geq \frac{1}{2} + \frac{pm}{1-p} + \sqrt{\frac{pm^2}{(1-p)^2} + \frac{1}{4}}$. These three cases are briefly illustrated in Example 2.3, together with a few related cases for which no analytic expressions were derived, hence for which the minimum required sample sizes were determined numerically.

Example 2.3

Suppose one requires $m = 10$ items for future use, and wishes to test n other items, which are exchangeable with those 10 items, in order to satisfy the quality criterion (2.8) with $p = 0.8$. Suppose that one really requires all 10 future items to function, so one sets $r = 10$. If one decides to only accept the 10 items if the sample of size n contains only good items, so one sets $s = n$, then the minimum required sample size is $n = 40$. If one decides to also accept the batch of 10 items if one out of the n tested items is faulty, so one sets $s = n - 1$, while keeping $r = m = 10$, then the minimum required sample size is $n = 86$, whereas even accepting the batch if two of the n tested items are faulty in the same situation leads to minimum required sample size

$n = 131$. If, however, one relaxes the quality criterion by setting $r = m - 1 = 9$, so one accepts that one of the 10 items for future use could be faulty, still with $p = 0.8$, and one decides to only accept these 10 items if all items tested are good ($s = n$), then the minimum required sample size is $n = 12$. These figures are also presented in Table 2.3, together with minimum required sample sizes for some more related cases, including different values of p in the criterion (2.8). Tables 2.2, 2.4 and 2.5 present similar results for the cases $m = 5$, $m = 15$ and $m = 20$, respectively. These tables clearly show that these inferences are strongly influenced by choice of p, in particular for values of p close to 1 very large numbers of items must be tested. They also indicate that the minimum required sample sizes are nearly linear as function of m for all situations.

In most acceptance sampling situations, the idea is to sample *a few* items to decide on acceptance of *many* items. The NPI approach to acceptance sampling requires a large number of items to be tested, this is because the criterian used is actually aimed at achieving very high quality based on strong evidence from the test data only. In the NPI approach the required sample size n can be smaller if we allow more failed items in the future items. For example, see Table 2.2, if we want all the future items to function well, so $r = m = 5$, and if one sets $s = n$ with the highest considered quality requirement, $p = 0.99$, then at least 495 items have to be tested and all must function well. If we allow one nonfunctioning item in the future batch, so $r = 4$, then the number of items to be tested becomes much smaller, with 72 items as a minimum required sample size. Additionally, if we doubled the batch size to $m = 10$, as in Table 2.3 and keeping $p = 0.99$, at least 990 items have to be tested if one requires all future items to be good with $s = n$ and $p = 0.99$, whereas at least 86 items have to be tested if 1 future item is allowed to fail. Similar results can be clearly seen in Tables 2.4 and 2.5.

p	$s = n$ $r = 5$	$s = n - 1$ $r = 5$	$s = n - 2$ $r = 5$	$s = n$ $r = 4$	$s = n - 1$ $r = 4$	$s = n$ $r = 3$	$s = n - 1$ $r = 3$
0.50	5	13	21	2	5	1	3
0.70	12	27	41	4	9	2	5
0.80	20	43	66	6	12	3	6
0.90	45	93	141	10	19	5	9
0.95	95	193	291	16	30	7	13
0.98	245	493	741	28	50	11	19
0.99	495	983	1491	41	72	15	25

Table 2.2: Minimum required sample sizes n with $m = 5$

p	$s = n$ $r = 10$	$s = n - 1$ $r = 10$	$s = n - 2$ $r = 10$	$s = n$ $r = 9$	$s = n - 1$ $r = 9$	$s = n$ $r = 8$	$s = n - 1$ $r = 8$
0.50	10	25	40	4	10	3	7
0.70	24	52	81	8	18	5	10
0.80	40	86	131	12	25	7	13
0.90	90	186	281	21	40	11	20
0.95	190	386	580	33	62	16	28
0.98	490	986	1481	58	104	25	42
0.99	990	1986	2981	86	153	33	56

Table 2.3: Minimum required sample sizes n with $m = 10$

p	$s = n$ $r = 15$	$s = n - 1$ $r = 15$	$s = n - 2$ $r = 15$	$s = n$ $r = 14$	$s = n - 1$ $r = 14$	$s = n$ $r = 13$	$s = n - 1$ $r = 13$
0.50	15	37	59	6	15	4	10
0.70	35	78	120	12	26	7	15
0.80	60	128	196	18	37	10	20
0.90	135	278	421	32	61	17	31
0.95	285	578	871	51	94	24	42
0.98	735	1478	2221	88	159	38	65
0.99	1485	2978	4471	131	233	51	86

Table 2.4: Minimum required sample sizes n with $m = 15$

p	$s=n$ $r=20$	$s=n-1$ $r=20$	$s=n-2$ $r=20$	$s=n$ $r=19$	$s=n-1$ $r=19$	$s=n$ $r=18$	$s=n-1$ $r=18$
0.50	20	49	78	9	20	5	13
0.70	47	103	160	17	35	10	21
0.80	80	170	261	25	49	14	27
0.90	180	371	561	43	81	22	41
0.95	380	771	1161	68	126	33	58
0.98	980	1971	2961	119	213	51	88
0.99	1980	3971	5961	176	312	70	116

Table 2.5: Minimum required sample sizes n with $m = 20$

2.5 Acceptance sampling with non-destructive tests

In Section 2.4, items tested could not be used again. In this section, we consider the NPI method for acceptance sampling in situations where items that have been tested can be used again. We assume as before that tests are perfect, in that they reveal with certainty if a tested item is good or bad, and that good (bad) items remain good (bad). Suppose we have a batch of items of total size t, and our quality criterion will be based on the requirement that at least v of these t items should function. The question is again how this requirement should be specified, prior to testing taking place. To link this setting to the notation used before, we will again determine the test sample size n and the minimum number s of these tested items that have to be good. This leaves $m = t - n$ items untested, of which we require at least $r = v - s$ to be good with a chosen minimum value p of the corresponding lower probability. The quality criterion that we set for this situation is, with $0 \leq s \leq n \leq t$ and $0 \leq r \leq t - n$,

$$\underline{P}(Y_{n+1}^{t} \geq r | Y_{1}^{n} \geq s) \geq p \tag{2.9}$$

The key difference with the quality criterion (2.8) for destructive testing is that the number of items which are not tested is not a constant, but decreases if the number of items tested, n, increases, as their sum is constant. Clearly, if all items in the batch have been tested, so $n = t$, there is no uncertainty left about the total number

of good items. This criterion (2.9), combined with the NPI lower probability (2.5), leads to the requirement that (n, s) are chosen such that

$$\binom{t}{n}^{-1} \sum_{j=r}^{t-n} \binom{s-1+j}{j} \binom{t-s-j}{t-n-j} \geq p \tag{2.10}$$

Requirement (2.10) does not generally lead to a simple closed form expression for the sampling schemes (n, s) that satisfy quality criterion (2.9). However, for one particular situation this criterion leads to an interesting result, namely when we want all items in the batch to be good, hence we set $s = n$ and $r = m = t - n$. For this extreme requirement, criterion (2.8) implies that the minimum required sample size should satisfy

$$n \geq pt \tag{2.11}$$

so at least proportion p of the total batch must be tested, and obviously none of these is allowed to fail for this criterion to be met. This result is easily derived from (2.10), but also follows from the fact that the NPI lower probability for item $k+1$ to be good, given all k items tested before are good, is equal to $k/(k+1)$, and applying this with sequential conditioning (so for k from n to $t-1$) on items being good leads to the NPI lower probability for the event that the next $t - n$ items are all good, given that the first n items were good, to be equal to $\prod_{l=0}^{t-n-1} \frac{n+l}{n+l+1} = \frac{n}{t}$. Clearly, for this lower probability to be at least p, we require minimum sample size pt. We consider this to be a nice result, which is particularly elegant when the possible interpretation of lower probability in terms of betting behaviour is considered (see Section 1.2). For example, let us consider a batch of size $t = 100$. According to this NPI approach when combined with the betting behaviour for lower and upper probabilities, one would be willing to pay up to p for the gamble which pays 1 if all items in the batch are good and 0 otherwise, if $100p$ items have been tested and these were all good. From this perspective, also note that on the basis of any $n \geq 1$ tests which are all good, one would never bet against all t items being good according to the NPI theory. This is also nicely in line with the possible interpretation of lower probabilities as representing evidence in favour of the event considered, whereas upper probabilities represent evidence against this event.

Example 2.4

Table 2.6 presents minimum required sample sizes for non-destructive testing of items from a batch of total size $t = 100$, and with high quality criteria to be satisfied, in the sense that zero ($s = n, r = 100 - n$), one ($s = n - 1, r = 100 - n$; or $s = n, r = 99 - n$), two ($s = n - 2, r = 100 - n$; or $s = n - 1, r = 99 - n$; or $s = n, r = 98 - n$) or three ($s = n - 1, r = 98 - n$) of the 100 items are allowed to fail in total, for varying minimum required lower probabilities. Either all tested items are assumed to be successful ($s = n$), or at most one or two are allowed to fail ($s = n - 1$ or $s = n - 2$). It is clear from these required sample sizes that, in order to meet high quality criteria, one must test a large proportion of the total batch, with the required number of items tested reduced substantially if one allows one or two failing items among the not-tested items in the batch. On the other hand, allowing one or two failing items among those tested increases the required sample sizes substantially.

	$s =$	n	$n-1$	$n-2$	n	$n-1$	n	$n-1$
p	$r =$	$100-n$	$100-n$	$100-n$	$99-n$	$99-n$	$98-n$	$98-n$
0.50		50	71	80	30	50	21	39
0.70		70	84	89	45	64	33	51
0.80		80	90	93	55	72	42	58
0.90		90	95	97	69	81	54	68
0.95		95	98	99	78	87	63	75
0.98		98	99	100	86	92	73	82
0.99		99	100	100	90	94	78	86

Table 2.6: Minimum required sample sizes n, with $t = 100$

2.6 Two-stage acceptance sampling

Two-stage acceptance sampling, also called double sampling [34], is illustrated in Figure 2.2. It uses chosen proportions p_1 and q_1 for either accepting or rejecting the batch after the first stage. Suppose that one wants to accept the whole batch

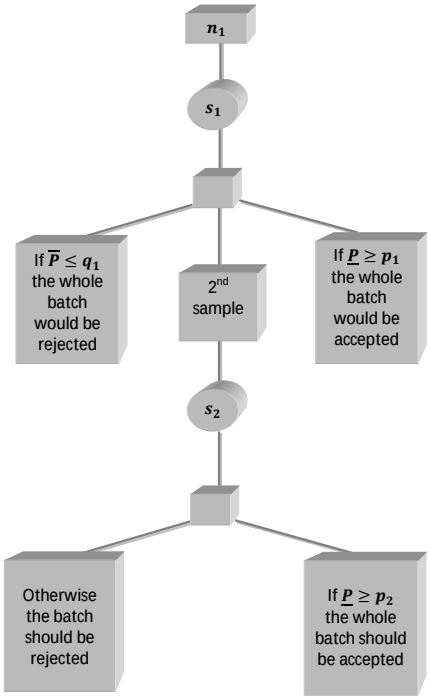

Figure 2.2: Sampling Procedure

provided that the lower probability for the event that at least r out of m future items function, given the number of functioning items in the first stage, is not less than a specific proportion, say p_1. If $\underline{P} < p_1$ the batch is not accepted at this stage, and the upper probability $\overline{P}$ must be checked. If $\overline{P} \leq q_1$ (the ratio to reject the batch) the batch is rejected, otherwise another sample of size n_2 is selected. Thereafter, if $\underline{P} \geq p_2$, the decision maker accepts the batch under consideration, else the batch is rejected. Clearly, in this section of the study it is more interesting to consider the upper probability as a criterion for making a decision for the first time.

Before testing takes place, one needs to decide on a suitable number of items to be tested in the first stage, with the knowledge that, if the test results do not clearly indicate a final decision (accept or reject the batch of m items), one can test a further n_2 items before making the final decision. With uncertainty quantified by lower and upper probabilities, a natural criterion for accepting the batch of m items, after the first n_1 tests have led to s_1 successes, is

$$\underline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} = s_1) \geq p_1 \tag{2.12}$$

for some chosen values of r, s_1 and p_1. In this NPI approach, the following equality holds (see Section 2.3.1),

$$\underline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} \geq s_1) = \underline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} = s_1) \tag{2.13}$$

which, as the lower probability in (2.12) is increasing in s_1, ensures that once n_1, s_1 and p_1 have been chosen before the tests take place, the required quality is certainly achieved if the actual number of successes in the n tests is larger than s_1. If criterion (2.12) is satisfied, that is if the observed number of items that function successfully in the test is at least s_1, then this would indicate that the quality of the first n_1 tested items is high enough to accept the batch of m items without further testing.

An alternative final decision after n_1 items have been tested is to reject the batch of m items. It is logical to take this decision if there are more failed items in the test than is judged to be acceptable, in such a manner that it indicates that it is unlikely that the batch of m items will have sufficient good items. A convenient way to express this, if testing n_1 items at the first stage has revealed s_1 successes, is by use of the criterion

$$\overline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} = s_1) \leq q_1 \tag{2.14}$$

for some chosen values of r and q_1. Notice here the use of the NPI upper probability, which can be interpreted as only suggesting to reject the batch after the first stage if the evidence against the event $Y_{n_1+1}^{n_1+m} \geq r$ is considered strong enough to make this event unlikely. As before, the choice of the values r, s_1 and q_1 will be required before testing takes place if they are to guide the choice of n_1, at which stage the criterion used will actually be

$$\overline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} \leq s_1) \leq q_1 \tag{2.15}$$

Similar to Equality (2.13) for the NPI upper probabilities the following equality holds

$$\overline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} \leq s_1) = \overline{P}(Y_{n_1+1}^{n_1+m} \geq r | Y_1^{n_1} = s_1) \tag{2.16}$$

It is interesting to compare Inequality (2.15) to Inequality (2.7). In both cases, the upper probability based on data in the form of a bound on the number of

successes, is equal to the upper probability based on data taking on the precise value which satisfies the bound and is most supportive for the event of interest. It should be emphasized that, throughout this NPI approach to acceptance sampling, test information of the form $Y_1^{n_1} \geq s_1$ or $Y_1^{n_1} \leq s_1$ is considered in order to define the methods, so before data are available. Once test results are available, of course the exact number of successes is used.

Taking into account that the upper probability in (2.14) is increasing in s_1, an actual decision to reject the batch after at most s_1 successful units in the first n_1 tested items, is supported by an upper probability of at most q_1.

In this setting, a final decision to accept the batch after the first stage of testing is based on a minimum required value for the lower probability for the event that at least r items will function successfully in the batch of size m, while a final decision to reject the batch at that stage is based on a maximum allowed value for the upper probability for that event. Of course, the choice of p_1 and q_1 should be considered with care as discussed later in this section, in all but rather trivial situations they will be such that, for at least some possible test results in the first stage, either the final decision to accept or the final decision to reject the batch of m items will be made after the first stage of testing. If the final decision is not made after the first stage of testing, then the decision will be to test n_2 further items and base the final decision on the combined results of the $n_1 + n_2$ tested items.

The key results for NPI for Bernoulli random quantities, namely the upper probability (2.3) and lower probability (2.4), are based on the data with only few further assumptions, and related inferences are not affected by a specific stop-criterion used for sampling or similar considerations, as long as one is happy with the exchangeability assumption for data and future observations that is implicit to Hill's assumption $A_{(n)}$ on which NPI is based [7,28]. This occurs similarly in Bayesian statistics, where it is also known as the 'likelihood principle', but in classical frequentist statistics, in particular when hypothesis tests are considered, the situation is different as data that could have occurred but did not play a role in such inferences, and such other possible data from the experiment depend on the specific stop-criterion used.

An advantage of NPI is that its central inferential assumption, Hill's $A_{(n)}$, is

explicitly a post-data assumption, which allows study of the data before the actual predictive inferences, in order to judge whether or not $A_{(n)}$ is deemed reasonable. As an example, suppose that one tests 10 items sequentially, and the first 5 are all good followed by 5 failures. Although this is not in itself an argument against the use of the NPI method for Bernoulli quantities, as used in this study, one may wish to study in a bit more detail whether or not something could have changed halfway through the testing process. As always in statistics, it is impossible to exclude any subjectivity in the modelling assumptions, and in NPI this is with regard to the post-data exchangeability assumption $A_{(n)}$ for the future observations and the data.

In addition, and again similar to Bayesian statistics, one can apply these same upper and lower probabilities if the total number of observations are $n_1 + n_2$, no matter whether the data were collected all together, or first n_1 data were observed followed by n_2. Hence, generalizing the results for single-stage sampling in the previous section to multi-stage sampling is conceptually trivial, and it might appear to be attractive to just perform one test at a time, and stop whenever an appropriate criterion has been satisfied. However, this might undermine the exchangeability assumption implicit in the NPI approach, so care must be taken when doing so.

There may be situations in which it is important to determine the number of items to be tested beforehand, for example if testing takes much time or is expensive. In such situations, the NPI approach is relatively weak, as it cannot give a strong indication on the number of tests that should be performed, which is a direct consequence of only making few modelling and inferential assumptions. Nevertheless, it may be attractive to consider a multi-stage procedure, in which first n_1 items are tested, in a similar setting as discussed in Section 2.4 (we only consider destructive testing in this section), and depending on the test results the batch of size m is either accepted or rejected, or a further n_2 items are tested. If attention is restricted to two-stage acceptance sampling, the decision after this second stage, so with a total of $n_1 + n_2$ items tested, will be either to accept or reject the batch. This can straightforwardly be generalized to further stages, the discussion here is restricted to two stages.

The minimum required number of items to be tested at the first stage, in order

to have the possibility of reaching the final decision to accept the batch of m items without a second stage of testing, is the minimum value of n_1 for which criterion (2.12) is satisfied with $s_1 = n_1$. This is intuitively logical, as a test result without failing items is the most favourable information for this event of interest. As mentioned earlier, when single-stage testing was discussed, this is satisfied if

$$n_1 \geq \frac{p_1 m}{1 - p_1} \tag{2.17}$$

Similarly, the minimum required number of items to be tested at the first stage in order to reject the batch of m items without further testing, is derived by setting $s_1 = 0$ in criterion (2.14), as a test that reveals only failing items is clearly the worst case. This leads to the condition

$$\prod_{j=1}^{m} \frac{j}{n_1 + j} \leq q_1 \tag{2.18}$$

Conditions (2.17) and (2.18) might provide some insights into appropriate choice of n_1, as if neither of these two conditions are satisfied then the second stage of testing would always be required. This is pretty weak advice if guidance is required on choice of an appropriate value for n_1, but the NPI approach does not provide stronger support for this choice.

Generally, in respect of the second sampling stage, the lower probability can be defined as function of (n_1, n_2, m). First let us assume that all items in both selected samples are functioning well, so $s_1 = n_1$ and $s_2 = n_2$. If we are interested in having all m future items functioning, so $r = m$, then the lower probability of interest is

$$\begin{aligned} \underline{P}(Y_{n_1+n_2+1}^{n_1+n_2+m} = m | Y_1^{n_1+n_2} = n_1 + n_2) =& \binom{n_1 + n_2 + m}{m}^{-1} \binom{n_1 + n_2 + m - 1}{m} \\ =& \frac{n_1 + n_2}{n_1 + n_2 + m} \end{aligned} \tag{2.19}$$

In order to find the minimum required sample size n_2 for accepting the batch if $s_1 = n_1$, this general form (2.19) can be used, hence $\frac{n_1+n_2}{n_1+n_2+m} \geq p_1$, so $n_2 \geq \frac{p_2}{1-p_1} m - n_1$.

If all inspected items in the first selected sample were functioning well, $s_1 = n_1$, and we wish to have all m future items to function well, then upper probability will be

$$\overline{P}(Y_{n_1+1}^{n_1+m} = m | Y_1^{n_1} = n_1) = \binom{n_1+m}{n_1}^{-1} \binom{n_1+m}{n_1} = 1 \tag{2.20}$$

which means that the upper probability $\overline{P}$ reaches to the peak with $s_1 = n_1$ and $r = m$. Thus, the batch would never be rejected. But if we had only one failing item in n_1, $s_1 = n_1 - 1$.

$$\overline{P}(Y_{n_1+1}^{n_1+m} = m | Y_1^{n_1} = n_1 - 1) = \binom{n_1+m}{n_1}^{-1} \binom{n_1+m-1}{n_1-1} = \frac{n_1}{(n_1+m)} \leq q_1 \tag{2.21}$$

then the sample size n_1 for rejecting the batch in this case is $n_1 \leq \frac{q_1}{(1-q_1)} m$. If we did not have any successful item in the first sample, $s_1 = 0$, then the upper probability for the event that all future items will function, $r = m$, would be

$$\overline{P}(Y_{n_1+1}^{n_1+m} \geq m | Y_1^{n_1} = 0) = \binom{n_1+m}{n_1}^{-1} = \frac{n_1! m!}{(n_1+m)!} \leq q_1 \tag{2.22}$$

Example 2.5

Table 2.7 shows a numerical example to illustrate the idea of two-stage acceptance sampling, where D refers to our decision, which is either accept (A) or reject (R) the batch of $m = 10$ items. If the results of the first sample of size n_1 are not conclusive with regard to accepting or rejecting (i.e. $\underline{P} < p_1$ and $\overline{P} > q_1$), a second sample of size n_2 is taken. To make it clear, suppose we set as criterion that, with lower probability 0.8, at least $r = 8$ items out of a batch of $m = 10$ items function in the process of interest after testing, and we need to base our decision on whether or not to accept a batch on test results of items exchangeable to these 10. Suppose further that we can get 10 items tested at the first testing stage, and if the results prove to be inconclusive we can test a further 10 items, after which the final decision to accept or reject the batch of $m = 10$ items is required. So, after the first stage of testing, criterion (2.12) with $p_1 = 0.8$ is applied for the decision to accept the batch directly, and we suppose further that the decision to immediately reject the batch after the first stage of testing is based on criterion (2.15) with $q_1 = 0.5$. Then, if the outcome of the first stage testing of 10 items is inconclusive according to these two criteria, stage two of testing is applied, with a further 10 items tested. In this case,

exchangeability is assumed for all 30 items involved, so the 20 tested items and the 10 in the batch of interest. Let us, after this test, again assume that we accept the batch of $m = 10$ items if the lower probability for at least $r = 8$ items functioning is at least 0.8, now of course based on the information of the 20 items tested. Hence, the acceptance criterion after stage 2 becomes $\underline{P}(Y_{21}^{30} \geq 8 | Y_1^{20} = s_1 + s_2) \geq 0.8$, with s_2 the number of items that function successfully in the second-stage test. Of course, the random quantity involved has now become Y_{21}^{30}, reflecting that we are interested in functioning of 10 future items following tests on 20 items.

