Durham E-Theses

Nonparametric Predictive Inference for Selection and Ranking

ALSHIHRY, FATIMAH,IBRAHIM

Nonparametric Predictive Inference for Selection and Ranking

Fatimah Ibrahim Alshihry

A Thesis presented for the degree of

Doctor of Philosophy

Department of Mathematical Sciences

University of Durham

England

October 2025

Dedicated to

My family.

My lovely children: Feras, Saud and Jaser.

Nonparametric Predictive Inference for Selection and Ranking

Fatimah Ibrahim Alshihry

Submitted for the degree of Doctor of Philosophy

October 2025

Abstract

This thesis introduces Nonparametric Predictive Inference (NPI) for selection and ranking, based on a single future observation from each group, and focuses on two main contributions. First, the development of NPI procedures for various selection and ranking events. Secondly, the application of different loss functions to quantify the loss incurred from non-optimal selection and ranking decisions.

Initially, NPI is applied to rank the best groups within subsets. A selected subset refers to one or more independent groups that are better than the rest, where better means that all future observations from the groups in the selected subset exceed all the future observations from the non-selected groups. The 'independent group' means that information about the random quantities for one group does not provide any information about the random quantities for another group. Two special cases are considered: selecting a ranked subset of the two best groups and the three best groups. For the subset consisting of two groups, the exact NPI lower and upper probabilities are derived for the event that these are the two best groups, while for the subset consisting of three groups, approximate NPI lower and upper probabilities are derived for the event that these are the three best groups.

The thesis further explores a more general case of selection and ranking, addressing how to rank buckets containing one or more independent groups. Here, 'bucket' refers to a cluster or grouping of independent groups. This approach tackles two key questions: how to allocate groups to buckets, and how to determine the optimal number of buckets—defined as the allocation that maximises or minimises the

NPI lower and upper probabilities for a given event. Various allocation methods are evaluated, including those based on measures such as the median. Additionally, the NPI-Bootstrap method is used to estimate probabilities, to approximate the probability of the event of interest itself, rather than its lower and upper probabilities. Throughout the thesis, data from the literature illustrate and support the methods.

In this thesis, the NPI method is applied across various selection and ranking events, using different loss functions to quantify the loss incurred from non-optimal selection and ranking decisions. Uncertainty is quantified by calculating the NPI lower and upper expected losses for the events corresponding to these scenarios. In the selection scenario, zero-one, linear, and quadratic loss functions are used in both pairwise and multiple comparisons. Several selection events are considered, including selecting the best group, selecting the subset of best groups, and selecting the subset that includes the best group. In ranking scenarios, zero-one and general multi-level loss functions are applied to ranked subsets of best groups. The zero-one loss function provides a binary measure of whether the ranking is correct, while the general multi-level loss function allows for a more nuanced evaluation by assigning penalties based on the specific ranking of groups according to the next future observation per group. For the general event of selection and ranking, linear and quadratic loss functions are used to evaluate the ranking of groups assigned to different buckets. The effect of the use of different loss functions on the selection and ranking decisions is illustrated by examples.

October 6, 2025

Declaration

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

Acknowledgements

I am deeply grateful for the numerous blessings bestowed upon me by Allah, the Lord of the universe, especially throughout the process of composing this thesis.

My deepest gratitude and heartfelt appreciation go to my supervisors, Prof. Frank Coolen and Prof. Tahani Coolen-Maturi, for their unwavering support, expert guidance, and invaluable advice throughout this PhD journey. Their combined expertise, dedication, and encouragement have been instrumental in shaping this thesis. I am profoundly grateful for their mentorship, which has been a cornerstone of my academic growth and the successful completion of this work.

I extend my deepest appreciation to my father, Ibrahim, who has been a significant source of motivation and support, and to my mother, Norah, for her boundless love and prayers. I am also grateful to all my brothers and sisters for their encouragement and unwavering belief in me.

I am indebted to my cherished husband for his unwavering confidence and continuous motivation, and to my sons, Feras, Saud, and Jaser, whose constant love and cheerful presence brightened my days.

I owe a debt of gratitude to my colleagues and friends whose assistance and companionship were invaluable during moments of difficulty and challenge.

Finally, I am grateful for the financial and psychological support provided by King Abdulaziz University in Saudi Arabia, as well as the Saudi Arabian Cultural Bureau and all of their London staff. Also, I would like to thank Durham University for providing the facilities that have enabled me to study smoothly.