In this setting, suppose first that all 10 items in stage one of testing functioned successfully. As $\underline{P}(Y_{11}^{20} \geq 8 | Y_1^{10} = 10) = 0.89 \geq 0.8$, we accept the batch of 10 items without further testing. If, instead, there was one item that failed this test, then the test was inconclusive, as $\underline{P}(Y_{11}^{20} \geq 8 | Y_1^{10} = 9) = 0.71 < 0.8$ and $\overline{P}(Y_{11}^{20} \geq 8 | Y_1^{10} = 9) = 0.89 > 0.5$, and therefore the second stage of testing will be used for a further 10 items. Suppose now that, in this second stage, all 10 tested items function, so in total 19 out of 20 tested items functioned. This leads to $\underline{P}(Y_{21}^{30} \geq 8 | Y_1^{20} = 19) = 0.90 \geq 0.8$, and now the batch of 10 items, which have not been tested but are exchangeable with the 20 that were tested, will be accepted. If there had been 9 functioning items in the second-stage test, then the batch would also have been accepted as $\underline{P}(Y_{21}^{30} \geq 8 | Y_1^{20} = 18) = 0.81 \geq 0.8$, but for any smaller number of good items it would have been rejected. It is nice to note here that 9 successful items out of 10 was not deemed to be sufficiently strong evidence to decide to accept the batch after the first stage of testing, but 18 successful items out of 20 is (just) sufficient to accept the batch. In Table 2.7 the results are given for some other values for r, s_1 and s_2 with the same considered p_1, q_1 and p_2 also.

$r = m - 2 = 8$						
s_1	$\underline{P}$	$\overline{P}$	D	s_2	$\underline{P}$	D
$s_1 = n_1 = 10$	0.89	-	A			
$s_1 = n_1 - 1 = 9$	0.71	0.89	n_2	$s_2 = n_2 = 10$	0.90	A
				$s_2 = n_2 - 1 = 9$	0.81	A
$s_1 = n_1 - 2 = 8$	0.50	0.71	n_2	$s_2 = n_2 = 10$	0.81	A
				$s_2 = n_2 - 1 = 9$	0.69	R
$r = m - 1 = 9$						
s_1	$\underline{P}$	$\overline{P}$	D	s_2	$\underline{P}$	D
$s_1 = n_1 = 10$	0.76	1	n_2	$s_2 = n_2 = 10$	0.89	A
$s_1 = n_1 - 1 = 9$	0.76	0.76	n_2	$s_2 = n_2 = 10$	0.74	R
				$s_2 = n_2 - 1 = 9$	0.59	R
$s_1 = n_1 - 2 = 8$	0.29	0.50	R			
$r = m = 10$						
s_1	$\underline{P}$	$\overline{P}$	D	s_2	$\underline{P}$	D
$s_1 = n_1 = 10$	0.50	1	n_2	$s_2 = n_2 = 10$	0.67	R
				$s_2 = n_2 - 1 = 9$	0.43	R
$s_1 = n_1 - 1 = 9$	0.23	0.50	R			
$s_1 = n_1 - 2 = 8$	0.10	0.24	R			

Table 2.7: Two stage acceptance sampling with $n_1 = n_2 = m = 10$, $p_1 = p_2 = 0.80$ and $q_1 = 0.50$

Example 2.6

This example illustrates the methods of Sections 2.4, 2.5 and 2.6 for a large batch of size $m = 200$ future items. In case of destructive tests, as presented in Section 2.4, Table 2.10 shows the minimum required sample size n, which should be tested with different quality requirements p, r and s. For example, with an acceptable quality level of $p = 0.50$, at least 781 items need to be tested if one will accept the batch if two or fewer items fail, $s = n - 2$, and one wishes all 200 items in the batch to function. While a minimum required lower probability of $p = 0.95$ in this case requires at least $n = 29610$ items to be tested.

Tables 2.8 and 2.9 present the minimum required sample size n if the test was not with destruction, as presented in Section 2.5. Suppose that one requires all 200 future items to function, $r = 200 - n$, if the sample of size n contains not more than two failed items, $s = n - 2$, with $p = 0.50$, then at least 100 items must be tested. If we increase p to 0.95, at least 198 items have to be tested. However, when one reads down the columns in Tables 2.10 and 2.8, we see the effect of the required quality level p, how many successful items in the tested items s and in the future batch r, and the minimum required size of our sample to accept the batch under inspection batch. When one reads across the tables below we see how quality affects the sample size, as the sample size increases when the quality requirement is higher.

On the other hand, Table 2.11 shows the minimum required sample size n_2 to accept the batch in the second stage. If we consider n_1 as in Table 2.10 in case $s = n$ and $r = 198$ with different p, the required sample sizes n_2 are given in Table 2.11. As can be seen the total combined sample sizes required to accept the batch is still the same. In other words, in the single sampling plan, the batch of size 200 can be accepted if a sample of size 200 has been tested with $s = n$, $r = 200$ and $p = 0.5$, but if we tested only 52 items the batch should be rejected then a second sample n_2 is selected. To accept the batch in this stage 148 items have to be tested $52 + 148 = 200$. While, if $s = n - 1$, $r = m - 1 = 199$ and $p = 0.9$, 820 items should be tested in the first stage, while if we tested 230 in that stage, $820 - 230 = 590$ items should be tested as a second sample n_2. For all cases, the sample sizes in case of $s = n$ and $r = m - 2 = 198$ in Table 2.10 are used as n_1 with different p.

	$s=$	n	$n-1$	$n-2$	n	$n-1$	n	$n-1$
p	$r=$	$200-n$	$200-n$	$200-n$	$199-n$	$199-n$	$198-n$	$198-n$
0.50		100	142	160	60	100	42	78
0.60		120	156	166	78	116	60	90
0.70		140	168	178	90	128	66	102
0.80		160	180	186	110	144	84	116
0.90		180	190	194	138	162	108	136
0.95		190	196	198	156	174	126	150
0.98		196	198	200	172	184	146	164
0.99		198	200	200	180	188	156	172

Table 2.8: Minimum required sample sizes n, with $t = 200$

	$s=$	n	$n-1$	$n-2$	n	$n-1$	n	$n-1$
p	$r=$	$180-n$	$180-n$	$180-n$	$150-n$	$150-n$	$110-n$	$110-n$
0.50		7	15	23	3	7	2	4
0.60		9	18	26	4	8	2	4
0.70		11	21	30	5	9	2	5
0.80		15	26	35	6	11	3	6
0.90		20	32	42	8	14	4	7
0.95		26	39	49	10	16	5	9
0.98		33	46	56	13	20	7	10
0.99		38	51	62	15	22	8	12

Table 2.9: Minimum required sample sizes n, with $t = 200$

p	$s=n$ $r=200$	$s=n-1$ $r=200$	$s=n-2$ $r=200$	$s=n$ $r=199$	$s=n-1$ $r=199$	$s=n$ $r=198$	$s=n-1$ $r=198$
0.50	200	484	781	83	200	52	126
0.70	467	1030	1600	165	351	99	207
0.80	800	1700	2610	247	496	142	279
0.90	1800	3710	5610	432	820	230	423
0.95	3800	7710	11610	693	1275	342	602
0.98	9800	19710	29610	1212	2175	535	911
0.99	19800	39710	59610	1796	3120	725	1215

Table 2.10: Minimum required sample sizes n with $m = 200$

p	$s=n$ $r=200$	$s=n-1$ $r=200$	$s=n-2$ $r=200$	$s=n$ $r=199$	$s=n-1$ $r=199$	$s=n$ $r=198$	$s=n-1$ $r=198$
0.50	148	432	729	31	148	0	74
0.70	368	931	1501	66	252	0	108
0.80	658	1558	2486	105	354	0	137
0.90	1570	3480	5380	202	590	0	193
0.95	3458	7368	11268	351	933	0	260
0.98	9265	19175	29075	677	1640	0	376
0.99	19075	38985	58885	1071	2395	0	490

Table 2.11: Minimum required sample sizes n_2 with $m = 200$ and predetermined n_1

2.7 Concluding remarks

The nonparametric predictive approach to basic acceptance sampling problems, as presented in this chapter, differs from more established frequentist methods as the emphasis is explicitly on random quantities representing functioning of actual items, instead of statistical tests and related inferences on model parameters. The Bayesian approach also offers the option of predictive inference, yet there one would condition explicitly on the event $Y_1^n \geq s$, in which case all these possible values for the numbers of successes in the n tests are taken into account, with the prior distribution over these values still influencing the inferences. The approach presented in this chapter, as indeed NPI in general [8], aims explicitly at reduction of modelling and other assumptions such as prior information, while still allowing some useful inferences. The use of lower and upper probabilities to quantify uncertainty supports this aim, and leads to the attractive opportunity to focus on lower probabilities for quality criteria, which can be considered to include some robustness in the inferences. The price one seems to pay for such robustness is the requirement, in most cases, of a large number of tested items, with few failures, in order to decide to accept a sample. However, if one considers that all information underlying these inferences comes from the test data, this should not be surprising if one really wishes ambitious quality criteria to be met, and if these criteria are formulated in terms of the quality of the actual items that will be used and that have not yet been tested.

The NPI approach presented in this chapter, in line with NPI in general, does not make use of prior assumptions on the data that will be observed or on specific features of a population distribution. The central $A_{(n)}$ assumption has an explicitly post-data nature, meaning that one judges whether or not it is deemed acceptable after the data have appeared. Although this provides attractive inferences which are little influenced by assumptions, as a consequence one cannot derive strong guidance on required numbers of items to be tested, which is possible when using classical frequentist or Bayesian statistical methods. As it is often difficult to judge, for the latter more established methods, the influence of additional modelling assumptions, there is a strong case for simultaneously using NPI and several other methods, where the latter can provide guidance on sample size considerations, while, once the data

are available, NPI can be used either directly for the inferences of interest, or to judge whether inferences resulting from the alternative methods are significantly influenced by the additional modelling assumptions underlying those methods. Such other methods also enable considerations, as part of the process to determine suitable sample sizes for testing, of so-called producer and consumer risks, related to the possibilities that a batch which is of sufficient quality gets rejected and that a batch which does not meet the quality criterion gets wrongly accepted. Again, such considerations have not yet been considered in the NPI approach. Exploration of how such considerations can be taken into account at later stages in multi-stage testing is an interesting challenge for future research.

Chapter 3

Sequential Acceptance Decisions

3.1 Introduction

In this chapter we present a new heuristic approach to sequential acceptance problems which generalize the well-known secretary problem, also known as the marriage problem. We assume that there are $N \geq 2$ candidates from which one wishes to select c candidates, with $1 \leq c \leq N-1$. We assume that all candidates can be ranked, without ties, with regard to a criterion of interest, such that the best candidate has rank 1 and the worst candidate rank N. The candidates are observed in a random sequence, with all sequences equally likely, and upon observation the relative rank of a candidate with regard to all previously observed candidates becomes known without observational error. It should be emphasized that no further information is assumed to become available about candidates, for example no real-valued measurement or score related to the quality of the candidate will become available. Actually, as we will comment on in Subsection 4.2.2 when discussing future research topics , such additional information would not necessarily change the core of our heuristic method significantly. In Sections 3.2, 3.3 and 3.4 we will assume that candidates are observed one at a time, in Section 3.5 this is generalized to simultaneous observation of subgroups of candidates. Upon observation of candidates it must be decided immediately whether to accept or discard the candidate, it is assumed throughout the chapter to be impossible to discard accepted candidates, or accept discarded candidates, at a later stage. Of course, the aim is to select c good candidates, and

it is important to consider how 'good' can be measured.

Sequential acceptance problems have attracted much attention in the literature. In particular, the special case with $c = 1$, known by a variety of names such as 'secretary problem' and 'marriage problem', has been solved by Lindley [33] using dynamic programming, both under the explicit assumption that one is only interested in selecting the candidate with rank 1 and the case with a loss function that is linear in the rank of the selected candidate.

Freeman [23] presented an extensive overview of the literature till the early 1980's, including attention to aspects such as uncertain acceptance of a candidate, recall opportunities, unknown total number of candidates, aiming for the minimum expected rank of the selected candidate and more general loss or utility functions. Freeman also discusses contributions in which more than one choice is made, with a variety of criteria. In particular, Henke [27] considered selection of candidates with the aim to minimise the sum of their actual ranks, which we also aim for in this study. Henke proved that a candidate should be accepted if his relative rank, compared to the ranks of the candidates already observed, is below a threshold value which depends on the total number of candidates, the number of candidates required, the number of candidates already accepted, and the stage in the process. Henke provides a system of recurrence relations that determine these threshold values. In principle, this problem can be formulated as a dynamic programming problem and hence the optimal selection strategy can be determined, so it seems that the problem is solved. However, once N and c are not small, the number of paths included in the dynamic programming formulation is extremely large, making computation of the optimal strategy impossible.

These sequential acceptance problems have also been studied from different perspectives, with probability theory typically used to study performance of some rather basic rules, focusing on topics such as the expected number of candidates to be observed until one has accepted c candidates for a specific rule. For example, Krieger *et al* [31] present probabilistic properties for several rules including a percentage rule, where a candidate is accepted if his relative rank at time of observation belongs to a fixed percentage of best ranks among those already selected, for example

if the percentage is set at 50% it implies that acceptance of a candidate improves the median of the ranks, at that moment, of all selected candidates. Krieger, Pollak and Samuel-Cahn [31] study properties such as the expected number of candidates selected at any stage in the process using such fixed percentage rules. In Section 3.6, we compare our new heuristic approach with a variation to this fixed percentage rule.

In this chapter, we propose a new heuristic method for solving such sequential acceptance problems, which will not generally provide the theoretically optimal solution but which is computationally very straightforward and also quite flexible for variations such as candidates arriving in groups. The approach is related to the frequentist statistical framework of nonparametric predictive inference (NPI) [3, 8], and the performance of the resulting strategies is investigated via simulations. We introduce our NPI-based heuristic method in Section 3.2. In Section 3.3 we investigate our new method for the basic case where only one candidate is to be selected. In Section 3.4 we generalize this to selection of $c \geq 1$ candidates. Whilst in Sections 3.2, 3.3 and 3.4 we assume that candidates are observed sequentially one at a time, in Section 3.5 we consider the situation with groups of candidates being observed together, with the acceptance decisions for all members of such a group made simultaneously. We will in particular consider whether or not it is always an advantage to see multiple candidates in groups. In Section 3.6 we compare our approach to an alternative heuristical method which uses a fixed-percentage. Another comparison is considered in Section 3.7, to namely methods that improve the subset of accepted candidates, where a candidate is selected only if his relative rank is not more than the median or the mean of the relative ranks of the already selected. In Section 3.8 we present a related alternative approach with a randomised decision process, where the value of the threshold p is chosen randomly, we also compare this method to our NPI heuristic approach. We end this chapter with some concluding remarks in Section 3.9.

3.2 NPI-based heuristic approach

We consider the scenario described in Section 3.1, with N candidates observed one at a time, of which we must select c candidates. Consider stage j, that is we are observing the j-th candidate and learn his relative rank r_j compared to the previous $j-1$ candidates, so $r_j \in \{1, \ldots, j\}$. Suppose that from the previous $j-1$ candidates a have been accepted, with $0 \leq a < c$, of course the process stops once c candidates have been selected. As we will see in Section 3.4, this does not mean that the remaining candidates can be totally neglected, as their ranks are still required in order to determine the ranks among all N candidates of the c selected candidates, these ranks among all candidates will be called 'absolute ranks'. We denote the absolute rank of candidate j by R_j, where it is crucial to emphasize that R_j will only be known to us, when going through the process of selecting candidates, once all N candidates have been observed.

The heuristic method proposed in this chapter is based on the probability of the event that among the remaining $N - j$ candidates, there are at least $c - a$ who are better than the current j-th candidate, so their absolute ranks are smaller than the absolute rank R_j of the current candidate. Let $Y_{j+1}^N(r_j)$ denote the number of candidates following the j-th candidate with relative rank r_j, who have lower absolute rank than him, then the probability of interest to us, and on which our heuristic method is based, is

$$P(Y_{j+1}^N(r_j) \geq c - a) = \binom{N}{j}^{-1} \sum_{l=c-a}^{N-j} \left[\binom{r_j - 1 + l}{l} \binom{N - r_j - l}{N - j - l} \right] \tag{3.1}$$

This probability follows by direct combinatorial arguments based on the fact that all sequences of candidates, and hence of their absolute ranks, are assumed to be equally likely, based on the $A_{(n)}$ assumption applied to multiple future observations. Note that for the events considered in this chapter, $A_{(n)}$ leads to precise probabilities. The term within the summation in (3.1) is easily seen to correspond to the case where there are precisely l future candidates with absolute rank lower than that of the j-th candidate, who is currently being considered, as it counts all combinations with l of the $N - j$ future candidates' absolute ranks being smaller, and the remaining $N - j - l$ of these absolute ranks being greater than the absolute rank of the j-th

candidate.

It should be emphasized that probability 3.1 is a direct predictive probability for the future $N - j$ observations, based on the first j observations, and is not a probability resulting from arguments of sampling variation as traditionally used in statistics. As such, this probability (3.1) fits in the NPI framework of statistics [3,8].

The heuristic solution for such a sequential acceptance problem proposed in this chapter is as follows. We choose a threshold probability value p, $0 < p < 1$, and we decide to *accept* the j-th candidate if

$$P(Y_{j+1}^{N}(r_j) \geq c - a) < p \tag{3.2}$$

and to *reject* him if

$$P(Y_{j+1}^{N}(r_j) \geq c - a) \geq p \tag{3.3}$$

So, the current candidate is rejected if the probability that there are at least $c - a$ better candidates among the future $N - j$ candidates is at least p, so if it is deemed likely (compared to choice of value of p) that there are sufficient stronger candidates yet to come, and he is accepted if this probability is smaller than p. This implies that for smaller values of p there is a tendency to accept fewer candidates early on, while for larger values of p earlier candidates are more likely to be accepted. In the limit case with $p \downarrow 0$ one ends up accepting only the final c candidates, while the limit case with $p \uparrow 1$ leads to the first c candidates being accepted.

It is important to study which values of p give good performances of our heuristic method, this is one of the topics addressed in the simulation studies in the following sections. In addition to possible arguments based on symmetry, one could argue in favour of setting $p = 1/2$ by considering the following scenario. Suppose we have to decide on acceptance of candidate $j = N - 1$ with rank r_{N-1}, and $c - a = 1$ so we must select one more candidate. Clearly, the random relative rank of the final candidate has equal probability $1/N$ to be any value $1, 2, \ldots, N$, and the probability that the absolute rank of candidate N is less than the absolute rank of candidate $N - 1$ is equal to r_{N-1}/N. Hence, setting $p = 1/2$ leads to candidate $N - 1$ being accepted if his relative rank r_{N-1} is less than $N/2$, else the final candidate will be accepted, which gives equal chances to each of the final two candidates to be

selected (neglecting the arbitrariness of the choice of the strict inequality in either (3.2) or (3.3), so what to do if $r_j = N/2$). While in most of this chapter we will keep p constant for all j, we also in Subsection 3.3.2 briefly explore varying p. The heuristic method presented in this chapter is different from percentage rules or other methods that have been proposed in the literature, and as computation of probability (3.1) is straightforward and this method does not require backward optimization as in the dynamic programming approach, the computational effort required to implement this heuristic method is very small, even for large values of N and c.

While the real benefit of this method will become clear in the problems involving selection of multiple candidates, considered in Sections 3.4 and 3.5, we start exploring its use and performance for the simple situation where only a single candidate needs to be selected.

It is mathematically straightforward to include loss functions in this framework, but for practical applications such loss functions may be rather artificial and only of indicative value. We prefer to consider and comment on the overall performance of the methods we propose, and those we compare our approach with, without focus on specific loss functions. When we consider selection of multiple candidates, which is the main reason for developing our heuristic approach, we will focus on the sum of the absolute ranks of the selected candidates, also without further use of loss functions.

All simulation studies reported in this chapter involve 10,000 runs[1], so each specific algorithm is applied to 10,000 cases, each case being a random permutation of the numbers $1, \ldots, N$ representing the absolute ranks of the N candidates. For example, for $N = 4$ one such possible permutation would be $(3, 4, 1, 2)$, in which case the absolute rank of the first candidate is $R_1 = 3$, of course he has relative rank $r_1 = 1$ as is always the case for the first candidate. The second candidate has $R_2 = 4$ and $r_2 = 2$. The third candidate has $R_3 = 1$ and $r_3 = 1$, and finally the fourth candate has $R_4 = 2$ and $r_4 = 2$. It should be emphasized that only the relative ranks r_j are observed, the absolute ranks R_j are only known once all N

[1]The statistical software R was used

candidates have been observed. We only report on results of a single simulation of 10,000 runs for each case. Such simulations were repeated several times for each scenario, but the results showed very little variation so the presented cases give a clear description of the performances of the considered approaches in the specific scenarios. Different scenarios were also studied, the main conclusions from those were in line with those from the cases reported here. If several procedures (e.g. different values of p in our approach) are compared and the results are presented in a single table or figure, then the same 10,000 runs have been used for all procedures to eliminate the effect of random fluctuations due to different permutations being used.

3.3 Selecting one candidate

Lindley [33] used dynamic programming to solve the basic problem of selecting a single candidate under a 0-1 loss (or utility) function, reflecting the situation that one only wishes to select the best candidate (hence with absolute rank 1). The solution to this problem, which has also been derived by other methods [23], is easily found due to the fact that one is only happy with the best candidate and does not distinguish between all other candidates, hence obviously one only accepts a candidate if his relative rank is 1. The solution is derived by defining $a_j = \sum_{i=j}^{N-1} i^{-1}$ and j^* the unique integer for which $a_{j^*-1} \geq 1 > a_j^*$, then the first $j^* - 1$ candidates are not accepted and the first candidate thereafter who is better than all previous candidates (hence has relative rank 1) is accepted. The probability of selecting indeed the best candidate, following this strategy, is approximately $e^{-1} = 0.368$ for large N. Of course, following this policy one may end up having to accept the final candidate whatever his (relative and absolute) rank.

Lindley [33] also discussed the use of a loss or utility function that is linear in the rank, hence minimising the expected rank of the single accepted candidate, an approach that reflects more closely the view we take in this study. Lindley proved that the optimal procedure is of the, intuitively logical, form to accept a candidate if his relative rank is at most some threshold value which depends on the stage j

of the process and which can be calculated from a system of recurrence equations. Already for this problem with only a single candidate to be selected, solving the recurrence equations is a substantial computational burden, and Lindley and others (see [23]) proposed approximate methods using differential equations to derive these thresholds.