Contents

Chapter 1

Introduction

1.1 Motivation

When comparing multiple populations, it is often important to decide which performs better. For example, in medical trials where several treatments are tested, decision makers may wish to identify the most effective treatment, the safest option, or the one with the fewest side effects.

The objective of selection is to identify the best population, a subset of the best populations, or a subset that includes the best population. In addition, there are situations where the goal is not to select the best populations but to rank them based on a specific performance measure. For example, in medical trials where multiple treatments are tested, it may be important not only to identify the best performing treatments, but also to rank them based on their effectiveness. This helps decision makers determine which treatments should be prioritised for further investigation. The origins of statistical theory for solving selection and ranking problems trace back to the $1950s$, particularly in fields like agriculture and clinical studies, where testing the homogeneity of multiple populations was common [63]. Homogeneity implies that the means, variances, or other statistical properties of the populations are the same. When statistical tests indicated differences in their means, that is, so homogeneity of the means was rejected statistically, natural questions emerged: How can the uncertainty about being the best be quantified, and which one is the best? For example, when selecting the best drug among several options, the homogeneity

test can only indicate if the drugs are not equivalent, not which one is the best, where best, for example, could refer to the safest, most effective or having the fewest side effects. Homogeneity tests are insufficient to resolve the issue of selecting the best population. Therefore, selection and ranking procedures have been developed to solve these challenges for decision makers [31].

The term best is defined according to the problem at hand and may vary in different applications. For example, in some cases, the best result could indicate a smaller number of fatalities in a drug experiment. In contrast, in other scenarios, it might refer to drugs that induce a higher reduction in tumor size.

Selection and ranking procedures are appropriate for answering questions such as: Which of the different drugs produces the best response? Which subgroup of different drugs produces a better response than a placebo? Which two of multiple types of advertising media reach the highest proportion of potential buyers of a particular product? Which of the different learning techniques produces the best comprehension? How do a group of multiple candidates for a certain political office rank in popularity at a specific point in time? Which of the different types of ski bindings has the lowest accident rate? [31]

Historical applications of selection and ranking procedures include areas such as poultry science, drug studies, selection of advertising medium, investment of stocks, and selection of laboratory kit suppliers, demonstrating their importance in various fields for making informed and effective decisions [31].

In this thesis, the focus is on developing NPI for different selection and ranking problems. NPI is a statistical framework based on Hill's assumption [46, 47]. In the NPI approach, the uncertainty is quantified via lower and upper probabilities for events of interest. Due to its ease of implementation and minimal assumptions, NPI has been applied in various areas of statistics and reliability [10, 13, 19].

Coolen [9] introduced the NPI method to compare two independent groups and identify the best group that is most likely to provide the largest future observation. Then, Coolen and van der Laan [18] extended the NPI method to compare multiple independent groups by identifying both the best group and a subset that contains best multiple groups, but without considering the ranking within that subset. The

approach considered in this thesis is different from that of Coolen and van der Laan [18], as it focuses on ranking the best groups within the subset, providing more detailed insights into how they compare to each other.

The thesis employs the well-known concept of a loss function to measure the loss incurred from non-optimal selection and ranking. This loss function guides the selection and ranking of future observations within a ranked subset of the best groups, reflecting the quality of the decision-making process. In the context of selection, zero-one, linear, and quadratic loss functions are used to quantify the severity of making a non-optimal choice. Similarly, in the context of ranking, a loss function quantifies the severity of an incorrect ranking, with zero-one and general multi-level loss functions used for evaluation.

This chapter provides a review of the main ideas and concepts used in the thesis. Section 1.2 delves into the literature on selection and ranking from a classical statistics perspective. Section 1.3 offers an overview of statistical decision theory, while Section 1.4 briefly introduces the concept of NPI. Section 1.5 examines selection and ranking from an NPI perspective, and Section 1.6 outlines the thesis's structure.