When selecting multiple candidates, the computational burden prevents the dynamic programming approach to be used for situations where N and c are not small, in which case our heuristic approach, with its computational simplicity, becomes attractive, as will be discussed in Section 3.4. We first consider our approach for the problem of selecting a single candidate, both to consider its performance for this basic problem and to get some understanding of the choice of the threshold probability value p.

We consider the case with $N = 10$ candidates of who one candidate must be selected, so $c = 1$. For each of the 10,000 runs in the simulations, our procedure was applied with p taking the values $0.1, 0.2, \ldots, 0.9$ and, additionally, $p = 0.632$ was also included for reasons explained in Subsection 3.3.1. Figure 3.1 presents boxplots for each of these procedures. Table 3.1 gives the detailed results, namely the number of times the selected candidate had each absolute rank, together with the cumulative number of selected candidates, indicated by $\sum$, which is the number of selected candidates with absolute rank at most the given number. Clearly, the procedure functions poorly for small values of p, with best performance for p in about the range 0.5 to 0.8, where the actual best performance depends of course on the chosen quality criterion. With $p = 0.632$ the best candidate is most frequently selected, if interest is generally in aiming at low rank of the selected candidate then also $p = 0.6$ and $p = 0.7$ lead to very good performance of our method, with the results for $p = 0.5$ and $p = 0.8$ not far behind. The cumulative numbers in Table 3.1 show that, while $p = 0.8$ does not lead to the best candidate being selected as often as for $p = 0.4$ to $p = 0.7$, it performs very well if one would be happy selecting someone with absolute rank not greater than 6. Actually, if one would be happy selecting any candidate except the worst one (with absolute rank 10), then $p = 0.9$ performs best as this never led to selecting the worst candidate in our 10,000 runs.

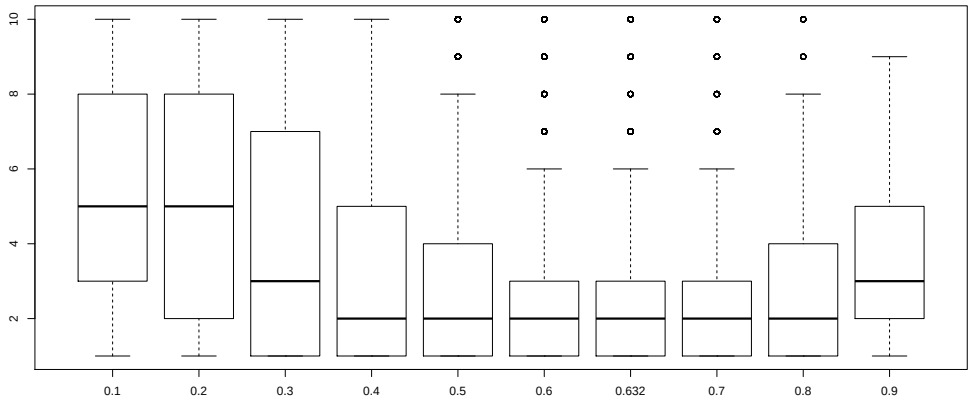

Figure 3.1: NPI for $N = 10, c = 1$

	Absolute rank									
	1	2	3	4	5	6	7	8	9	10
$p = 0.1$	1042	1017	1036	971	1001	899	1002	990	1046	996
$\sum$	1042	2059	3095	4066	5067	5966	6968	7958	9004	10000
$p = 0.2$	1890	1043	934	870	879	796	895	876	934	883
$\sum$	1890	2933	3867	4737	5616	6412	7307	8183	9117	10000
$p = 0.3$	2630	1781	939	653	675	613	669	670	694	676
$\sum$	2630	4411	5350	6003	6678	7291	7960	8630	8324	10000
$p = 0.4$	3002	2207	1559	709	449	379	414	413	424	444
$\sum$	3002	5209	6768	7477	7926	8305	8719	9132	9556	10000
$p = 0.5$	3426	2195	1440	965	541	281	282	279	282	309
$\sum$	3426	5621	7061	8026	8567	8848	9130	9409	9691	10000
$p = 0.6$	3336	2519	1715	999	576	317	131	128	130	149
$\sum$	3336	5855	7570	8569	9145	9462	9593	9721	9851	10000
$p = 0.632$	3486	2527	1665	960	500	367	262	67	76	90
$\sum$	3486	6013	7678	8638	9138	9505	9767	9834	9910	10000
$p = 0.7$	3213	2585	1750	1138	653	334	174	46	51	56
$\sum$	3213	5798	7548	8686	9339	9673	9847	9893	9944	10000
$p = 0.8$	2802	2411	1882	1308	826	463	219	60	14	15
$\sum$	2802	5213	7095	8403	9229	9692	9911	9971	9985	10000
$p = 0.9$	2191	1982	1694	1394	1062	772	498	296	111	0
$\sum$	2191	4173	5867	7261	8323	9095	9593	9889	10000	10000

Table 3.1: NPI approach for $N = 10$ and $c = 1$

Further insights into this selection process are provided by Table 3.2, which gives at each possible stage of the process the maximum relative rank of a candidate who will be selected, still with $N = 10$ and $c = 1$ so obviously given that no earlier candidate has been selected, hence using $a = 0$ in the probability (3.1). This shows that for $p = 0.1$, always the final candidate gets selected, with possible absolute rank $1, \ldots, 10$ and probability $1/10$. For $p = 0.9$, it is possible that the worst candidate gets selected if each r_j, for $j = 2, \ldots, 9$, exceeds the largest value at which the corresponding candidate j would be selected, and of course it requires $r_{10} = 10$. It is easy to check that this occurs for 216 out of the 10! possible and equally likely permutations representing the absolute ranks of the sequence of candidates, the most obvious one being the case with candidates arriving precisely in the order of the absolute ranks, so $j = R_j$. In this case with $p = 0.9$, the probability that the worst candidate gets selected in a single run is 5.95×10^{-5}, hence in one simulation study with 10,000 independent runs the probability of none of these runs leading to the worst candidate being selected is about 0.55 (derived by approximating the Binomial distribution by the Poisson distribution with parameter 0.595). Indeed, in other simulation runs with the same values $N = 10$, $c = 1$ and $p = 0.9$ we did see small numbers of cases in which the worst candidate had been selected.

j	Select j if $r_j \leq$									
	$p = 0.1$	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
1	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	1
3	-	-	-	-	-	-	-	-	1	1
4	-	-	-	-	-	-	1	1	1	2
5	-	-	-	-	-	1	1	2	2	2
6	-	-	-	-	1	1	1	2	2	3
7	-	-	-	1	1	2	2	3	3	4
8	-	-	1	2	2	3	3	4	5	6
9	-	1	2	3	4	5	6	6	7	8
10	10	10	10	10	10	10	10	10	10	10

Table 3.2: NPI for $N = 10, c = 1$: optimal procedure

Figures 3.2 and 3.3 provide results of a similar simulation of 10,000 runs, but with $N = 30$ and $N = 60$, respectively. The conclusions are in line with those for $N = 10$, but it should be noted that $p = 0.1$ now leads to a slightly better performance than 'fully random', as no longer the final candidate is selected with certainty. In our procedure applied with $c = 1$, candidate N gets selected with certainty if and only if candidate $N - 1$ would not even get selected if he has rank $r_{N-1} = 1$ and no earlier candidates were selected (so $a = 0$). In this case, probability (3.1) is equal to $1/N$, which is of course immediately clear as it is just the probability for the event that the final candidate has absolute rank $R_N = 1$. So, it follows that for $p \leq 1/N$ our method leads to automatic selection of the final candidate, so in the case with $N = 10$ discussed above the value $p = 1/10$ was precisely the largest possible value of p for which the tenth candidate is always selected.

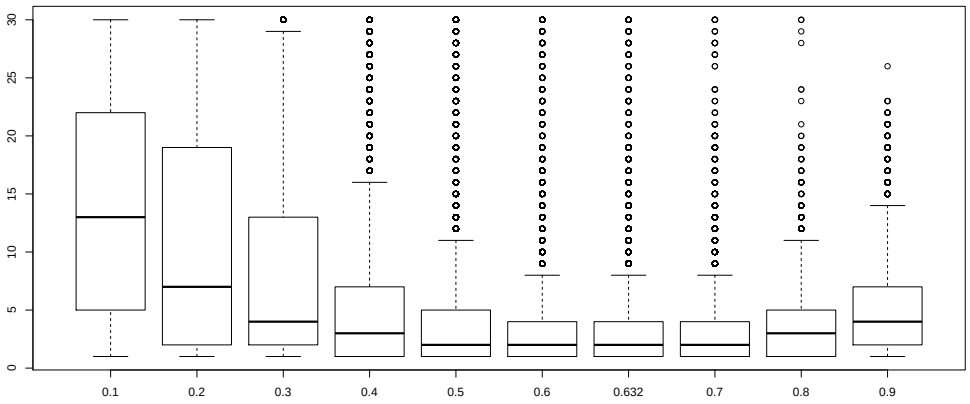

Figure 3.2: NPI for $N = 30, c = 1$

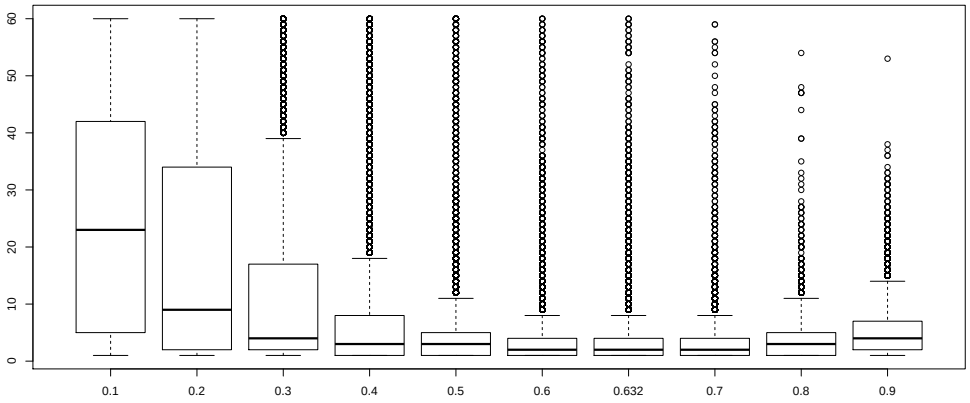

Figure 3.3: NPI for $N = 60, c = 1$

Table 3.3 presents some further summary statistics of these simulations with $N = 10$, $N = 30$ and $N = 60$, namely the means and standard deviations of the absolute ranks of the selected candidates per run, together with the number of best ($R_j = 1$) and worst ($R_j = N$) selections. Of course, to select (one of) the best candidate(s) out of $N = 60$ is a substantially harder problem than with $N = 10$ candidates, but we see that the performance of our procedure is actually quite close for these two cases when considering how often the best candidate is selected. Chow [6] showed that, when following the optimal dynamic programming procedure, the expected absolute rank of the selected candidate in the case with N candidates and $c = 1$ has limiting value 3.8695 for $N \to \infty$. This expected value increases as function of N. This suggests that our NPI approach for $N = 60$ and with $p = 0.7$ performs well as the mean absolute rank of the selected candidates for the 10,000 simulation runs is 3.61.

(N, c)	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$N = 10, c = 1$										
$Mean$	5.48	4.99	4.17	3.39	3.02	2.69	2.60	2.60	2.77	3.36
sd	2.90	3.06	3.05	2.69	2.41	1.96	1.85	1.71	1.68	2.03
Best	1042	1890	2630	3002	3426	3336	3486	3213	2802	2191
Worst	996	883	676	444	309	149	90	56	15	0
$N = 30, c = 1$										
$Mean$	13.89	10.57	8.15	6.06	4.70	3.72	3.49	3.29	3.41	4.74
sd	9.26	9.47	8.66	7.25	5.86	4.35	3.80	3.12	2.63	3.69
Best	994	1710	2369	2789	3053	3140	3071	2969	2611	1834
Worst	330	248	178	109	65	20	16	6	2	0
$N = 60, c = 1$										
$Mean$	25.07	18.12	12.80	8.18	5.78	4.44	3.97	3.61	3.69	5.27
sd	19.19	18.86	16.43	12.49	9.20	6.67	5.73	4.36	3.26	4.57
Best	933	1699	2292	2714	3004	3107	3070	2952	2507	1721
Worst	143	109	71	36	21	4	3	0	0	0

Table 3.3: NPI for different N, c: summary statistics

3.3.1 NPI for Best-Only (NPI-BO)

As mentioned before, our heuristic method is developed particularly for problems where multiple candidates have to be selected from a substantial total number of candidates. The case with a single candidate to be accepted ($c = 1$), as presented in this section, is mainly included to get a feeling for the approach and the influence of the value of p. We included the value $p = 0.632$ in our simulations for the following reason. If we change our method, still for $c = 1$, such that in addition to the criterion given by inequalities (3.2) and (3.3) we only accept candidate j if $r_j = 1$, so if he is the best candidate seen thus far, then we can compare our approach with Lindley's optimal solution for the problem with interest only in selecting the best candidate, using the 0-1 loss function as described at the start of this section [33]. We call this variation to our method 'NPI-BO' (for Best Only).

The value j^* in Lindley's method, which is such that the j^*-th candidate is the first who will be selected if $r_{j^*} = 1$, can be used together with inequalities (3.2) and (3.3), the latter together with the fact that candidate $j^* - 1$ would be rejected even with $r_{j^*-1} = 1$. This leads to the following two inequalities

$$p_l^* = \binom{N}{j^*}^{-1} \sum_{l=1}^{N-j^*} \binom{N-1-l}{N-j^*-l} < p \tag{3.4}$$

and

$$p_u^* = \binom{N}{j^*-1}^{-1} \sum_{l=1}^{N-j^*+1} \binom{N-1-l}{N-j^*+1-l} \geq p \tag{3.5}$$

These inequalities provide an interval such that for $p \in (p_l^*, p_u^*]$ the results of the NPI-BO method and Lindley's method with 0-1 loss function are the same (in both these methods the final candidate is accepted if no earlier candidate was accepted). For $N = 10$ we have $j^* = 4$ and $(p_l^*, p_u^*] = (0.6, 0.7]$ while for $N = 60$ we have $j^* = 23$ and $(p_l^*, p_u^*] = (0.617, 0.633]$ and for $N = 1000$ we have $j^* = 369$ and $(p_l^*, p_u^*] = (0.631, 0.632]$. More detailed numerical study suggested that p_l^* is increasing and p_u^* is decreasing as functions of N, and we further note that these values appear to converge to $0.632 = 1 - e^{-1}$ for $N \to \infty$. Analytic justification of these suggestions is difficult due to the required use of j^*, which is not available in closed form. However, Lindley's method is known to select the best candidate with probability $e^{-1} = 0.368$

for $N \to \infty$. So, if candidate j^* with $r_j^* = 1$ is accepted because the probability that there will be a better candidate in the future is about 0.632 for very large values of N, then indeed this candidate is the best candidate overall with probability of about 0.368. The choice of $p = 0.632$ for our method, as included in this chapter, is therefore based on the fact that our method NPI-BO and Lindley's method with the 0-1 loss function coincide for this value for, it appears, all values of N.

First, we compare our heuristic method NPI with Lindley's method (L), the results of simulations for $N = 10$ and $N = 30$ for different values of p including 0.632 are presented in Figures 3.4 and 3.5. If $N = 10$, as is shown in Figure 3.4, NPI performs worse than Lindley's method for $p \leq 0.4$ and $p = 0.9$, but better for $p = 0.6, 0.632, 0.7$ and about similarly for $p = 0.5$ and $p = 0.8$ as can also be seen in Table 3.4. The best possible selection in this case, namely the candidate with absolute rank 1, was selected by Lindley's dynamic programming method 3979 out of 10,000 times, while the best candidate gets selected by NPI only 3002 times with $p = 0.4$. This quality of selection became better when p increased to $p = 0.5$ and $p = 0.6$ with best performance for $p = 0.632$, where the optimal selection was made 3486 times, before getting worse again for $p = 0.8$, as shown in Table 3.9. Obviously, overall, NPI performs better than Lindley's method in the frequency of selecting the second or third best candidates, candidates with absolute ranks 2 or 3, for all values of p. NPI selects the worst candidate, so with absolute rank 10, less frequently, especially for $p \geq 0.5$.

Another simulation of 10,000 runs is presented in Figure 3.5, for the case $N = 30$. Figure 3.5 shows the overall changes in the quality of selection for both the NPI approach and Lindley's method. The overall figure is nearly identical to the results for $N = 10$ in Figure 3.4.

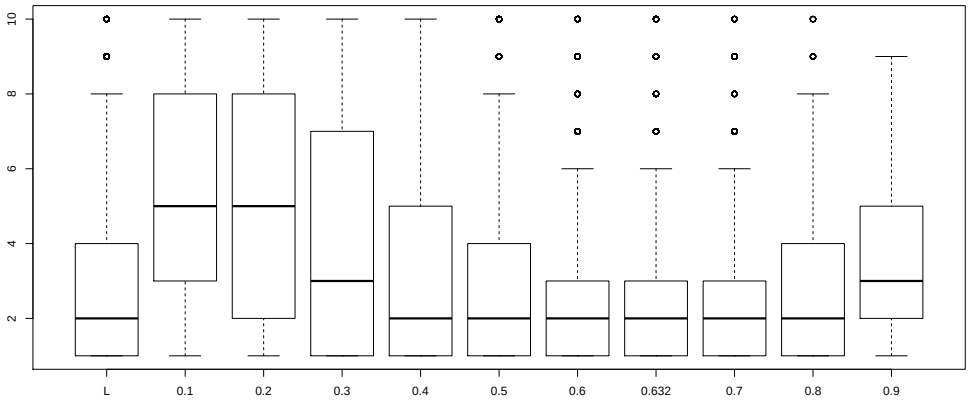

Figure 3.4: Lindley's method (L) and NPI for N=10, c=1

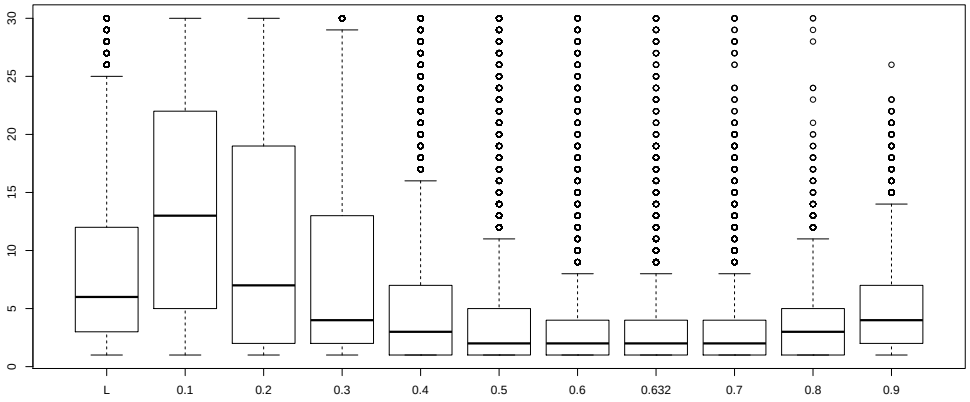

Figure 3.5: Lindley's method (L) and NPI for N=30, c=1

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	1480	633	631	275	183	134	157	169	160	162	3979
2	976	510	225	128	35	28	31	29	32	37	2031
3	337	378	285	63	10	6	9	11	2	17	1118
4	136	187	121	216	4	4	3	7	4	0	682
5	49	119	63	16	216	1	4	2	1	0	471
6	18	76	57	8	1	206	1	0	0	0	367
7	6	80	50	33	0	0	209	0	0	0	348
8	0	77	37	0	0	0	0	200	0	0	314
9	0	81	49	0	0	0	0	0	225	0	355
10	0	66	41	0	0	0	0	0	0	228	335
Total	3002	2207	1559	709	449	379	414	413	424	444	10000

Table 3.4: NPI and Lindley's method with $p = 0.4$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	2064	431	478	315	204	95	99	94	101	98	3979
2	881	753	164	105	81	9	7	10	8	13	2031
3	296	362	355	58	28	4	3	5	0	7	1118
4	123	161	136	223	32	1	1	2	3	0	682
5	39	111	73	58	190	0	0	0	0	0	471
6	16	76	53	45	5	172	0	0	0	0	367
7	7	77	54	37	1	0	172	0	0	0	348
8	0	77	37	32	0	0	0	168	0	0	314
9	0	81	49	55	0	0	0	0	170	0	355
10	0	66	41	37	0	0	0	0	0	191	335
Total	3426	2195	1440	965	541	281	282	279	282	309	10000

Table 3.5: NPI and Lindley's method with $p = 0.5$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	2515	301	475	302	175	87	36	28	27	33	3979
2	503	1131	168	129	62	38	0	0	0	0	2031
3	196	297	465	88	38	34	0	0	0	0	1118
4	79	169	154	223	29	28	0	0	0	0	682
5	27	130	99	55	132	28	0	0	0	0	471
6	12	98	81	43	34	99	0	0	0	0	367
7	4	100	70	44	32	3	95	0	0	0	348
8	0	102	57	32	23	0	0	100	0	0	314
9	0	104	79	45	24	0	0	0	103	0	355
10	0	87	67	38	27	0	0	0	0	116	335
Total	3336	2519	1715	999	576	317	131	128	130	149	10000

Table 3.6: NPI and Lindley's method with $p = 0.6$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	3486	0	243	133	62	30	25	0	0	0	3979
2	0	1748	92	95	42	25	29	0	0	0	2031
3	0	101	847	75	30	34	31	0	0	0	1118
4	0	95	67	442	25	28	25	0	0	0	682
5	0	102	71	27	216	28	27	0	0	0	471
6	0	91	75	32	23	115	31	0	0	0	367
7	0	97	67	41	28	21	94	0	0	0	348
8	0	102	57	32	23	33	0	67	0	0	314
9	0	104	79	45	24	27	0	0	76	0	355
10	0	87	67	38	27	26	0	0	0	90	335
Total	3486	2527	1665	960	500	367	262	67	76	90	10000

Table 3.7: NPI and Lindley's method with $p = 0.632$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	3213	0	357	215	129	49	16	0	0	0	3979
2	0	1658	113	122	91	27	20	0	0	0	2031
3	0	114	810	73	54	45	22	0	0	0	1118
4	0	115	69	416	35	28	19	0	0	0	682
5	0	121	66	47	190	31	16	0	0	0	471
6	0	115	74	39	28	90	21	0	0	0	367
7	0	121	63	59	35	10	60	0	0	0	348
8	0	111	56	52	29	20	0	46	0	0	314
9	0	120	74	57	34	19	0	0	51	0	355
10	0	110	68	58	28	15	0	0	0	56	335
Total	3213	2585	1750	1138	653	334	174	46	51	56	10000

Table 3.8: NPI and Lindley's method with $p = 0.7$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI										Total
	1	2	3	4	5	6	7	8	9	10	
1	1809	779	617	379	222	116	48	9	0	0	3979
2	107	974	389	274	167	70	43	7	0	0	2031
3	119	85	484	199	116	77	28	10	0	0	1118
4	59	84	62	251	105	58	23	4	0	0	682
5	116	80	51	30	108	84	24	8	0	0	471
6	117	67	59	31	18	47	25	3	0	0	367
7	113	87	54	35	25	10	15	9	0	0	348
8	94	81	60	37	19	7	6	10	0	0	314
9	121	91	50	42	18	16	3	0	14	0	355
10	111	83	56	30	28	8	4	0	0	15	335
Total	2802	2411	1882	1308	826	463	219	60	14	15	10000

Table 3.9: NPI and Lindley's method with $p = 0.8$, $N = 10$ and $c = 1$

Secondly, we performed simulations to compare NPI-BO to Lindley's method for $N = 10$ and $N = 30$, the results are presented in Figures 3.6 and 3.7. In order to compare NPI-BO to Lindley's method for the two cases $N = 10$ and $N = 30$, Figures 3.8 and 3.9 show the comparison between NPI-BO and Lindley's method (L) in both cases respectively, with $p = 0.5 - 0.8$. When $N = 10$ it should be noted that the overall quality of selections for L were almost identical with NPI-BO for $p = 0.632, 0.70$ and 0.80, while NPI-BO performs worse than L for $p \leq 0.6$. Tables 3.10-3.13 show the results for the case $N = 10$. For example, Lindley's method selected the optimal candidate 3979 times, while NPI-BO for $p = 0.5$ only did so 3766 times, with an improvement to 3958 times for $p = 0.6$ as illustrated in Tables 3.10 and 3.11. Then NPI-BO reaches the optimal solution to get exactly the same results with Lindley's method for $p = 0.632$ and $p = 0.7$ as detailed in Table 3.12 before a gradual descent in performance for $p = 0.8$ in Table 3.13, where the best candidate was selected 3652 times, together with a substantial improvement in selecting the second, third, fourth and fifth candidates and a smaller number of selecting the worst candidate with absolute rank 10.