1.2   Statistical selection and ranking procedures

Over the past seventy years, research in the statistical literature has been devoted to various theoretical aspects of many types of selection and ranking procedures. A pioneer in the field of selection and ranking problems is Bechhofer [3] who contributed to the early development of methods for comparing and selecting the best population among multiple populations based on ranking means of Normal populations with known variances. The focus then shifted to the area of subset selection procedures initiated by Gupta [37]. In various practical scenarios, it is often necessary to select multiple best populations rather than just one. Gupta [37] discussed the process of selecting a subset that includes the best population with specific parametric distributions. Gupta and others have contributed to the field by exploring subset selection approaches across various distributions, such as Normal distribution and Binomial distribution. For the Normal distribution, the best population is defined

as the one with the largest mean, while in the Binomial distribution, it is defined as the one with the highest probability of success [37, 39, 43].

An interesting aspect of the selection and ranking problem discussed in the literature is the determination of the subset size. In some cases, selecting only one population from a set may not be ideal. For example, if there are two similarly extreme options in the positive direction, it might be preferable to select both rather than just one. The size of the selected subset can be either random or fixed, but in this thesis, it is assumed to be fixed, so this distinction does not matter for our method. This flexibility in subset selection, where the subset size is random, can be advantageous for certain types of problem. The subset size is determined by the observed data using specific procedures and formulas available in the literature [37, 39, 43]. However, there are some situations where having a subset of a fixed size is preferred [24]. This ensures that the number of populations in the subset remains constant rather than varying. The decision to use a fixed or random subset size depends on the experimenter's objectives. If the experimenter prefers a fixed size of subset, promoting the benefits of a procedure with a random subset size does not offer an advantage [31].

Selecting a subset of random size from Normal distributions that contains the best population with respect to means and a common known variance was presented by Gupta [37, 38]. Similarly, for the case of a common unknown variance proposed by Gupta [37], Gupta and Sobel [42]. A subset selection problem for Binomial distributions is addressed in Gupta and Sobel [43]. Many tables applied for subset selection problems associated with Binomial distribution are presented by Gupta et al. [39].

However, selecting a subset of fixed size of Normal distributions with respect to means with a common known variance is proposed by Desu and Sobel [24]. Desu and Sobel's research is connected to earlier work by Mahamunulu [59], where the interest is centred on populations that have large values of the ranking parameter. Typically, the populations with the largest values of the ranking parameters are referred to as the best populations. Mahamunulu [59] considered a problem of selecting a fixed size of subset, from a given set of all populations, which contains a subset of the

best populations.

For subset selection problems involving Normal distributions with unknown and possibly unequal variances, Dudewicz and Dalal [27] proposed a two stage procedure to select the population with the largest mean. In the first stage, a small sample of observations is taken from each population, and the mean of these observations is calculated for each population. In the second stage, additional observations are collected from each population. Subsequently, a new weighted mean is calculated using all the observations, ensuring that the weights assigned to the initial and additional observations are consistent. Finally, the population with the largest weighted mean from the second stage is selected as the best. However, this method is not considered in this thesis.

Furthermore, Bechhofer et al. [4] have extensively investigated sequential procedures related to the selection of a subset of best multiple populations, addressing both ranked and unranked scenarios. In another paper, Sobel [70] explored the problem of selecting a subset of multiple populations and confirming that at least one of them belongs to the best multiple populations. Moreover, Sobel [69] considered the problem of selecting the best multiple populations out of the k populations, each characterized by a cumulative distribution function F_i for $i = 1, \ldots, k$. The selection of these best multiple populations is based on a specific criterion, which is the population with the largest q-quantile, where q is a given number such that $0 < q < 1$.

Sobel [69] developed nonparametric procedures similar to those discussed by Rizvi and Sobel [64], based on n independent observations from each of the k populations. Desu and Sobel [25] addressed the nonparametric version of selecting a fixed size of subset that includes the population with the largest q-quantile. More recently, Kumar and Grover [54] proposed a nonparametric subset selection procedure for location parameters that does not rely on specific distributional assumptions. Their method, based on U-statistics, preserves important properties such as monotonicity and subset-inclusion, and performs well under various simulated settings. In a related study, Kumar [53] introduced a general class of subset selection procedures applicable to both location and scale parameters.

Designing an experiment in statistics requires specifying the characteristics in advance. This is not directly relevant to this thesis, as the method used here, NPI, works differently and will be discussed later in Section 1.4. Some methods in the literature, such as those by Tong [71] and Bechhofer [3], explicitly focus on experimental design.

Tong [71] addressed the scenario of selecting populations that are better than a control or standard. Tong [71] described a treatment control experiment, which is a common design used in various research and experimentation scenarios. This design applies to the situation in which each of the k treatments is compared to a single control or standard treatment. The number of treatments in this type of experiment is k treatment and a single control or standard treatment.