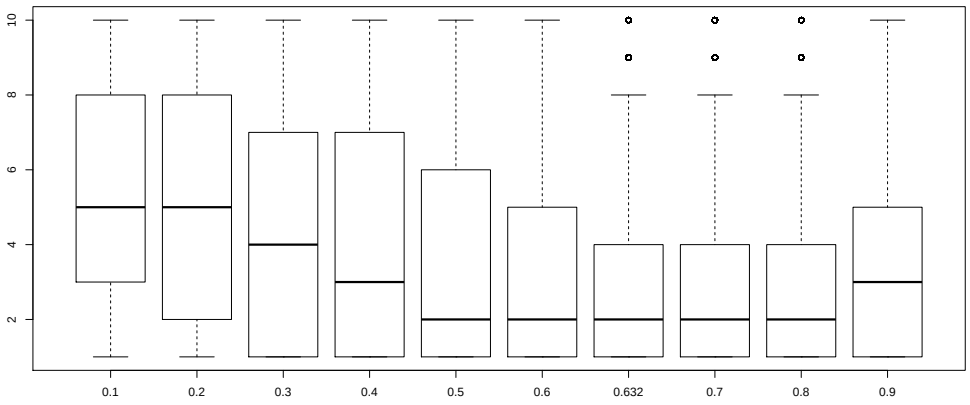

Figure 3.6: NPI-BO for N=10, c=1

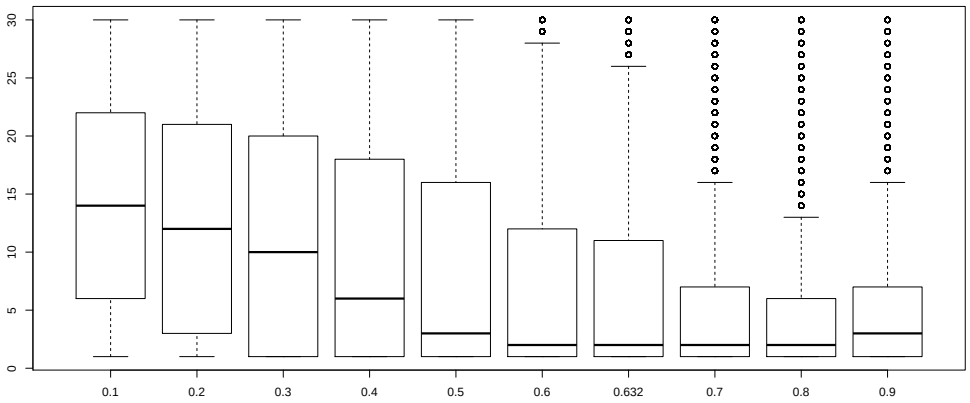

Figure 3.7: NPI-BO for N=30, c=1

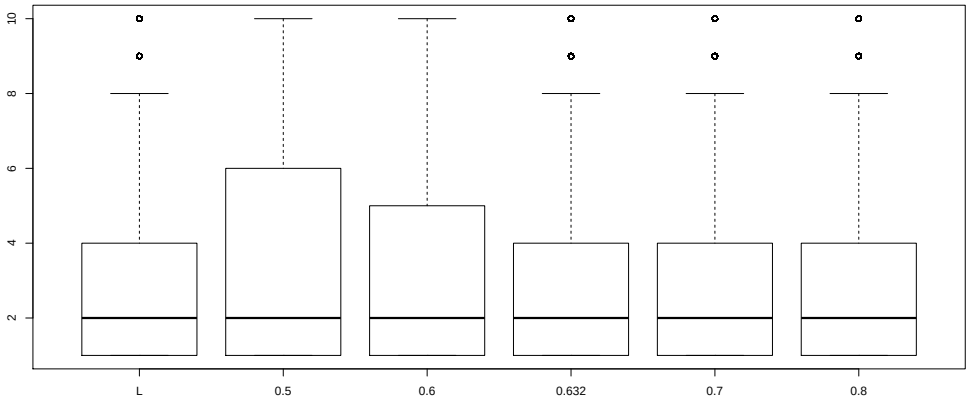

Figure 3.8: Lindley's method and NPI-BO for N=10, c=1 with $p = 0.5 - 0.8$

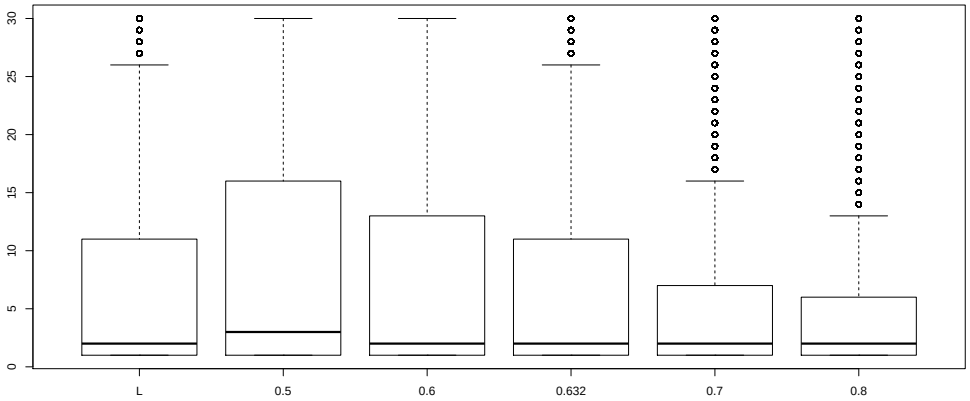

Figure 3.9: Lindley's method and NPI-BO with p=0.5-0.8 when N=30, c=1

Lindley	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	2266	204	207	193	180	166	186	180	206	191	3979
2	1003	914	14	20	16	13	8	12	13	18	2031
3	301	267	499	7	4	7	6	9	6	12	1118
4	129	80	91	359	3	4	3	4	6	3	682
5	42	38	35	22	329	0	0	0	3	2	471
6	18	11	11	7	5	314	1	0	0	0	367
7	7	4	6	2	1	0	328	0	0	0	348
8	0	0	0	0	0	0	0	314	0	0	314
9	0	0	0	0	0	0	0	0	355	0	355
10	0	0	0	0	0	0	0	0	0	335	335
Total	3766	1518	863	610	538	504	532	519	589	561	10000

Table 3.10: Lindley's method and NPI-BO with $p = 0.5$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	2983	123	112	117	109	89	110	111	116	109	3979
2	646	1385	0	0	0	0	0	0	0	0	2031
3	206	207	705	0	0	0	0	0	0	0	1118
4	80	74	90	438	0	0	0	0	0	0	682
5	27	28	28	28	360	0	0	0	0	0	471
6	12	7	6	11	11	320	0	0	0	0	367
7	4	3	3	3	4	3	328	0	0	0	348
8	0	0	0	0	0	0	0	314	0	0	314
9	0	0	0	0	0	0	0	0	355	0	355
10	0	0	0	0	0	0	0	0	0	335	335
Total	3958	1827	944	597	484	412	438	425	471	444	10000

Table 3.11: Lindley's method and NPI-BO with $p = 0.6$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3979	0	0	0	0	0	0	0	0	0	3979
2	0	2031	0	0	0	0	0	0	0	0	2031
3	0	0	1118	0	0	0	0	0	0	0	1118
4	0	0	0	682	0	0	0	0	0	0	682
5	0	0	0	0	471	0	0	0	0	0	471
6	0	0	0	0	0	367	0	0	0	0	367
7	0	0	0	3	0	0	348	0	0	0	348
8	0	0	0	0	0	0	0	314	0	0	314
9	0	0	0	0	0	0	0	0	355	0	355
10	0	0	0	0	0	0	0	0	0	335	335
Total	3979	2031	1118	682	471	367	348	314	355	335	10000

Table 3.12: Lindley's method and NPI-BO with $p = 0.632$ or $p = 0.7$, $N = 10$ and $c = 1$

Lindley	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	2659	779	267	142	77	35	16	4	0	0	3979
2	107	1358	298	151	67	28	19	3	0	0	2031
3	119	0	740	139	69	29	15	7	0	0	1118
4	95	0	0	467	81	26	11	2	0	0	682
5	116	0	0	0	313	30	12	0	0	0	471
6	117	0	0	0	0	235	14	1	0	0	367
7	113	0	0	0	0	0	227	8	0	0	348
8	94	0	0	0	0	0	0	220	0	0	314
9	121	0	0	0	0	0	0	0	234	0	355
10	111	0	0	0	0	0	0	0	0	224	335
Total	3652	2137	1305	899	607	383	314	245	234	224	10000

Table 3.13: Lindley's method and NPI-BO with $p = 0.8$, $N = 10$ and $c = 1$

From the above discussion, NPI and NPI-BO provide good solutions to the sequential selection problem in case we are interested in selecting only one candidate, $c = 1$, who we wish to be the best, so with absolute rank 1. Thus we finally compare NPI to NPI-BO in a simulation with 10,000 runs for $N = 10$. For $p = 0.4$, Table 3.14 shows that NPI-BO gives the optimal candidate more often than NPI with 3330 times for NPI-BO and 3002 times for NPI, while NPI was better in selecting candidates with absolute ranks 2 or 3 and NPI selected the worst candidate less often, 444 times for NPI and 662 times for NPI-BO. For $p = 0.5$, Table 3.15 illustrates an improvement in both methods where the best candidate gets selected 3766 times by NPI and 3426 times by NPI-BO. In Table 3.16, for $p = 0.6$, NPI recorded a slight fall in selecting the best candidate to 3336 with an improvement in NPI-BO to 3958 times. For $p = 0.632$ and $p = 0.7$, NPI-BO reached the peak by giving the optimal choice 3979 times, while NPI gave 3486 (for $p = 0.632$) and 3213 (for $p = 0.7$) times out of 10,000 as in Tables 3.17 and 3.18, which show also a continuous improvement in NPI in terms of selecting the candidate with absolute rank 10. If $p = 0.8$, NPI scored a lower number of selecting the worst candidate, only 15 times with a dramatic fall in selecting the best, only 2802 as well as a slight decrease in the performance of NPI-BO to score 3652 times to select the best and a slight improvement in selecting the worst, 224 times.

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3002	0	0	0	0	0	0	0	0	0	3002
2	0	1028	148	157	136	135	145	141	181	136	2207
3	237	166	591	72	71	74	95	83	88	82	1559
4	91	104	83	431	0	0	0	0	0	0	709
5	0	0	0	0	449	0	0	0	0	0	449
6	0	0	0	0	0	379	0	0	0	0	379
7	0	0	0	0	0	0	414	0	0	0	414
8	0	0	0	0	0	0	0	413	0	0	413
9	0	0	0	0	0	0	0	0	424	0	424
10	0	0	0	0	0	0	0	0	0	444	444
Total	3330	1298	822	660	656	588	654	637	693	662	10000

Table 3.14: NPI and NPI-BO with $p = 0.4$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3426	0	0	0	0	0	0	0	0	0	3426
2	0	1229	123	126	109	110	115	122	144	117	2195
3	203	143	617	59	61	63	79	66	77	72	1440
4	76	86	72	369	55	50	56	52	86	63	965
5	61	60	51	56	313	0	0	0	0	0	541
6	0	0	0	0	0	281	0	0	0	0	281
7	0	0	0	0	0	0	282	0	0	0	282
8	0	0	0	0	0	0	0	279	0	0	279
9	0	0	0	0	0	0	0	0	282	0	282
10	0	0	0	0	0	0	0	0	0	309	309
Total	3766	1518	863	610	538	504	532	519	589	561	10000

Table 3.15: NPI and NPI-BO with $p = 0.5$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3336	0	0	0	0	0	0	0	0	0	3336
2	0	1467	132	135	128	119	126	146	145	121	2519
3	319	137	620	91	97	93	90	78	101	89	1715
4	173	132	99	302	42	40	56	44	61	50	999
5	87	60	50	35	184	27	35	29	34	35	576
6	43	31	43	34	33	133	0	0	0	0	317
7	0	0	0	0	0	0	131	0	0	0	131
8	0	0	0	0	0	0	0	128	0	0	128
9	0	0	0	0	0	0	0	0	130	0	130
10	0	0	0	0	0	0	0	0	0	149	149
Total	3958	1827	944	597	484	412	438	425	471	444	10000

Table 3.16: NPI and NPI-BO with $p = 0.6$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3486	0	0	0	0	0	0	0	0	0	3486
2	0	1748	101	95	102	91	97	102	104	87	2527
3	243	92	847	67	71	75	67	57	79	67	1665
4	133	95	75	442	27	32	41	32	45	38	960
5	62	42	30	25	216	23	28	23	24	27	500
6	30	25	34	28	28	115	21	33	27	26	367
7	25	29	31	25	27	31	94	0	0	0	262
8	0	0	0	0	0	0	0	67	0	0	67
9	0	0	0	0	0	0	0	0	76	0	76
10	0	0	0	0	0	0	0	0	0	90	90
Total	3979	2031	1118	682	471	367	348	314	355	335	10000

Table 3.17: NPI and NPI-BO with $p = 0.632$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	3213	0	0	0	0	0	0	0	0	0	3213
2	0	1658	114	115	121	115	121	111	120	110	2585
3	357	113	810	69	66	74	63	56	74	68	1750
4	215	122	73	416	47	39	59	52	57	58	1138
5	129	91	54	35	190	28	35	29	34	28	653
6	49	27	45	28	31	90	10	20	19	15	334
7	16	20	22	19	16	21	60	0	0	0	174
8	0	0	0	0	0	0	0	46	0	0	46
9	0	0	0	0	0	0	0	0	51	0	51
10	0	0	0	0	0	0	0	0	0	56	56
Total	3979	2031	1118	682	471	367	348	314	355	335	10000

Table 3.18: NPI and NPI-BO with $p = 0.7$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference NPI-BO										Total
	1	2	3	4	5	6	7	8	9	10	
1	2802	0	0	0	0	0	0	0	0	0	2802
2	0	1753	85	84	80	67	87	81	91	83	2411
3	350	91	1049	62	51	59	54	60	50	56	1882
4	237	123	60	683	30	31	35	37	42	30	1308
5	145	100	47	24	402	18	25	19	18	28	826
6	81	42	48	32	24	195	10	7	16	8	463
7	32	24	13	12	12	11	102	6	3	4	219
8	5	4	3	2	8	2	1	35	0	0	60
9	0	0	0	0	0	0	0	0	14	0	14
10	0	0	0	0	0	0	0	0	0	15	15
Total	3652	2137	1305	899	607	383	314	235	234	224	10000

Table 3.19: NPI and NPI-BO with $p = 0.8$, $N = 10$ and $c = 1$

3.3.2 NPI with Non-Constant p (NPI-NC)

In this chapter we use a constant p for all candidates j when applying our method. One could argue in favour of using a function $p(j)$ instead, where an increasing function seems intuitively most attractive as this is relatively kinder to later candidates than to early ones, so it would avoid accepting too many candidates early on while at the same time not delaying acceptance of later candidates too much. The use of such non-constant $p(j)$ in our heuristic approach is studied via simulations in this section. We consider the case when p varies linearly from 0.4 to 0.6, as described in Equation (3.6), and also from 0.5 to 0.8 as in Equation (3.7)

$$p(j) = 0.4 + \frac{0.2}{N-1}(j-1) \tag{3.6}$$

$$p(j) = 0.5 + \frac{0.3}{N-1}(j-1) \tag{3.7}$$

The same 10,000 independent runs for $N = 10$, which were used in the previous subsection, are used again in order to compare NPI-NC to NPI approach. This comparetive study is based firstly on Equation (3.6) and presented in Tables 3.20-3.24 for $p = 0.5 - 0.8$ respectively. NPI performs better than NPI-NC in terms of selecting the best candidate when $p = 0.5, 0.6$ and 0.632 where NPI-NC recorded 3241 times, while NPI gives the optimal candidate 3426 times for $p = 0.5$ as shown in Table 3.20, 3336 times for $p = 0.6$ as in Table 3.21, 3486 times for $p = 0.632$ as in Table 3.22, with a gradual decrease in selecting the best if $p = 0.7$ and $p = 0.8$ comparing to NPI-NC, which is seen in Tables 3.23 and 3.24. Also, in most cases NPI provided the second best candidate more than NPI-NC and NPI does not select the worst candidate as often as NPI-NC does.

Now we consider the same simulation in case $0.5 \leq p \leq 0.8$ as in Equation (3.7), the frequencies of the selected absolute ranks from the best candidate (absolute rank 1) to the worst candidate (absolute rank 10) are 3260, 2456, 1794, 1109, 635, 341, 183, 152, 37, and 33, respectively. These results show a slight improvement in the performance of NPI-NC than the previous considered criteria which presented in equation 3.6. However, even with the slight improvement in NPI-NC by considering

p between 0.5 and 0.8, NPI method is still better than NPI-NC.

NPI	Nonparametric predictive inference for nonconstant p NPI-NC										Total
	1	2	3	4	5	6	7	8	9	10	
1	3241	0	0	97	27	61	0	0	0	0	3426
2	0	2009	0	108	32	46	0	0	0	0	2195
3	0	0	1215	131	34	60	0	0	0	0	1440
4	0	0	106	771	35	53	0	0	0	0	965
5	0	0	66	27	398	50	0	0	0	0	541
6	0	0	54	0	39	188	0	0	0	0	281
7	0	0	39	0	53	0	190	0	0	0	282
8	0	0	35	0	47	0	0	197	0	0	279
9	0	0	37	0	57	0	0	0	188	0	282
10	0	0	42	0	47	0	0	0	0	220	309
Total	3241	2009	1594	1134	769	458	190	197	188	220	10000

Table 3.20: NPI and NPI-NC with $p = 0.5$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference for nonconstant p NPI-NC										Total
	1	2	3	4	5	6	7	8	9	10	
1	2373	201	246	153	130	88	31	36	33	45	3336
2	497	1463	155	138	102	52	28	33	25	26	2519
3	266	263	1097	59	20	10	0	0	0	0	1715
4	84	61	78	770	6	0	0	0	0	0	999
5	19	20	17	12	508	0	0	0	0	0	576
6	2	1	1	2	3	308	0	0	0	0	317
7	0	0	0	0	0	0	131	0	0	0	131
8	0	0	0	0	0	0	0	128	0	0	128
9	0	0	0	0	0	0	0	0	130	0	130
10	0	0	0	0	0	0	0	0	0	149	149
Total	3241	2009	1594	1134	769	458	190	197	188	220	10000

Table 3.21: NPI and NPI-NC with $p = 0.6$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference for nonconstant p NPI-NC										Total
	1	2	3	4	5	6	7	8	9	10	
1	1773	367	421	300	242	126	70	59	58	70	3486
2	813	1074	187	159	109	74	23	33	26	29	2527
3	409	377	766	59	29	14	2	4	0	5	1665
4	157	110	134	546	6	5	0	1	0	0	960
5	42	42	37	33	345	1	0	0	0	0	500
6	15	7	12	11	8	207	21	33	27	26	367
7	32	32	37	26	30	31	74	0	0	0	262
8	0	0	0	0	0	0	0	67	0	0	67
9	0	0	0	0	0	0	0	0	76	0	76
10	0	0	0	0	0	0	0	0	0	90	90
Total	3241	2009	1594	1134	769	458	190	197	188	220	10000

Table 3.22: NPI and NPI-NC with $p = 0.632$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference for nonconstant p NPI-NC										Total
	1	2	3	4	5	6	7	8	9	10	
1	1500	367	421	300	242	126	70	59	58	70	3213
2	813	863	262	216	143	104	45	50	37	52	2585
3	518	439	580	86	67	32	7	7	6	8	1750
4	237	172	179	410	48	23	18	15	17	19	1138
5	116	110	93	82	227	25	0	0	0	0	653
6	34	35	31	20	23	127	10	20	19	15	334
7	23	23	28	20	19	21	40	0	0	0	174
8	0	0	0	0	0	0	0	46	0	0	46
9	0	0	0	0	0	0	0	0	51	0	51
10	0	0	0	0	0	0	0	0	0	56	56
Total	3241	2009	1594	1134	769	458	190	197	188	220	10000

Table 3.23: NPI and NPI-NC with $p = 0.7$, $N = 10$ and $c = 1$

NPI	Nonparametric predictive inference for nonconstant p NPI-NC										Total
	1	2	3	4	5	6	7	8	9	10	
1	666	485	488	378	283	173	84	73	77	95	2802
2	1022	348	324	238	183	102	48	59	39	48	2411
3	711	506	233	149	112	63	34	26	23	25	1882
4	424	336	270	139	56	32	13	9	16	13	1308
5	231	188	151	134	60	24	6	12	7	13	826
6	119	95	80	64	47	39	1	2	9	7	463
7	51	41	42	29	18	22	3	6	3	4	219
8	17	10	6	3	10	3	1	10	0	0	60
9	0	0	0	0	0	0	0	0	14	0	14
10	0	0	0	0	0	0	0	0	0	15	15
Total	3241	2009	1594	1134	769	458	190	197	188	220	10000

Table 3.24: NPI and NPI-NC with $p = 0.8$, $N = 10$ and $c = 1$

3.4 Selecting multiple candidates

We now consider the problem of selecting c candidates out of the total N candidates, where all candidates are observed sequentially, and the decision to accept or reject each candidate must be made instantaneously based only on the relative rank of this candidate compared to the previous candidates. So we still restrict attention to candidates arriving one-at-a-time, the generalization to candidates arriving in groups is discussed in Section 3.5. Of course, we hope to select the c candidates with absolute ranks $1, \ldots, c$, the worst possible outcome would be to end up with the c candidates with absolute ranks $N-c+1, \ldots, N$. In order to measure the quality of the group of candidates, we will throughout this chapter use the sum of the absolute ranks of the selected candidates, aiming at minimisation of this sum. Hence this sum can be considered as a loss function; if one prefers to formulate decision problems with utility functions, then choose minus this sum as utility function and aim at maximisation of utility. The heuristic approach presented here can also be used for other loss functions that take all the ranks of the selected people into account and that aims generally speaking at minimisation of these ranks, e.g. possibly using a weighted sum. For such different criteria, the main reason why a heuristic approach is needed is the same, we discuss this next before we analyse the performance of the heuristic approach using simulations.

This sequential acceptance problem can be formulated as a dynamic programming problem with N stages, related to the N candidates, with at each stage the decision to either accept or reject the candidate, with observations of the relative rank r_j just before the decision at stage j, so about candidate j, is made. The relative rank is represented as a random quantity[2] $r_j \in \{1, \ldots, j\}$, it should be noted that only after observing the r_j for all N candidates we know the absolute ranks R_j, for all candidates $j = 1, \ldots, N$. To represent this optimisation problem as a dynamic programming problem, the whole problem without taking some con-

[2]We use the same notation r_j for the relative rank of candidate j both for its random value before observation and for its observed value, and similar for his absolute rank R_j. This avoids introducing further notation and interpretation will be clear from the context.

straints into account would require a graph with $2^{N-1} \times N!$ paths, where the first factor relates to all decisions on acceptance or rejectance and the second factor to the number of possible sequences of relative ranks. We take into account here that for the final candidate the acceptance decision will be clear due to the strict requirement to select c candidates. This requirement actually reduces the graph further, because once c candidates have been accepted one cannot accept further candidates, yet the ranks of the remaining candidates are still needed in order to determine the absolute ranks of the selected candidates. Even with optimal reduction of the graph, the number of paths to go through to solve the dynamic programming problem is too large for practical implementation in all cases except when both N and c are small. This explains the need for heuristic methods, like the one presented in this chapter.

Although our study does not involve an actual application, there are many decision problems of this nature, where large values of N are not uncommon. For example, a university may have the opportunity to offer studentships to 10 students, with 100 students who meet the requirements visiting over several months to meet with selectors, with the latter eager to make offers quickly possible since students might otherwise accept offers elsewhere. Of course, in practice such problems may contain more aspects which complicate the modelling and optimisation process, for example information may be available beyond ranks, or on multiple attributes, and due to historical information one may have some further judgements about the absolute rank of a candidate prior to observing all candidates. Nevertheless, often such further aspects end up in a combined score representing an overall ranking, in which case heuristic methods are likely to remain necessary and the method presented in this chapter remains relevant. It would be interesting to extend the currently proposed heuristic method to deal with possible further information, we leave this as a topic for future research.

We performed extensive simulation studies for our heuristic method in cases where c candidates must be selected sequentially from a group of N candidates. Throughout, each case (N, c) was simulated 10,000 times for each of the values of p considered. We consider first the case with $N = 5$ candidates and two candidates

must be selected, so $c = 2$. For each of the 10,000 runs in the simulations, our procedure was applied with p taking the values $0.1, 0.2, \ldots, 0.9$ and, $p = 0.632$ was also included as in Section 3.3. Table 3.25 shows the number of times the selected candidate had each absolute rank. It can be clearly seen that the procedure functions poorly for small values of p, with best performance for p about 0.4 and 0.5, where the actual best performance depends of course on the chosen quality criterion.

p	Absolute rank							Mean	sd
	3	4	5	6	7	8	9		
0.1	1655	1342	2130	2015	1507	683	668	5.51	1.7
0.2	2156	1303	2012	1837	1341	683	668	5.36	1.81
0.3	2381	1867	2477	1791	978	338	168	4.88	1.51
0.4	3148	1831	2471	1382	838	162	168	4.61	1.47
0.5	3148	1831	2471	1382	838	162	168	4.61	1.47
0.6	2316	1679	2408	1461	1166	480	490	5.10	1.71
0.7	2138	1640	2435	1464	1353	480	490	5.17	1.71
0.8	1966	1567	2443	1603	1451	480	490	5.24	1.69
0.9	981	1033	2035	1960	1975	1013	1003	6.00	1.73

Table 3.25: NPI for $N = 5, c = 2$: absolute rank of selected candidate

Table 3.26 gives summaries of simulation results for the cases with (N, c) equal to $(5, 2)$, $(10, 3)$, $(10, 7)$, $(20, 6)$, $(30, 6)$ and $(30, 10)$. For each simulated scenario, the statistic of interest is the sum of the absolute ranks of the c selected candidates. Table 3.26 provides the mean and standard deviation (sd) of the 10,000 simulations for each case, together with the number of times the selected group was actually the best possible, so the candidates with absolute ranks $1, \ldots, c$ were selected. Figures 3.10-3.15 present these simulation results in box plots. It is interesting that these box plots all have quite similar shapes for corresponding values of p, with best performances typically for values of p in the range 0.5 to 0.7. The optimal values for the means and 'best' in Table 3.26 are also almost all achieved by values of p in this range, except for the maximal value of 'best' for the case $(10, 3)$ which occurred for $p = 0.4$. The computation time for these simulations was effectively neglectable due to the straightforward nature of our approach with only a single probability to

be calculated for each candidate. This also holds for substantially larger values of N and c, for example the comparisons with alternative heuristic methods in Section 3.6 are performed for $N = 200$ and $c = 20, 40, 60, 80$ without any computational problems.

(N, c)	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$N = 5, c = 2$										
Mean	5.98	5.50	5.35	4.85	4.57	4.57	5.13	5.12	5.18	5.25
sd	1.73	1.74	1.80	1.49	1.45	1.45	1.72	1.72	1.71	1.69
Best	1022	1683	2175	2423	3201	3201	2281	2281	2152	1995
$N = 10, c = 3$										
Mean	13.68	12.30	11.11	10.38	10.00	9.80	9.87	10.16	11.96	12.69
sd	4.52	4.22	3.78	3.62	3.24	3.05	3.08	3.13	4.20	4.26
Best	504	750	1145	1607	1558	1530	1472	1257	842	585
$N = 10, c = 7$										
Mean	34.60	33.85	33.65	31.84	31.92	32.30	32.57	32.96	34.13	35.69
sd	4.27	4.20	3.34	3.05	3.10	3.59	3.63	3.79	4.37	4.56
Best	665	909	1003	1512	1466	1455	1353	1263	1033	577
$N = 20, c = 6$										
Mean	43.14	38.15	34.61	32.39	31.54	31.19	31.19	31.81	37.14	40.24
sd	12.20	10.73	9.45	8.18	7.49	7.00	6.97	7.29	10.35	11.73
Best	130	237	350	482	481	483	498	425	238	144
$N = 30, c = 6$										
Mean	53.77	45.54	39.58	36.20	34.55	33.78	33.68	35.47	37.30	48.70
sd	18.61	16.34	13.90	11.73	10.28	9.31	9.13	10.48	11.48	17.28
Best	77	197	272	348	388	403	332	234	80	393
$N = 30, c = 10$										
Mean	95.58	84.67	78.66	75.74	73.73	73.60	73.83	79.27	83.33	93.68
sd	21.96	18.73	15.84	13.86	12.55	12.13	12.19	15.54	18.03	21.70
Best	45	95	153	172	195	186	186	136	92	39

Table 3.26: NPI for different N, c: summary statistics

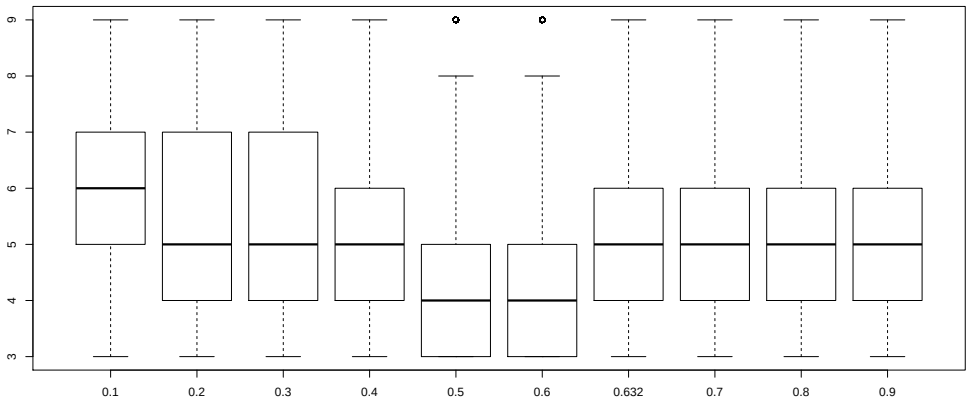

Figure 3.10: NPI with N=5, c=2

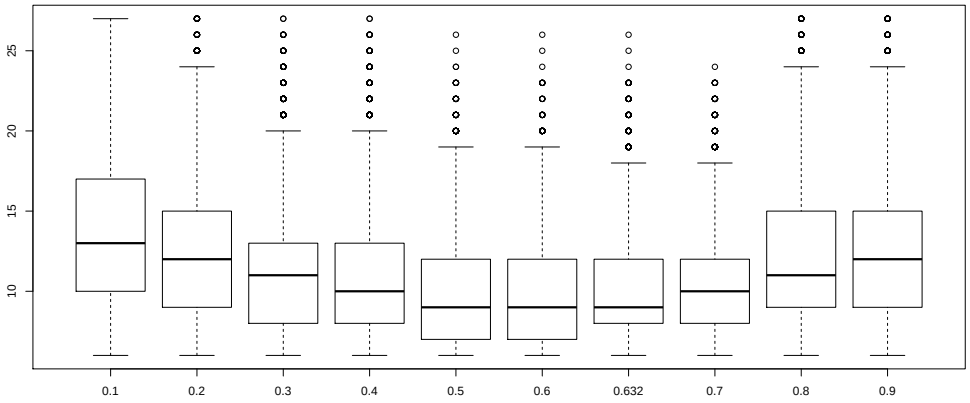

Figure 3.11: NPI with N=10, c=3

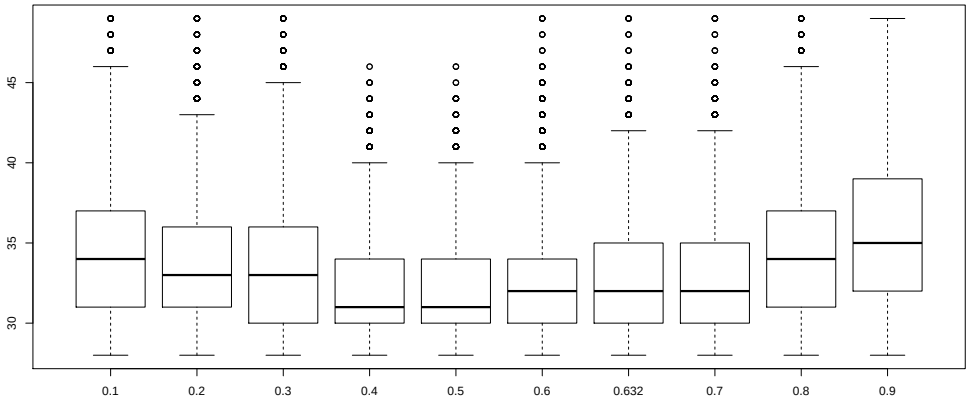

Figure 3.12: NPI with N=10, c=7

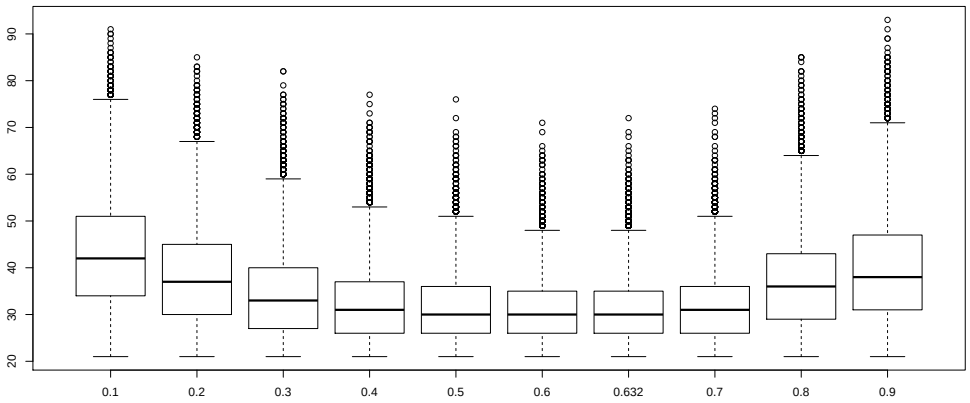

Figure 3.13: NPI with N=20, c=6

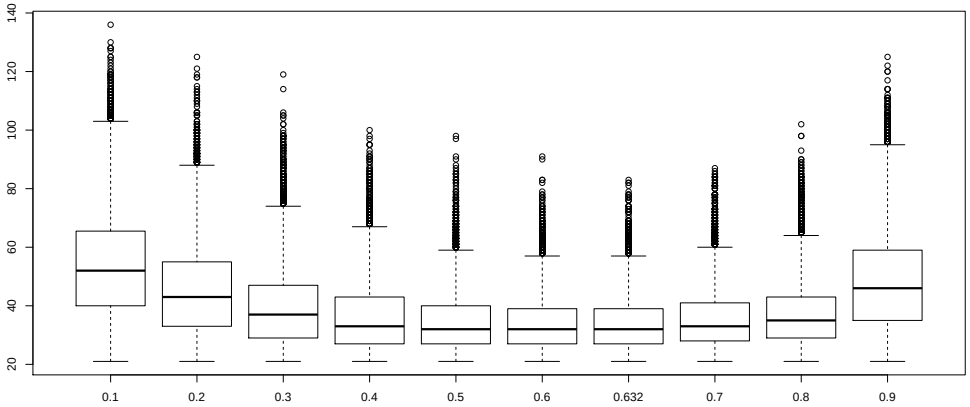

Figure 3.14: NPI with N=30, c=6

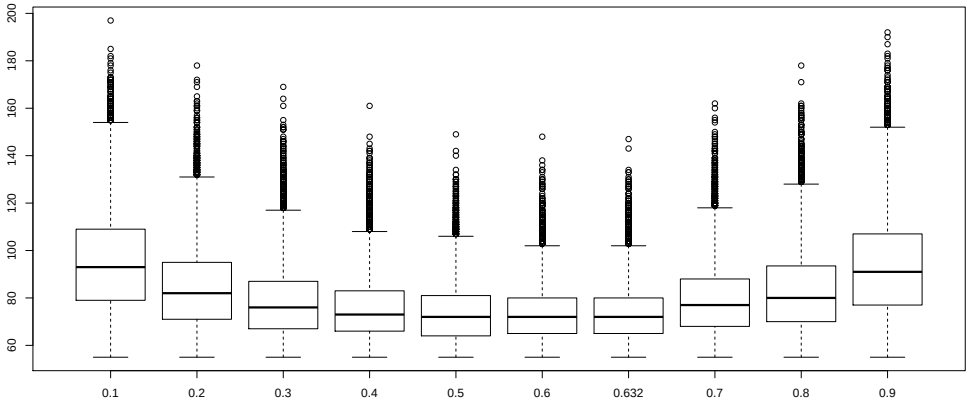

Figure 3.15: NPI with N=30, c=10

3.5 Selecting multiple candidates observed in groups

In Sections 3.3 and 3.4, candidates were observed one at a time, and the decision to either accept candidate j or not to accept him was made before observing candidate $j+1$. In this section we explore the generalization of this situation with candidates being observed in groups, and on the basis of this information acceptance decisions on the individual candidates are made. For example, in the problem to offer 10 studentships with 100 candidates, as briefly discussed in Section 3.4, it may be the case that 5 candidates visit the university on a single day, at the end of which any offers for studentships have to be made. It is straightforward to adapt our heuristic method to situations like this, where it does not matter whether the groups are all of the same size or of varying sizes. In our heuristic procedure, we see the relative ranks, among all previous candidates and those in this group, of all members of the group who are considered simultaneously. We then decide first about the best candidate in this group, using the same criterion as before in this chapter with the important change that the number of future candidates is of course based on the remaining candidates after the current group. If this strongest candidate is accepted, we consider the second strongest, and so on. Of course, if for example the second strongest candidate in the current group is not accepted, we do not need to look at further candidates from this group. To make a decision either to accept or reject the best person in the current group the predictive precision probability (2.5) is used with respect of the grouping procedure, as follows

$$P(Y^N_{(md)+1}(r_j^{md}) \geq c-a) = \binom{N}{md}^{-1} \sum_{l=c-a}^{N-(md)} \left[\binom{r_j^{md}-1+l}{l} \binom{(md)-r_j^{md}+N-(md)-l}{N-(md)-l} \right] \tag{3.8}$$

where N is the total number of candidates, j is the current candidate, m is the size of the checked group, d is the current day/stage, c is the number of candidates wanted to be selected, a is the number of selected candidates so far and r_j^{md} is the relative rank of the candidate considered among the first md candidates.

It should be noted in this probability (3.8) that the term a will change to $a+1$ within the group if a candidate has been accepted, then to $a+2$ if another one is

accepted, and so on.

Intuitively, one may expect that it is best to the selectors to observe as many candidates together as possible, of course with the optimal situation of observing all candidates together in a single group in which case $r_j = R_j$ for all candidates (j does not really indicate a 'stage' anymore here, but can still be used to label the candidates) so it is trivial to select those with absolute ranks $1, \ldots, c$. We consider one case in detail to illustrate a specific feature of such selection of candidates when they are observed in groups.

Example 3.3

Suppose that $N = 24$ candidates appear in the following order of absolute ranks: 12, 2, 7, 3, 17, 14, 9, 10, 19, 4, 8, 16, 1, 18, 20, 15, 5, 6, 24, 21, 11, 13, 23, 22. Tables 3.27-3.32 present the result of the selection process with candidates arriving in groups of size m, for different values of m and with our NPI method applied with $p = 0.3, 0.4, ..., 0.8$, respectively. For example, Table 3.29 presents the absolute ranks of the selected candidates, presented in the order in which they were selected, and also the sum of these absolute ranks with $p = 0.5$. While the case with $p = 0.6$, presented in Table 3.30, gives a better performance for $m = 1, 2, 3, 6$ and leads to identical solutions for the other values of m considered for this sequence of the candidates, we focus on the case with $p = 0.5$ as it has the interesting feature of a worse selection for $m = 6$ than for $m = 4$, hence illustrating that it is not always beneficial to see candidates in larger groups. Of course, this is due to the information for individual candidates not always being identical when using different values for m. In this specific case, having groups of size $m = 4$ leads to sum of absolute ranks of 29 while $m = 6$ leads to 36, so the larger group size gives a worse result. Clearly, this is due to the candidate with absolute rank 6, who with $m = 4$ is in the fifth group and is included based on the consideration that there are only 4 further candidates to follow after this group and that there are still 2 candidates needed to be accepted (prior to deciding on this candidate). For $m = 6$, this candidate is in the third group, so when he is considered there are still 6 candidates to follow after his group with two candidates to be selected, it then is better not to select this

candidate according to our criterion with $p = 0.5$. The last two cases in Table 3.29, with $m = 8$ and $m = 12$, show that the order in which candidates are accepted can vary for the same two candidates. The candidate with absolute rank 4 is in group 2 for $m = 8$, together with the candidate with absolute rank 1. Therefore, the latter is accepted first because we look at a group of candidates and decide on the strongest of them first. For $m = 12$, the candidate with absolute rank 1 is in the second group and is thus considered later in the process than the candidate with absolute rank 4.

$\sum$	Who						m
66	4	1	5	11	23	22	1
66	4	1	5	11	23	22	2
56	4	1	5	11	13	22	3
36	2	4	1	5	11	13	4
36	2	4	1	5	11	13	6
29	2	1	4	5	6	11	8
28	2	3	1	5	6	11	12

Table 3.27: $p = 0.3$

$\sum$	Who						m
51	4	1	5	6	13	22	1
49	4	1	5	6	11	22	2
51	4	1	5	6	13	22	3
29	2	4	1	5	6	11	4
36	2	4	1	5	11	13	6
21	2	3	1	4	5	6	8
21	2	3	4	1	5	6	12

Table 3.28: $p = 0.4$

$\sum$	Who						m
51	4	1	5	6	13	22	1
49	4	1	5	6	11	22	2
47	2	4	1	5	13	22	3
29	2	4	1	5	6	11	4
36	2	4	1	5	11	13	6
21	2	3	1	4	5	6	8
21	2	3	4	1	5	6	12

Table 3.29: $p = 0.5$

$\sum$	Who						m
40	2	4	1	5	6	22	1
40	2	4	1	5	6	22	2
28	2	3	4	1	5	13	3
29	2	4	1	5	6	11	4
26	2	3	4	1	5	11	6
21	2	3	1	4	5	6	8
21	2	3	4	1	5	6	12

Table 3.30: $p = 0.6$

$\sum$	Who						m
40	2	4	1	5	6	22	1
42	2	4	8	1	5	22	2
28	2	3	4	1	5	13	3
31	2	4	8	1	5	11	4
26	2	3	4	1	5	11	6
22	2	3	7	1	4	5	8
22	2	3	4	7	1	5	12

Table 3.31: $p = 0.7$

$\sum$	Who						m
46	12	2	4	1	5	22	1
26	2	3	4	1	5	11	2
24	2	3	9	4	1	5	3
21	2	3	4	1	5	6	4
26	2	3	4	1	5	11	6
22	2	3	7	1	4	5	8
22	2	3	4	7	1	5	12

Table 3.32: $p = 0.8$

Tables 3.33-3.38 present the optimal selections for the case that the candidates arrive in the reversed order to the one given above, and with $p = 0.3, 0.4, ..., 0.8$, respectively. Let us consider the case with $p = 0.4$ in Table 3.34, which is included as it shows that the case $m = 4$ might lead to worse selection than the case $m = 2$, which may be surprising because here the groups for $m = 2$ are not split up when using $m = 4$ instead (as was the case with the above example with $m = 4$ and $m = 6$). Clearly, the difference is in the selection of the candidate with absolute rank 9, who is included for $m = 4$ as when he is considered there are only 4 more candidates to come, but for $m = 2$ there are still 6 more candidates to come and he is not selected, which happens to be beneficial as it allows the candidate with absolute rank 3 to be selected instead, hence the result for $m = 2$ is better in this case than for $m = 4$.

$\sum$	Who						m
28	11	6	5	1	3	2	1
26	11	5	1	4	3	2	2
26	11	5	1	4	3	2	3
26	11	5	1	4	2	3	4
26	11	1	5	4	2	3	6
22	5	1	4	2	3	7	8
22	1	5	2	3	4	7	12

Table 3.33: $p = 0.3$

$\sum$	Who						m
30	11	6	5	1	4	3	1
26	11	5	1	4	3	2	2
30	11	5	6	1	4	3	3
32	11	5	1	4	9	2	4
26	11	1	5	4	2	3	6
21	5	6	1	4	2	3	8
21	1	5	6	2	3	4	12

Table 3.34: $p = 0.4$

$\sum$	Who						m
39	13	11	6	5	1	3	1
30	11	5	6	1	4	3	2
37	13	11	5	1	4	3	3
29	11	5	6	1	4	2	4
31	11	1	5	4	8	2	6
21	5	6	1	4	2	3	8
21	1	5	6	2	3	4	12

Table 3.35: $p = 0.5$

$\sum$	Who						m
39	13	11	6	5	1	3	1
49	22	11	5	6	1	4	2
39	13	11	5	6	1	3	3
29	11	5	6	1	4	2	4
36	11	13	1	5	4	2	6
21	5	6	1	4	2	3	8
21	1	5	6	2	3	4	12

Table 3.36: $p = 0.6$

$\sum$	Who						m
40	13	11	6	5	1	4	1
49	22	11	5	6	1	4	2
40	13	11	5	6	1	4	3
29	11	5	6	1	4	2	4
36	11	13	1	5	4	2	6
29	5	6	11	1	4	2	8
28	1	5	6	11	2	3	12

Table 3.37: $p = 0.7$

$\sum$	Who						m
58	22	13	11	6	5	1	1
49	22	11	5	6	1	4	2
40	13	11	5	6	1	4	3
40	11	13	5	6	1	4	4
40	11	13	1	5	6	4	6
29	5	6	11	1	4	2	8
28	1	5	6	11	2	3	12

Table 3.38: $p = 0.8$

Figures 3.16-3.22 show the results of a simulation study for 10,000 runs for selecting 6 candidates out-of 24 candidates, $N = 24, c = 6$. Figure 3.16 considers the case of observing the candidates sequentially one at a time $m = 1$, which is exactly the same as the discussed situation in Section 3.4. However, Figure 3.16 shows that NPI gives good performance especially with p between 0.5 and 0.7 with the lowest mean 32.7, 32.7, 32.8 and 33.4 respectively, as presented in the summary statistics Table 3.39 where the best quality of selection (selecting the best 6 candidates, $\sum = 21$) has been selected 410 times for $p = 0.5$, 435 times for $p = 0.6$, 414 times for $p = 0.632$ and only 324 times for $p = 0.7$.

In Figure 3.17 the size of groups is $m = 2$ which means that each time two candidates are observed together in 12 stages. It gives a slightly better performance than observing the candidates one-at-a-time, with a quite lower mean and standard

deviation *sd* for all values of p as is shown in Table 3.39, where the mean is 30.9 for $p = 0.5$ with 566 times the optimal case selected ($\sum = 21$). For $m = 3$ the 24 candidates are observed in 8 groups of 3 candidates, the performance continued to improve except for $p = 0.9$ which recorded a slight worse performance than in the case of $m = 2$, however this could have resulted from the fact that by increasing the size of group better candidates can be lost. The high value $p = 0.9$ means that we accept more early candidates, so sometimes the better candidates are in later groups while we accept more in the beginning of the process. The same occurred again for $m = 4$ and $p = 0.8$ as presented in Figure 3.19 and in Table 3.39 where the mean increased from 31.5 for $m = 3$ to 32.2 for $m = 4$, while a better performance is clearly shown for all other values of p, with lowest mean and *sd* for $p = 0.6, 0.632$ and highest frequency of selecting the optimal 6 candidates by scoring 975 times for each.

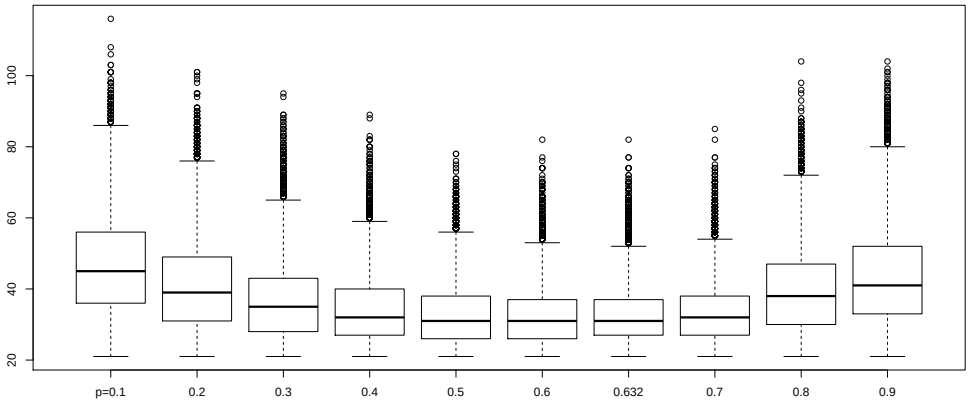

Figure 3.16: candidates in groups with $N = 24, c = 6$ and $m = 1$

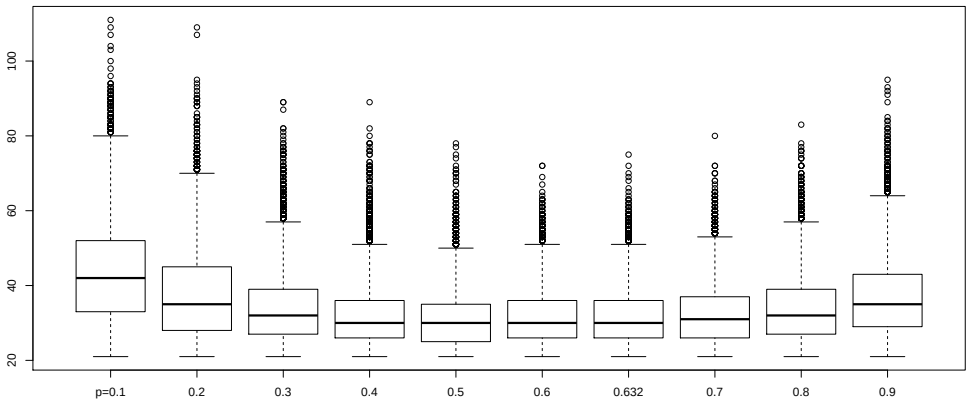

Figure 3.17: candidates in groups with $N = 24, c = 6$ and $m = 2$

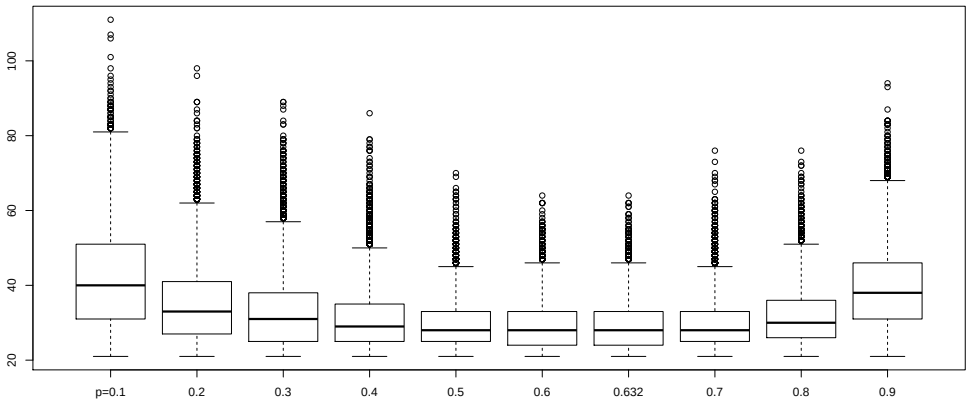

Figure 3.18: candidates in groups with $N = 24, c = 6$ and $m = 3$

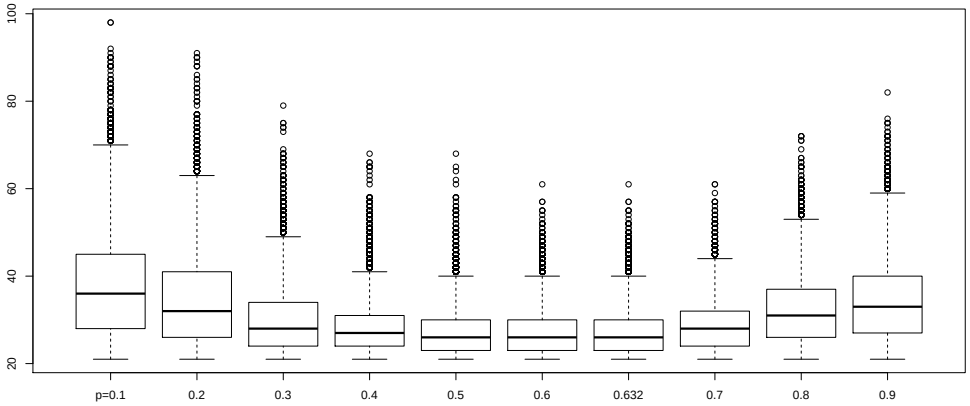

Figure 3.19: candidates in groups with $N = 24, c = 6$ and $m = 4$

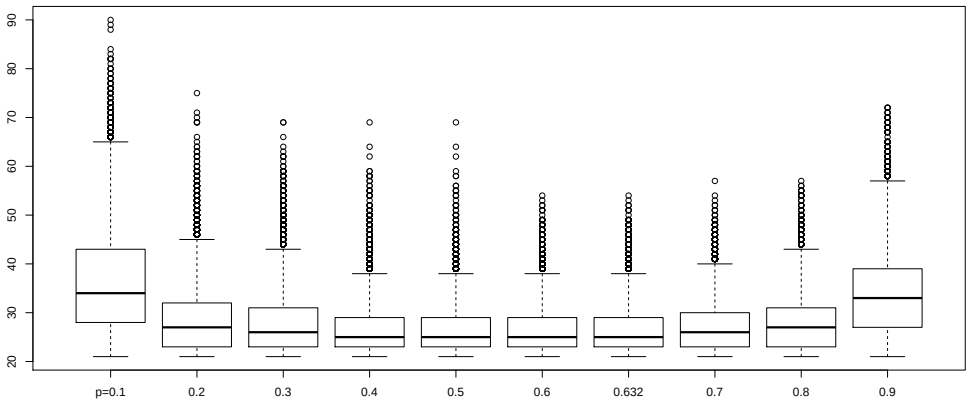

Figure 3.20: candidates in groups with $N = 24, c = 6$ and $m = 6$

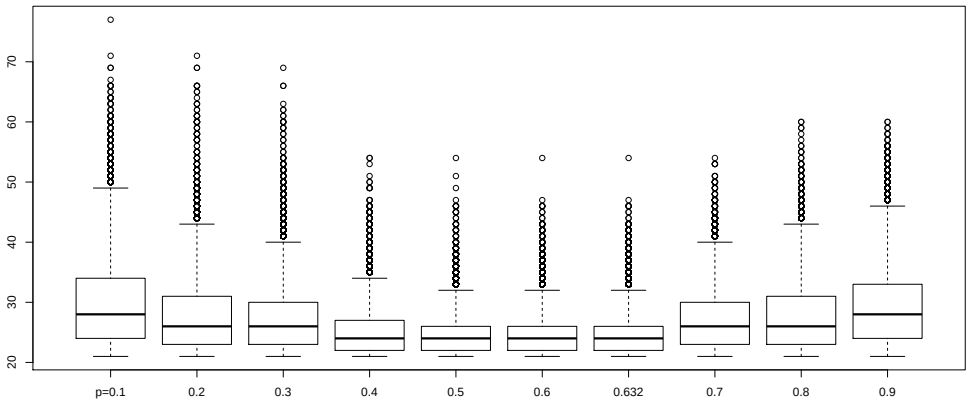

Figure 3.21: candidates in groups with $N = 24, c = 6$ and $m = 8$

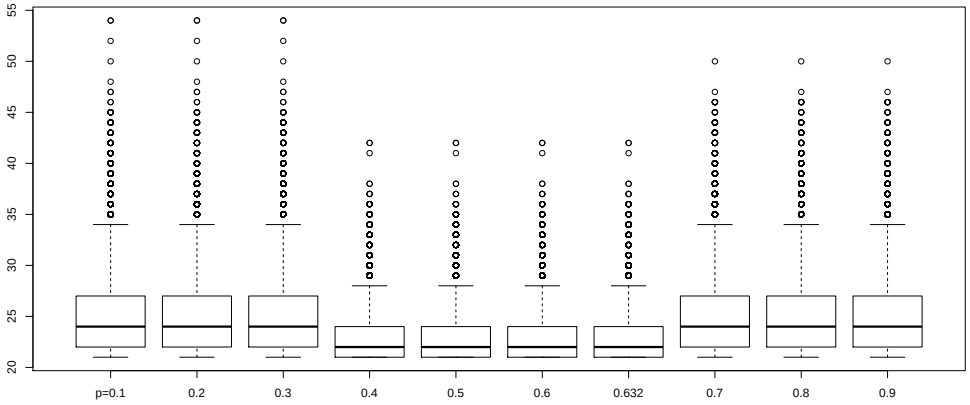

Figure 3.22: candidates in groups with $N = 24, c = 6$ and $m = 12$

In Figure 3.20 the 24 arrived candidates are observed in 4 equal groups, each with 6 candidates, where the means and standard deviations of the selections are rather lower than for all previous cases $m = 1, 2, 3, 4$ with better selection for the optimal case for most values of p where the frequencies started at 278 times for $p = 0.1$, peaked at 1430 for $p = 0.5$ and ended at 448 for $p = 0.9$ which is a significant improvement in the performance. The size of groups m is increased in Figure 3.21 to be $m = 8$, this increase leads to further improvement at all levels of p with selecting the best six candidates 2184 times for $p = 0.6, 0.632$ as can be seen in Table 3.39.

The performance in Figure 3.22 has reached the peak when all candidates are divided into only two groups, each of $m = 12$ candidates. The shape of Figure 3.22 is different than the usual cases as it is divided into three groups each group with the same performance. First group when $p = 0.1, 0.2, 0.3$, the second when $p = 0.4, 0.5, 0.6, 0.632$ and the third when $p = 0.7, 0.8, 0.9$ the results for each group are exactly the same with mean 24.9, 23 and 24.9, respectively. The reason for that is when $m = 12$ we have two groups, when $p \leq 0.3$ we accept later candidates which means that the 6 selected candidates are selected from the second group and in case $p \geq 0.7$ we usually accept earlier candidates which means that the 6 selected candidates are selected from the first group, while when $0.4 \leq p \leq 0.632$ the 6 candidates are selected from the two groups. To sum up, results show that larger group size m tends to be better with some exceptions in case of large p when we tend to accept early candidates.

In case $m = 12$, as we mentioned at the start of this section, to make a decision about whether or not to accept the j-th candidate, $j = 1, 2, ..., 12$ in the first group, the probability (3.8) is calculated for each candidate depending on his relative rank among all the candidates of the first group. Then the best candidate in the first group r_1^{12} is always accepted if $p > 0.0068$, as $P(Y_{13}^{24}(r_1^{12}) \geq 6) = 0.0068$. The second best r_2^{12} is always accepted as well if $p > 0.0774$, as $P(Y_{13}^{24}(r_2^{12}) \geq 5) = 0.0774$. While the candidate with relative rank 3, r_3^{12} is only accepted if $p > 0.3202$, as $P(Y_{13}^{24}(r_3^{12}) \geq 4) = 0.3202$. The candidates with relative ranks 4, 5 and 6 can be accepted if and only if $p > 0.6797$, $p > 0.9225$ and $p > 0.9931$, respectively. In

Figure 3.22, for values of p from 0.1 to 0.9 as considered before, there are only three different selections for all the runs in the simulations. For $0.1 \leq p \leq 0.3$ we accept 2 candidates from the first group and 4 candidates from the second group, if $0.4 \leq p \leq 0.632$ we accept 3 candidates from each group and if $0.7 \leq p \leq 0.9$ we accept 4 candidates from the first group then 2 candidates from the second group. Classically, to find the expected number of best selection in 10,000 runs we need to calculate the probability of the optimal selection in each option. Mathematically the first $p \leq 0.3$ and the third $p \leq 0.9$ options are exactly the same, where the probability is $\frac{\binom{12}{2}\binom{12}{4}}{\binom{24}{6}} = 0.2427$, so the expected number is 2427 which is very close to our results as presented in Table 3.39, where the best candidates selected 2453 times for $(0.1 \leq p \leq 0.3)$ and 2426 times for $(0.7 \leq p \leq 0.9)$. While if a moderate p is used then 3 candidates are accepted from each group, thus the probability is $\frac{\binom{12}{3}\binom{12}{3}}{\binom{24}{12}} = 0.3595$ and the expected number is 3596 which is also very close to our result 3598 as presented in Table 3.39.

In the case of either accepting later ($p \leq 0.3$) or earlier ($p \geq 0.7$) candidates the probability of the optimal selection is 0.2427. Let Y be the number of the optimal selection in the 10,000 orders, then Y has a Binomial distrubution with parameters $(10000, 0.2427)$ and approximately normally distributed with parameters $(2427, (42.87152)^2)$, then the confidence interval for the expected number of cases out of 10,000 in which the optimal selection is made is $[2343, 2511]$. In addition, the confidence interval for the expected number if $0.4 \leq p \leq 0.632$ is $[3502, 3690]$.

$N = 24, c = 6$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$m = 1$										
$Mean$	46.8	41.3	36.6	34.5	32.8	32.8	32.8	33.4	39.8	43.5
sd	14.7	13.3	11.4	10.1	8.7	8.4	8.3	8.7	12.0	13.9
$Best$	104	220	388	426	410	435	414	324	202	114
$m = 2$										
$Mean$	43.2	37.5	33.7	31.9	30.9	31.4	31.5	32.1	33.9	36.9
sd	13.6	12.1	9.8	9.51	7.5	7.6	7.6	8.0	8.9	10.6
$Best$	170	334	427	562	566	524	500	463	316	230
$m = 3$										
$Mean$	41.9	35.2	33.7	31.3	29.2	29.0	29.1	29.7	31.5	39.3
sd	13.4	10.9	9.9	8.2	6.4	6.1	6.2	6.7	7.6	10.9
$Best$	178	434	552	589	747	739	711	652	491	147
$m = 4$										
$Mean$	38	34.8	29.9	28.8	27.6	27.5	27.5	28.5	32.2	34.6
sd	12.3	11.1	7.6	6.1	5.6	5.3	5.3	5.8	8.06	9.22
$Best$	339	486	770	956	973	975	975	854	420	358
$m = 6$										
$Mean$	36.4	28.7	27.9	26.5	26.5	26.6	26.6	27.1	28.0	33.9
sd	10.9	7.1	6.6	5.2	5.0	5.0	5.0	5.2	5.9	8.9
$Best$	278	1114	1092	1397	1430	1292	1292	1048	1147	448
$m = 8$										
$Mean$	29.9	28.1	27.5	24.9	24.6	24.7	24.7	27.1	27.6	29.3
sd	7.8	6.9	6.1	4,02	3.7	3.8	3.8	5.4	5.8	7.0
$Best$	1055	1215	1178	2158	2025	2184	2184	1343	1091	1079
$m = 12$										
$Mean$	24.9	24.9	24.9	23.0	23.0	23.0	23.0	24.9	24.9	24.9
sd	4.4	4.4	4.4	5.13	5.13	5.13	5.13	4.3	4.3	4.3
$Best$	2453	2453	2453	3598	3598	3598	3598	2426	2426	2426

Table 3.39: Means and standard deviation with different m and p

The discussion above focuses on selecting multiple candidates observed in groups of equal size. Here, we apply our NPI method to select multiple candidates from randomly arriving candidates, the candidates now observed in groups of different sizes. We illustrate how this works by again considering the example with $N = 24$ candidates discussed before. Suppose that we are again interested in selecting $c = 6$ candidates. We consider the reverse order of absolute ranks as given in Example 3.2. The results with different p's and group sizes are presented in Tables 3.40-3.44, which give the sum $\sum$ of the absolute ranks of the selected candidates. These ranks are listed under 'Who' in order of their selection. In Table 3.40 the considered sizes of groups are $m = 2, 3, 5, 4, 6, 3, 1$. In this procedure, for $p = 0.1$, the first two candidates with absolute ranks 22 and 23 are observed first. Using probability (3.8), the best candidate in this first group (with relative rank 1, of course) there will be enough better candidates in the future groups, with probability at least 0.1, thus the best candidate in the first group is rejected, of course no further candidates need to be checked in this group. The second group is of size $m = 3$, which means that candidates with absolute ranks 13, 11 and 21 are observed together at the same time, however the best candidate in this group has absolute rank 11 and relative rank 1 and is still rejected. Then the third group is considered which is of size $m = 5$ and includes candidates with absolute ranks 24, 6, 5, 15 and 20. The best candidate in this group with absolute rank 5 has relative rank 1 and is accepted. Then the second best with relative rank 2 is checked and accepted as well, but the third with relative rank 5 has been rejected. Thus the fourth group is considered with size $m = 4$, which includes the best candidate who is the selected one in this group. The followed group includes 6 candidates the best with absolute rank 4 is accepted but the second best who with absolute rank 9 is rejected, thus we move to the sixth group. This group of size 3, the candidates with absolute ranks 3, 7 and 2, the best candidate in this group is accepted. Five candidates have been selected so far $a = 5$ and we still need another candidate $c = 6$. Despite that the sixth group includes the second and the third best, the last candidate is taken from the last group which of size 1. The sum of the selected absolute ranks therefore is $\sum = 30$. For a higher quality requirement $p = 0.5$, an early candidate who is the best in the

second group, with absolute rank 11 is accepted to be the first selected one, then the candidates with absolute ranks 6 and 5, 1, 4 and 2 are taken from the third, fourth, fifth and sixth respectively, to achieve a sum of absolute ranks $\sum = 29$. For the highest quality requirement $p = 0.99$, which is indicate to accept more earlier candidates, the first and second groups of size $m = 2$ and $m = 3$ candidates are all selected, with ranks 22, 23, 13, 11 and 21, the the best one in the followed group is selected to reach to 95 as the sum of the selected absolute ranks. In Table 3.44 we changed the order of two absolute ranks 17 and 3 to be 3 and 17, so the fifth group still includes 6 candidates, but now they are 4,19,10,9,14,3 and the sixth group still has 3 candidates, but now they are 17, 7, 2. Some of the results have been influenced by this change, for example, for $p = 0.1$ the quality of selection minimised to the optimal selection $\sum = 21$ instead of $\sum = 30$, as during the selection process the first 3 selected candidates are still the same (candidates with absolute ranks 6, 5, 1), but in the fifth group the best candidate in this group became with absolute rank 3 and the second best became with absolute rank 4, whom both have been taken. Then the best candidate in the last group is selected from the sixth group who has absolute rank 2. Also for $p = 0.3$, the fifth accepted candidate is with absolute rank 3 instead of 4, as the best candidate in this group became 3 and it is checked first. The random order of candidates in each group therefore can affect the selection process, then the quality of selection either Positively or negatively, depending on the position of each group.

Table 3.41 shows the same procedure with different size of groups, the considered sizes here are $m = 5, 5, 3, 2, 3, 3, 3$. For $p = 0.1$, the first group includes 5 candidates are all rejected, while two candidates with absolute ranks 5 and 6 are selected from the second group, the best candidate from the third group, 3 from the sixth group and the last two candidates are selected from the last group, with absolute ranks 2 and 7, with a quality of selection $\sum = 24$, which is better than the previous size of groups which considered in Table 3.40.

Despite that small p leads to accept later candidates, the quality of selection is improved when we consider early large size of groups as in Table 3.43. The considered sizes were $m = 7, 6, 5, 1, 2, 3$, the optimal selection $\sum = 21$, has been

made for $p = 0.1$ and $p = 0.2$. While, in Table 3.42 the large size is considered later, the optimal selection has been made once with $p = 0.1$. Furthermore, with the early large size of groups as in Table 3.43, high level of p gives a better quality of selection, comparing with the later large sizes as in Table 3.42. For example, for $p = 0.8$, $p = 0.9$ and $p = 0.99$ the presented quality of selection in Table 3.43 are 40, 40 and 74 respectively, while in Table 3.42 the corresponding quality of selections are 58, 58 and 95, respectively. Therefore, in order to get a better quality of selection especially with high and low p, it is better to observe the largest possible number in the beginning of the process.

$\sum$	Who						p
30	6	5	1	4	2	12	0.10
21	6	5	1	4	3	2	0.20
29	11	6	5	1	4	2	0.30
29	11	6	5	1	4	2	0.40
29	11	6	5	1	4	2	0.50
49	22	11	6	5	1	4	0.60
58	22	13	11	6	5	1	0.70
58	22	13	11	6	5	1	0.80
58	22	13	11	6	5	1	0.90
95	22	23	13	11	21	5	0.99

Table 3.40: $m = 2, 3, 5, 4, 6, 3, 1$

$\sum$	Who						p
24	6	5	1	3	7	2	0.10
21	6	5	1	4	3	2	0.20
28	11	6	5	1	3	2	0.30
30	11	6	5	1	4	3	0.40
30	11	6	5	1	4	3	0.50
39	13	11	6	5	1	3	0.60
40	13	11	6	5	1	4	0.70
40	13	11	6	5	1	4	0.80
40	13	11	6	5	1	4	0.90
78	22	13	11	21	6	1	0.99

Table 3.41: $m = 5, 5, 3, 2, 3, 3, 3$

$\sum$	Who						p
21	6	5	1	4	3	2	0.10
29	11	6	5	1	4	2	0.20
29	11	6	5	1	4	2	0.30
29	11	6	5	1	4	2	0.40
38	13	11	6	5	1	2	0.50
40	13	11	6	5	1	4	0.60
40	13	11	6	5	1	4	0.70
58	22	13	11	6	5	1	0.80
58	22	13	11	6	5	1	0.90
95	22	23	13	11	21	5	0.99

Table 3.42: $m = 1, 2, 3, 5, 3, 3, 7$

$\sum$	Who						p
21	6	5	1	4	3	2	0.10
21	6	5	1	4	3	2	0.20
27	6	5	1	8	4	3	0.30
30	11	6	5	1	4	3	0.40
35	11	6	5	1	8	4	0.50
40	13	11	6	5	1	4	0.60
40	13	11	6	5	1	4	0.70
40	13	11	6	5	1	4	0.80
40	13	11	6	5	1	4	0.90
74	22	13	11	21	6	1	0.99

Table 3.43: $m = 7, 6, 5, 1, 2, 3$

$\sum$	Who						p
21	6	5	1	4	3	2	0.10
21	6	5	1	4	3	2	0.20
28	11	6	5	1	3	2	0.30
30	11	6	5	1	4	3	0.40
30	11	6	5	1	4	3	0.50
48	22	11	6	5	1	3	0.60
58	22	13	11	6	5	1	0.70
58	22	13	11	6	5	1	0.80
58	22	13	11	6	5	1	0.90
95	22	23	13	11	21	5	0.99

Table 3.44: $m = 2, 3, 5, 4, 6, 3, 1$

3.6 Comparison to a fixed-percentage method

The heuristic method presented in this chapter is particularly aimed at problems with many candidates, for which dynamic programming cannot be applied. Of course, other heuristic rules have been suggested for this problem [23, 37, 39] and research on such methods has mostly been focused on their probabilistic properties [27, 31, 32]. We have compared our NPI method with several alternatives that are also computationally straightforward.

In this section we consider the straightforward rule that candidate j is accepted if and only if $r_j \leq \gamma \times j$ for some value $\gamma \in (0, 1)$, of course only up to the point when c candidates are selected and automatically selecting final candidates if there are no other options to ensure that we select c candidates in total. Let us call this the 'gamma-rule'.

Table 3.45 provides a study of this selection procedure for $N = 10, c = 3$ with various values of γ. It shows the maximum relative rank of a candidate who will be selected in this case. The first candidate will never be selected in this case as his relative rank $r_1 = 1$, is never less than γ, for $0 < \gamma < 1$. If $\gamma = 0.1$, the last 3 candidates $j = 8, 9$ and 10 are inevitably selected. For $\gamma = 0.2$ the fifth candidate can be selected if he has relative rank 1 or the following candidate can be accepted if he is the best candidate thus far and so on. It is worth to note that the worst candidate cannot be accepted with any value of γ, except of course if needed at the end of the process.

For a simulation study with 10,000 independent runs with $N = 10$ the results are shown in Table 3.46. The gamma-rule clearly performs better than NPI for $\gamma \geq 0.5$, where the optimal selection has been selected 2634, 2769, 2575, 2578, 2190 and 1660 for $\gamma = 0.5, 0.6, 0.632, 0.7, 0.8$ and 0.9, respectively.

According to this simulation for $N = 10, c = 3$ in Table 3.46 and further simulations, they all showed that the gamma-rule performs well for small values of N, but for larger values of N its performance is substantially worse than our NPI method. Tables 3.47, 3.48, 3.49, 3.50, 3.51 and 3.52 show the summary statistics of the sums of absolute ranks of the selected candidates in simulations for the case with $N = 10, c = 3$; $N = 60, c = 10$; $N = 100, c = 25$; $N = 200, c = 20$; $N = 200, c = 40$;

j	Select j if $r_j \leq$								
	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
1	-	-	-	-	-	-	-	-	-
2	-	-	-	-	1	1	1	1	1
3	-	-	-	1	1	1	2	2	2
4	-	-	1	1	2	2	2	3	3
5	-	1	1	2	2	3	3	4	4
6	-	1	1	2	3	3	4	4	5
7	-	1	2	2	3	4	4	5	6
8	A	1	2	3	4	4	5	6	7
9	A	1	2	3	4	5	6	7	8
10	A	2	3	4	5	6	7	8	9

Table 3.45: γ percentile rule for 3 out of 10

$N = 200, c = 60$ and $N = 200, c = 80$ respectively, with for simplicity the same values chosen for γ as for p in our NPI method. For example, when $N = 10, c = 3$ gamma-rule provides better candidates than NPI, especially with $\gamma \geq 0.5$ as illustrated in the summary statistics in Table 3.46, but these results gradually changed with increasing values N and c. Boxplots for the results of these simulations are presented in Figures 3.23, 3.24, 3.25, 3.26, 3.27, 3.28, respectively. Let us consider the best possible selection in the case of $N = 200, c = 60$, namely the candidates with absolute ranks 1 to 60, leads to sum of 1830, so our NPI method clearly performs very well for p in the range 0.4 to 0.7. In these 10,000 simulations for each value of p, this best possible selection was actually made three times, twice for $p = 0.5$ and once for $p = 0.7$. For the gamma-rule the performance is substantially weaker, and the optimal selection was never made. The main reason why the NPI method performs better in larger N is that it continuously adapts to the number of candidates already selected and the number still to come, which the gamma-rule does not do. While for small numbers of candidates this is not necessarily very important, for large numbers such as in this simulation the positive impact of these adaptations is very substantial.

$N = 10, c = 3$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	16.5	15.0	13.8	12.0	9.8	9.2	9.2	8.6	8.6	9.1
sd	4.4	4.4	4.4	4.2	3.7	3.4	3.3	2.8	2.3	2.4
$Best$	86	179	340	1250	2634	2769	2575	2578	2190	1660
$Worst$	101	58	37	16	6	1	1	0	0	0
NPI										
$Mean$	13.8	12.3	11.2	10.4	10.1	9.9	9.9	10.2	11.9	12.6
sd	4.5	4.2	3.8	3.6	3.3	3.1	3.1	3.1	4.3	4.3
$Best$	476	771	1127	1517	1494	1499	1444	1279	941	665
$Worst$	34	12	2	2	1	0	0	0	14	14

Table 3.46: Gamma-rule and NPI for $N = 10, c = 3$: summary statistics

$N = 60, c = 10$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	251.1	180.8	112.0	81.7	86.0	93.9	96.3	101.2	108.0	114.7
sd	49.7	48.7	43.8	18.6	12.9	15.2	15.8	17.3	19.1	20.7
$Best$	0	3	55	29	2	0	0	1	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	132.0	110.0	96.9	88.0	83.4	82.4	82.4	87.3	91.4	120.5
sd	43.3	36.4	30.6	25.2	20.8	19.3	19.1	23.2	26.0	40.8
$Best$	21	47	95	111	124	113	109	86	65	17
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.47: Gamma-rule and NPI for $N = 60, c = 10$: summary statistics

$N = 100, c = 25$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	1135.4	978.1	816.2	634.6	476.5	468.3	479.0	503.3	537.8	570.3
sd	122.4	120.0	120.0	121.1	82.1	41.0	45.7	41.7	52.1	57.7
$Best$	0	0	0	0	0	0	0	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	502.0	451.8	424.0	406.0	397.4	399.0	400.1	404.65	443.4	480
sd	94.0	76.4	63.0	52.0	45.0	45.5	46.4	49.5	71.9	89.7
$Best$	1	3	14	13	17	16	12	6	7	4
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.48: Gamma-rule and NPI for $N = 100, c = 25$: summary statistics

$N = 200, c = 20$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	1210	387.4	335.1	386.1	431.5	473.4	485.9	511.2	545.5	578.2
sd	241.9	165.9	36.7	47.7	56.3	64.2	66.6	71.4	77.6	83.7
$Best$	0	0	0	0	0	0	0	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	476.9	389.4	340.1	308.0	290.9	286.3	285.7	291.9	312.5	369.6
sd	155.4	124.6	100.9	79.0	63.0	55.8	54.2	58.5	75.4	111.0
$Best$	5	6	10	13	9	10	12	10	15	3
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.49: Gamma-rule and NPI for $N = 200, c = 20$: summary statistics

$N = 200, c = 40$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	3391	2684	1938	1225	1210	1324	1358	1430	1527	1620
sd	317	313	315	217	85	101	106	116	128	141
$Best$	0	0	0	0	0	0	0	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	1201	1086	1021	983	964	962	964	986	1016	1156
sd	200	158	127	103	87	85	86	105	126	187
$Best$	0	0	1	0	5	1	1	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.50: Gamma-rule and NPI for $N = 200, c = 40$: summary statistics

$N = 200, c = 60$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	5558	5020	4461	3850	3171	2580	2534	2617	2798	2965
sd	365	354	347	353	366	254	176	140	162	182
$Best$	0	0	0	0	0	0	0	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	2290	2155	2084	2040	2025	2025	2028	2047	2139	2233
sd	234	185	149	122	109	110	112	127	180	222
$Best$	0	0	0	0	2	0	0	1	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.51: Gamma-rule and NPI for $N = 200, c = 60$: summary statistics

$N = 200, c = 80$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
$\gamma - percentile$										
$Mean$	7698	7311	6908	6478	6003	5511	5348	4962	4459	4560
sd	392	383	372	366	368	372	377	389	290	204
$Best$	0	0	0	0	0	0	0	0	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0
NPI										
$Mean$	3739	3602	3523	3477	3461	3466	3490	3518	3569	3712
sd	251	201	162	134	121	123	137	157	187	244
$Best$	0	0	0	0	1	0	0	1	0	0
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.52: Gamma-rule and NPI for $N = 200, c = 80$: summary statistics

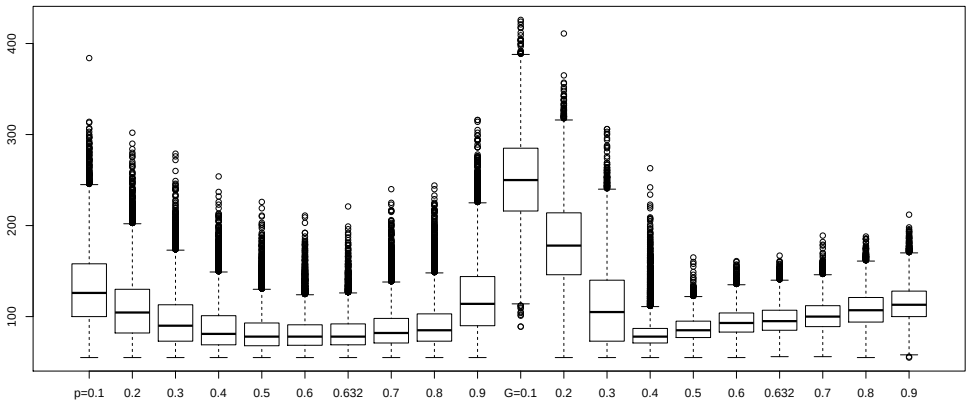

Figure 3.23: NPI (p) and Gamma-rule (G) for $N = 60, c = 10$

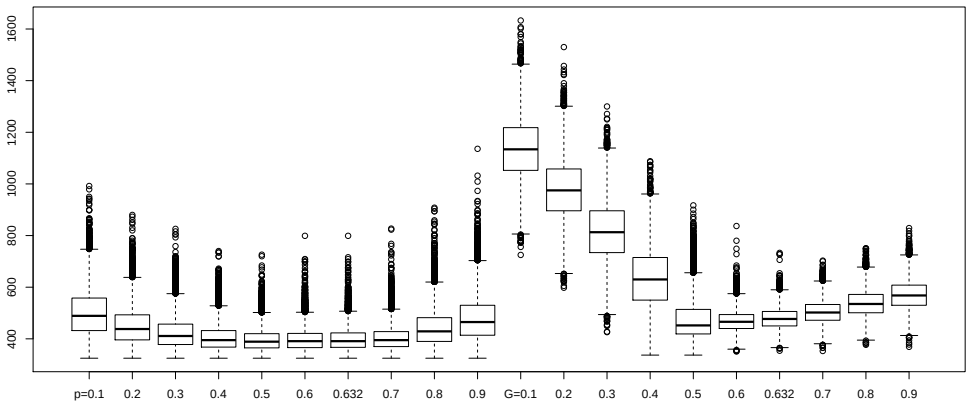

Figure 3.24: NPI (p) and Gamma-rule (G) for $N = 100, c = 25$

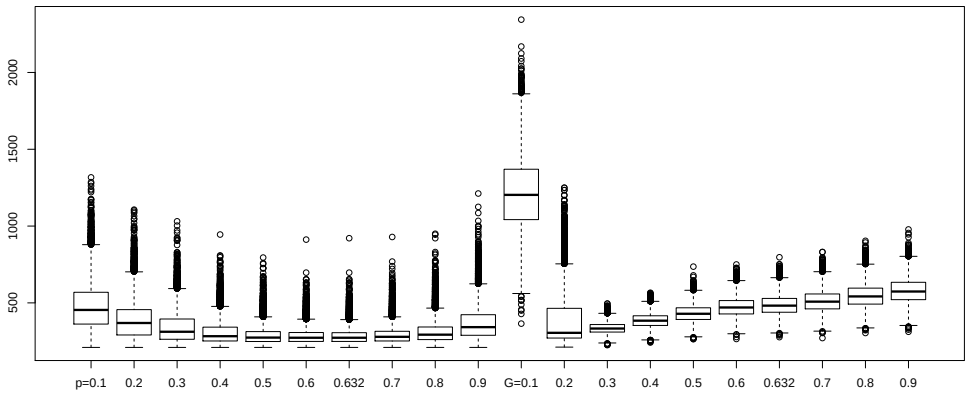

Figure 3.25: NPI (p) and Gamma-rule (G) for $N = 200, c = 20$

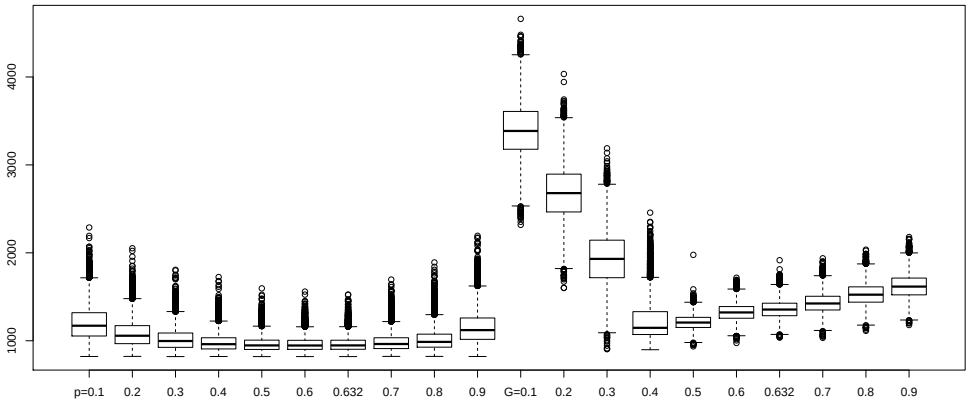

Figure 3.26: NPI (p) and Gamma-rule (G) for $N = 200, c = 40$

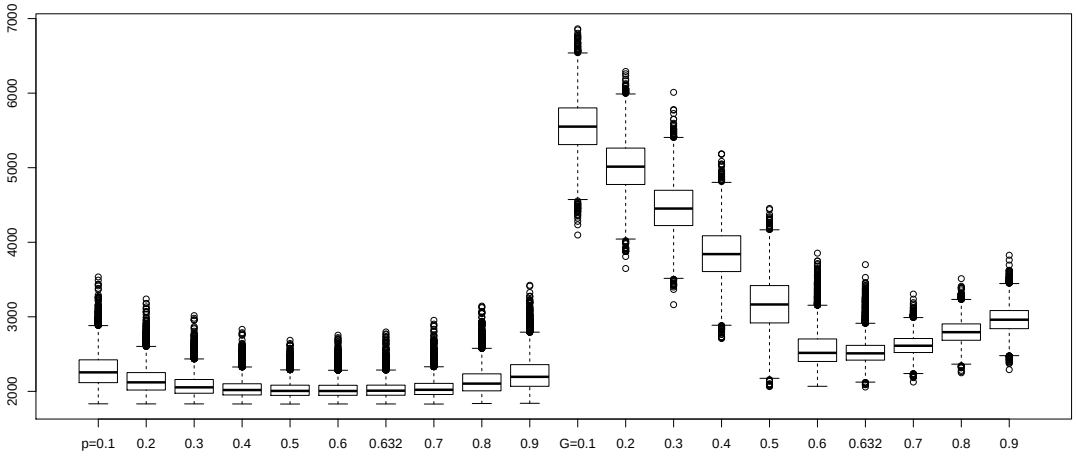

Figure 3.27: NPI (p) and Gamma-rule (G) for $N = 200, c = 60$

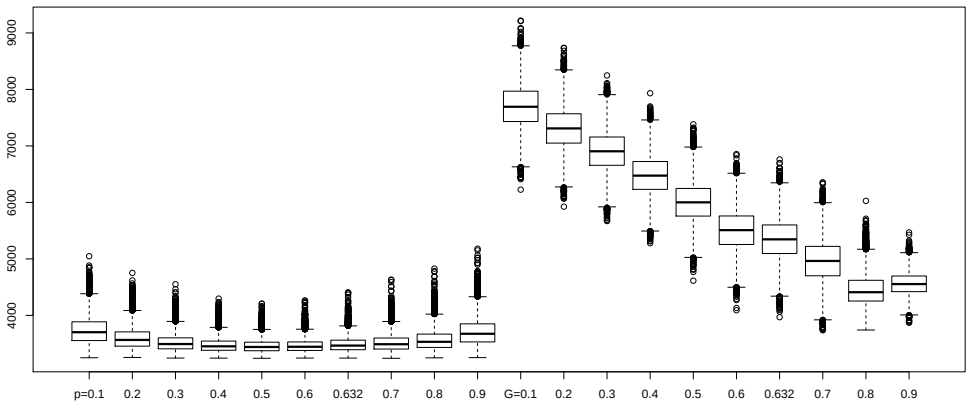

Figure 3.28: NPI (p) and Gamma-rule (G) for $N = 200, c = 80$

3.7 Improving the subset of accepted candidates

In this section, we consider the slightly more complicated rules which are such that candidate j is selected if and only if his relative rank is lower than or equal to the median or the mean of the relative ranks, at that stage j, of the already selected candidates. So for these rules the relative ranks of already selected candidates are updated at every stage, and a candidate is selected if that improves (or keeps the same) the median or mean of the relative ranks of all selected candidates. Again, with these rules one also stops once c candidates have been selected, and one accepts automatically the final candidates needed to ensure selection of c candidates.

To explain the idea, suppose that we want to select $c = 3$ candidates from a total of $N = 7$ candidates. Suppose further the absolute ranks for these 7 candidates are 5, 6, 4, 2, 1, 3 and 7, and we will observe them sequentially. Let us start by the median first, so candidate j, is accepted if and only if his relative rank is less than or equal to the median of the relative ranks of the already selected candidates. So, the first observed candidate is the best so far and his relative rank is equal to 1, which is equal to the median, thus the first observed is the first accepted candidate as it is always the case. The second candidate is worse than the first one, so the second candidate has relative rank 2, and the one selected candidate so far now still has relative rank 1. So the median is still 1, which means the second candidate is rejected. The third candidate is the best so far with relative rank 1 and the first accepted candidate now has updated relative rank 2, so the median of the accepted candidates is 2, which means the third candidate is selected as the second accepted candidate. The fourth observed candidate is better than all the observed ones, this also better than the already selected ones, so of course gets selected , leading to $\sum = 5 + 4 + 2 = 11$.

As for the mean, suppose again we have $N = 10$ candidates and we are interested in selecting only $c = 3$ candidates. Assume that the 10 candidates arrived randomly in the following order 3, 1, 9, 4, 2, 7, 8, 5, 6, 10. As before, the first observed candidate is accepted as his relative rank is equal to his mean. The second observed one is better than the first selected candidate, so his updated relative rank is 1 and the accepted candidate has updated relative rank 2. Thus the second observed

candidate is accepted. The third observed candidate has relative rank 3, as he is the worst yet, thus he is rejected, as the mean of the accepted candidate is 1.5, the fourth candidate is also rejected. The relative rank of the fifth candidate is 2 and the updated mean of the accepted ones is 2, which means the fifth is accepted and the process is stopped, with $\sum = 3 + 1 + 2 = 6$.

These two methods, indicated by 'Median' and 'Mean', were also compared to NPI via simulations, the results for the case with $N = 200$ and $c = 20$ are presented in Figure 3.29. As was the case when the gamma-rule was used, these two methods perform worse than the NPI method and this is mainly because they do not adapt to the number of candidates already selected or to the number of candidates still to come in the process.

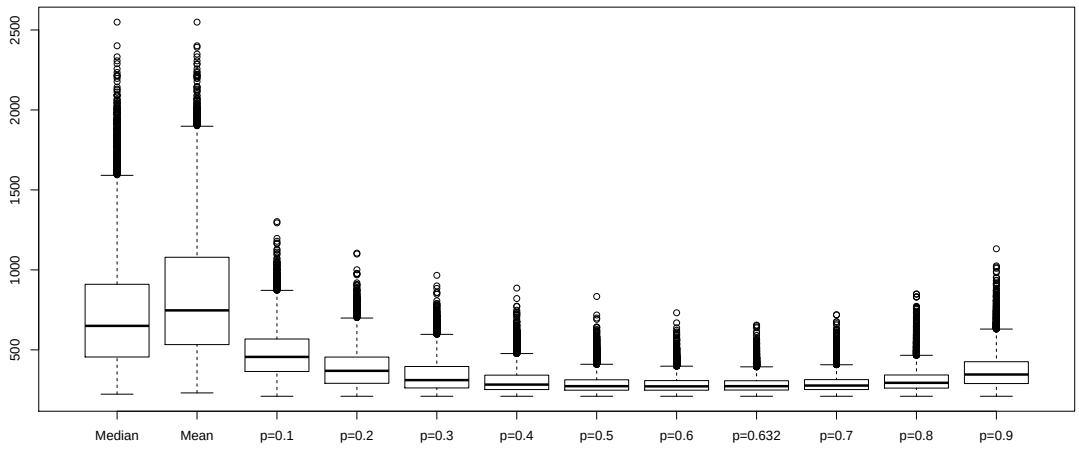

Figure 3.29: Median and mean improving methods and NPI for $N = 200, c = 20$

$N = 200, c = 20$	0.1	0.2	0.3	0.4	0.5	0.6	0.632	0.7	0.8	0.9
NPI										
$Mean$	477.6	387.8	339.1	307.8	290.3	285.2	284.8	291.7	313.2	372.3
sd	152.7	122.0	99.5	78.5	61.9	53.3	52.0	58.5	75.8	112.5
$Best$	2	7	10	6	17	12	8	8	3	3
$Worst$	0	0	0	0	0	0	0	0	0	0

Table 3.53: Means and standard deviation with different p, 20 out of 200

3.8 A randomised procedure

The heuristic NPI procedure presented in this chapter is fully based on the probability (3.1) for the event that there are at least $c - a$ (the remaining required number) candidates still to come who are better than the currently considered candidate j. The procedure is deterministic in the sense that candidate j is accepted or rejected with certainty, based on comparison of probability (3.1) with a chosen threshold value p. If the complementary event happens, namely that the number of remaining candidates better than candidate j is less than the required number yet to be selected, it would be best to select candidate j. Therefore, one could consider explicitly that candidate j is needed with probability

$$P(Y_{j+1}^N(r_j) < c - a) = 1 - P(Y_{j+1}^N(r_j) \geq c - a) \tag{3.9}$$

This suggests a randomised selection procedure, as alternative heuristic method to the one presented above, where candidate j is accepted with probability (3.9) and thus rejected with probability (3.1). We illustrate this procedure below for selection of candidates who are considered one-at-a-time, we also compare it to our deterministic heuristic method as introduced above. It can also be implemented for candidates arriving in groups, but it would of course not be sensible to then randomise for each candidate individually, as that may lead to the best candidate on one day being rejected and the second best candidate on the same day being accepted. Hence, to apply such a randomised heuristic method for the case with candidates arriving in groups, we recommend to first decide on the best candidate for the day via randomisation, and only if this candidate is accepted to consider selecting the next candidate too, and so on. This is easy to implement with a similar adaptation of the algorithm as presented in Section 3.5.

Restricting attention to the case of candidates arriving one-at-a-time, this randomised procedure has the important property that the probability of candidate j being accepted decreases as function of his relative rank r_j. Of course, the deterministic procedure proposed in Section 3.3 also has this property, but with the probability of candidate j being accepted equal to 0 or 1, for high or low values of r_j respectively. The performance of the randomised procedure is investigated below,

including a comparison with the deterministic procedure. One clear advantage is that it does not require choice of the somewhat arbitrary value of the threshold p used in the deterministic procedure, so arguably the randomised procedure is even more straightforward to implement.

To illustrate this idea a small example is provided. Suppose that a total of $N = 7$ candidates arrive in a random order as explained in Figure 3.30 and we are interested in selecting $c = 2$ candidates. The interviewer checks them one by one. The relative rank of the first candidate is 1, in order to make the decision to either accept or reject this candidate the probability 3.1 must be calculated. This is 0.86 which is less than the random selected value $q_1 = 0.93$ drawn randomly from the uniform distribution, so the first candidate is accepted, as indicated on the figure as (A). The second candidate is worse than the first candidate so he has relative rank 2, with probability 3.1 equal to 0.95 and a random selected value 0.39, which indicates to the second candidate is rejected as $P > q_2$. Despite that the third candidate is better than the first and second observed candidtes and his probability is equal to 0.57 which is not a good indicator for getting a better candidates in the future ones, he got a smaller value of $q_3 = 0.08$. So, the third is rejected as well as the fourth and the fifth, while the probability of getting a better candidate after the sixth is 0.29 with a random value $q_6 = 0.68$, so the sixth candidate is taken as the second selected candidate. As illustrated below (3.30) the third and last candidates have accepted with true ranks 1 and 6 respectively.

Figure 3.31 presents the results for a simulation study for 10,000 runs to select 40 candidates out of 200 candidates by our NPI method and the randomised procedure (R). This randomised procedure does not perform well for our sequential acceptance problems when compared to the NPI method. This is due to the fact that the effect of such randomisation is that excellent candidates have a positive probability of being rejected while poor candidates have a positive probability of being accepted, both of these decisions do not help to select the strongest group.

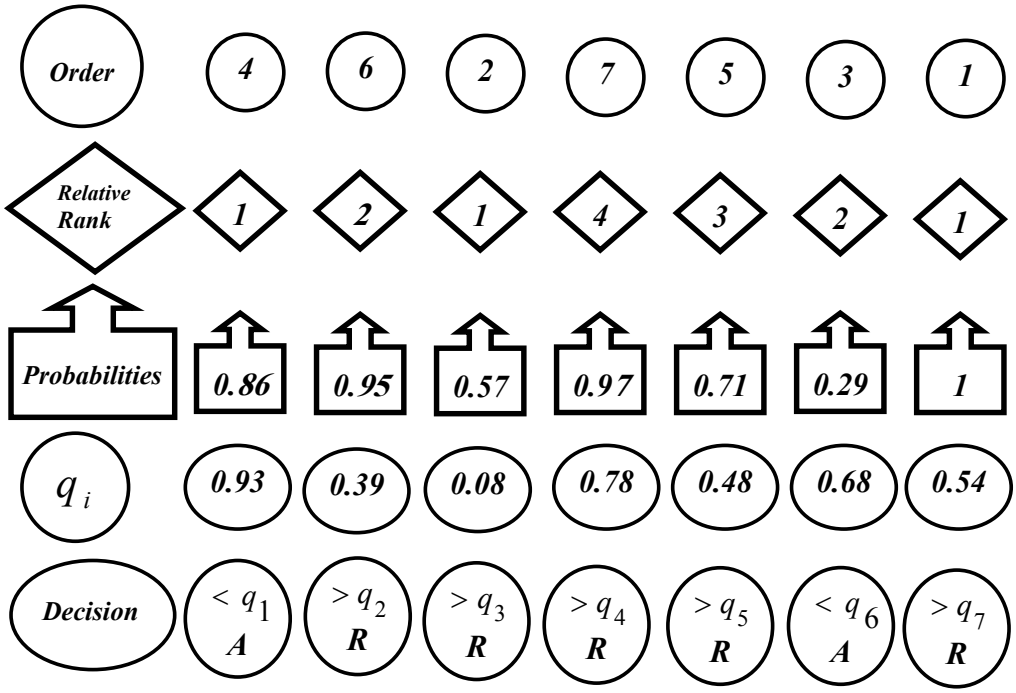

Figure 3.30: Randomised procedure 2 out of 7

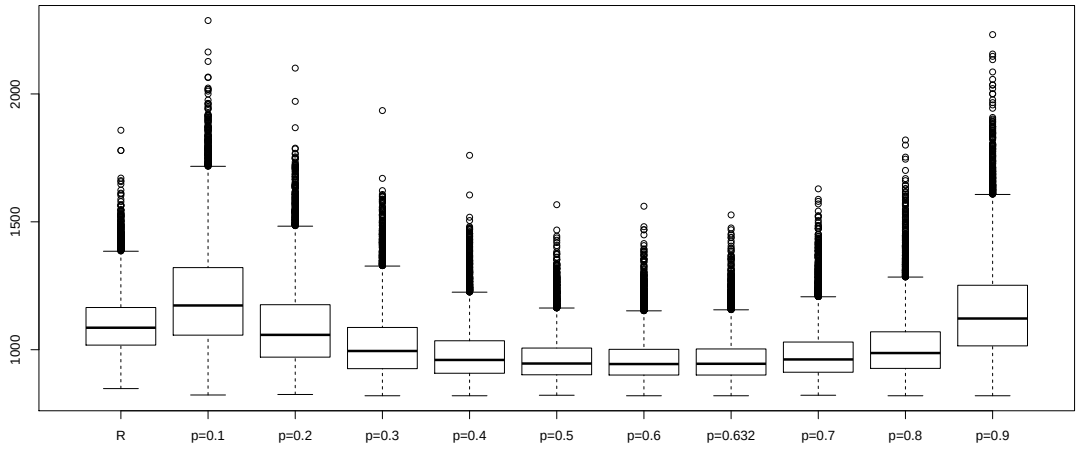

Figure 3.31: Simulation for randomised procedure and NPI for $N = 200$, $c = 40$

3.9 Concluding remarks

The nonparametric predictive approach to sequential acceptance problems, as presented in this chapter, provides a heuristic solution with the advantage of fast computation and has been shown to perform well. The success of this approach suggests that similar heuristic methods could be of use in other problems where exact optimisation methods are computationally infeasible as is often the case for so-called online decision problems [38]. This opens up an interesting area of research, particularly for decision problems that can naturally be formulated in terms of future observations, as the NPI approach has been developed particularly for inference on future observations. Identifying suitable problems and developing NPI-based heuristic solution methods provide interesting research challenges that may have substantial impact.

Chapter 4

Conclusions and Future Research

This chapter provides a short summary of the main results presented in this thesis, and discusses important challenges for future research.

4.1 Conclusions

In this thesis we have introduced Nonparametric Predictive Inference (NPI) for two acceptance decisions problems, where new solutions for acceptance sampling for attributes and for sequentially selecting one or more candidates from a group were presented and discussed in detail.

In Chapter 2 we introduced a generalisation for the NPI approach for Bernoulli random quantities presented by Coolen [7] when the interest is in the number of successes in future items, with known information on the number of successes in tested items. We presented the application of NPI to acceptance sampling for attributes, including two types of tests. The first type assumes that tested items cannot be used anymore, it is called 'destructive testing'. In the second type, a tested item that functioned well can still be used after the test, it is called 'non-destructive testing'. For a single sampling plan just one sample of size n was sampled from a batch of items of size N. An extension of the single stage sampling to two-stage sampling is also considered and discussed with numerical examples. In most acceptance sampling situations, the idea is to sample *a few* items to decide on acceptance of *many* items. In the NPI approach, however a large number of items need to be tested in

order to accept relatively few, this is due to the very high quality requirement and the fact that this must be proven by the data only.

In Chapter 3 the NPI approach to sequential acceptance problems was considered. First we considered selecting only one candidate from a group of N candidates who arrive in a random order. NPI has shown a good performance with the advantage of fast computation and flexibility. Some special cases were also provided and compared with our heuristic method. In case we are interested in selecting the best candidate, NPI-BO was suggested and compared with the optimal solution by Lindley. However, NPI-BO with a specific p gives an optimal performance to this problem exactly as Lindley's performance. Also, NPI with a non-constant quality requirement NPI-NC, was suggested in this chapter which has shown a good performance but not as good as the NPI with constant values. This approach has been extended in order to select $c > 1$ candidates. We started by considering the case of candidates arriving one-at-a-time. NPI provided a good performance especially with large N and c, compared to some other methods such as a fixed-percentage method, methods that improve the subset of accepted candidates and a randomised procedure. Of course, this is due to the fact that NPI takes into account how many candidates have been accepted so far and how many candidates one still needs to select.

Then we have explored the generalization of this situation with candidates being observed in either equal groups of size m or in if the groups varying in size. An example was used to show that due to the information for individual candidates not always being identical when using different group sizes, it is not always beneficial to see candidates in larger groups. Overall, in case of observing candidates in groups, NPI approach has achieved best performance for p in the range of $0.5 \leq p \leq 0.7$. Of course, in most cases it is better to see more candidates together as was shown in simulations.

4.2 Topics for future research

In this section we discuss a few possible areas for future research which build on the work presented in this thesis. All these areas would add to the NPI methodology for acceptance decisons.

4.2.1 Acceptance sampling

The results presented in Chapter 2 only consider very basic acceptance sampling scenarios, several possible generalizations provide interesting challenges for research. For example, inclusion of costs in (multi-stage) sampling provides interesting opportunities for research, while also careful consideration of the possible use of lower and upper probabilities in negotiations between producers and consumers raises interesting questions. For scenarios where observations remain categorical, but with more than two unordered categories, for example if specific failure modes are taken into account, the recently developed NPI theory for categorical data [11] can be applied. For real-valued observations, related theory of NPI with applications to statistical process control has been presented by [1, 2], generalization of this also provides interesting research challenges. Other statistical methods that quantify uncertainty via lower and upper probabilities have been presented in the literature, for example robust-Bayes like methods as presented by Walley [43, 44], and it will be interesting to apply such models to basic acceptance sampling problems and compare them with the results of the NPI approach presented in this chapter.

4.2.2 Sequential acceptance decisions

Generalization of the problem considered in Chapter 3 raises interesting questions. For example, if we do not only have the relative rank of a candidate but a real-valued measurement, then nothing would change with regard to probability (3.1) except we would get lower and upper probabilities [3] for a number of possibly interesting events, e.g. if we want to select a group with sum of absolute ranks at most a specific value, and we could consider other rules and criteria for 'good' selection. It is also of interest to study the use of more general loss functions (or utility functions)

related to such decision processes in our approach. Many variations to this sequential acceptance decision problem have been studied [23,40,41] and these also provide nice research challenges when considered in relation to our NPI-based heuristic method. Interesting topics include uncertainty in observations and multi-attribute rankings and utilities.

In addition to applications to a wider variety of problems, there are also interesting open questions, however, that could be addressed in future research. For example, one could attempt to design a (heuristic) solution to this problem such that the overall best candidate has (almost) the same probability of being accepted no matter at which stage j of the process he is considered. Deriving analytic solutions for such problems is difficult due to the obvious complexity in deriving such probabilities.

Bibliography

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

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

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

[4] Bayes T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, Royal Society of London*, **53**, 370–418; **54**, 296–325.

[5] Boole G. (1854). An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities. *Walton and Maberley; London*.

[6] Chow Y.S., Moriguti S., Robbins H., Samuels S.M. (1964). Optimal selection based on relative rank. *Israel Journal of Mathematics*, **2**, 81–90.

[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. (2006a). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, **15**, 21–47.

[9] Coolen F.P.A. (2006b). On probabilistic safety assessment in case of zero failures. *Journal of Risk and Reliability*, **220**, 105–114.

[10] Coolen F.P.A. (2007). Nonparametric prediction of unobserved failure modes. *Journal of Risk and Reliability*, **221**, 207–216.

[11] Coolen F.P.A., Augustin T. (2009). A nonparametric predictive alternative to the Imprecise Dirichlet Model: the case of a known number of categories. *International Journal of Approximate Reasoning*, **50**, 217–230.

[12] Coolen F.P.A., Coolen-Schrijner P. (2005). Nonparametric predictive reliability demonstration for failure-free periods. *IMA Journal of Management Mathematics*, **16**, 1–11.

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

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

[15] Coolen F.P.A., Elsaeiti M. (2009). Nonparametric predictive methods for acceptance sampling. *Journal of Statistical Theory and Practice*, **3**, 907-921.

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

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

[18] Dale B.G. (1999). Managing Quality (3rd ed.). *Blackwell; Oxford (UK).*

[19] Dempster A.P. (1963). On direct probabilities. *Journal of the Royal Statistical Society B*, **25**, 100–110.

[20] Dodge H.F., Romig H.G. (1998). Sampling Inspection Tables, Single and Double Sampling (2nd ed.). *Wiley; New York (US).*

[21] Dodge H. F. (1995). Chain sampling inspection plan. *Industrial Quality Control*, **4**, 10-13.

[22] Elsaeiti M., Coolen F. (2009). Nonparametric predictive inference for attribute acceptance sampling with destructive tests. In: *Proceedings of the 18th Advances in Risk and Reliability Technology Symposium*, L. Bartlett (Editor), pp. 212-220. Loughborough (UK).

[23] Freeman P.R. (1983). Dynamic programming and decision theory. *Applied Statistics*, **10**, 39-52.

[24] Gaafar L., Keats B. (1992). Statistical process control: a guide for implementation. *International Journal of Quality and Reliability Management*, **9, 4**, 9-20.

[25] Gilbert J., Mosteller F. (1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, **61**, 35-73.

[26] Goetsch D.L., Davis S.B. (2000). Quality Management: Introduction to Total Quality Management for Production, Processing, and Services. *Prentice Hall, Englewood Cliffs (US).*

[27] Henke M. (1973). Expectations and variances of stopping variables in sequential selection processes. *Journal of Applied Probability*, **10**, 786-806.

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

[29] Hill B.M. (1988). De Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference (with discussion). In: *Bayesian Statistics*, D.V. Lindley, J.M. Bernardo, M.H. DeGroot, A.F.M. Smith (Editor), **3**, 211–241. *Oxford University Press; Oxford (UK).*

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

[31] Krieger A.M., Pollak M., Samuel-Cahn E. (2007). Select sets: rank and file. *The Annals of Applied Probability*, **17**, 360-385.

[32] Krieger A. M., Pollak M., Samuel-Cahn E. (2008). Beat the mean: sequential selection by better than average rules. *Journal of Applied Probability*, **45**, 244-259.

[33] Lindley D.V. (1961). Dynamic programming and decision theory. *Journal of Applied Statistical Science*, **10**, 39-52.

[34] Montgomery D.C. (2004). Introduction to Statistical Quality Control (5th ed.). *Wiley; New York (US).*

[35] Moser L. (1956). On a problem of Cayley. *Scripta Mathematica.*, **22**, 289-292.

[36] Oakland J.S. (1986). Statistical Process Control. *Heinemann Professional Publishing Ltd; (UK).*

[37] Preater J. (2000). Sequential selection with a better-than-average rule. *Statistics & Probability Letters*, **50**, 187-191.

[38] Preater J. (2006). On-line selection of an acceptable pair. *Journal of Applied Probability*, **43**, 729-740.

[39] Samuels S. (1991). Secretary problems. In: *Handbook of Sequential Analysis.* B.K. Ghosh, P.K. Sen (Editor), Marcel Dekker, Inc., New York, Basel, Hong Kong, pp. 381–405.

[40] Tamaki M. (1979). Recognizing both the maximum and the second maximum of a sequence. *Journal of Applied Probability*, **16**, 803-812.

[41] Tamaki M. (1991). A secretary problem with uncertain employment and best choice of available candidates. *Operations Research*, **39**, 274-284.

[42] Trietsch D. (1999). Statistical Quality Control. *World Scientific.* Singapore

[43] Walley P. (1991). Statistical Reasoning with Imprecise Probabilities. *Chapman and Hall; London (UK).*

[44] Walley P. (1996). Inferences from multinomial data: learning about a bag of marbles (with discussion). *Journal of the Royal Statistical Society B.*, **58**, 3–57.

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

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

[47] Willemain T.R., Runger G.C. (1996). Designing control charts using an empirical reference distribution. *Journal of Quality Technology*, **28**, 31-38.

[48] Woodall W.H., Montgomery D.C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, **31**, 376-386.