In selection and ranking problems, the objectives can extend to complete ranking of all k populations based on their parameter values. For example, in the evaluation of k competing consumer or product goods, there might be a need to rank these goods concerning their means. The problem of completely ranking three populations based on their corresponding variances was presented in Bechhofer [3]. A contribution to completely ranking k populations according to their mean values was made by Carroll and Gupta [8]. The Gupta and Carroll method is suitable for larger sample sizes while Freeman et al. [30] are suitable for smaller sample sizes. In addition, the complete ranking of the variances of the Normal distribution was introduced by Schafer [66]. Finally, Bayesian methods were applied to the complete ranking problem by Govindarajulu and Harvey [35]. However, the main interest in this thesis lies in ranking a subset of the best groups.

The next two sections briefly review the indifference zone approach for ranking means of Normal populations and the subset selection approach.

1.2.1 Indifference zone approach

The indifference zone approach was introduced by Bechhofer [3]. The main objective is to select the best Normal distribution out of k independent Normal distributions. The best population here means the population with the largest mean. Let μ_i and σ_i^2 denote the mean and variance of the i'th population and the Normal distributions are

denoted as $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, for $i = 1, 2, ..., k$. In this case, the variance was assumed to be common and known, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$. The mean values $\mu_1, \mu_2, \ldots, \mu_k$ are arranged in ascending order, then, the ordered means can be written as

$$\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]} \tag{1.1}$$

It is assumed that it is not known which population is associated with $\mu_{[i]}$, where $i = 1, \ldots, k - 1$. It is also assumed that the best and second best populations here are the populations with the largest and second largest means respectively. The inference based on sample means, and the k ranked sample means are obtained as follows

$$\bar{x}_{[1]} \leq \bar{x}_{[2]} \leq \ldots \leq \bar{x}_{[k]} \tag{1.2}$$

Bechhofer [3] proposed designing an experiment to identify the best population in such a way that under specific conditions the probability of making a correct selection and ranking, $P(CSR)$, is at least a preassigned value, $P(CSR) \geq P^*$. The number of observations n required depends on the probability of making a correct selection and ranking. Note that the higher the probability of correct selection and ranking P^*, the greater the required sample size n.

In order to apply Bechhofer's method [3], the following steps are followed: Take n_i observations from the i'th population, the sample size of all populations is assumed to be common $n_i = n$. Next, compute the sample means $\bar{x}_{[1]}, \ldots, \bar{x}_{[k]}$. Make the ranking of the sample means as in Equation (1.2). Finally, the populations associated with $\bar{x}_{[k]}$ and $\bar{x}_{[k-1]}$ are identified as the largest and second largest populations respectively.

The specific condition mentioned in the previous paragraph, $P(CSR)$, states the distance between the largest and the second largest mean is at least a specified value δ^* as expressed by

$$\mu_{[k]} - \mu_{[k-1]} \geq \delta^* \tag{1.3}$$

this condition here is referred to as the least favourable configuration of the probability of population means.

For a specified P^*, the minimum required sample size n for each population is the smallest integer that is greater than or equal to the value of n calculated as

$$n = \left(\frac{d\sigma}{\delta^*}\right)^2 \tag{1.4}$$

if the right hand side of Equation (1.4) is not an integer, then round up to the next larger integer to determine n. Here, d is the multivariate Normal quantile constant needed to achieve the desired confidence P^* in Bechhofer's method.

When the d value in Equation (1.4) for a given P^* is unknown, linear interpolation on natural logarithmic values can be applied to estimate it. Three steps are followed to find it: First, the two closest values below and above the unknown d and its corresponding P^* values are identified. The natural logarithms of the corresponding values of d and $1 - P^*$ are then taken, since this transformation makes the relationship nearly linear. Linear interpolation is performed between these natural logarithmic values to determine the missing d. Finally, an antilogarithm of this result is taken to obtain d in its original scale.

1.2.2 Subset selection approach

There are many methods presented in the literature to select a subset of the best populations, but their objectives are not directly relevant to the method presented in this thesis. This is because the approach in Gupta [38] focuses on selecting the best populations without considering their ranking, whereas the method presented in this thesis is concerned with ranking the best populations. Gupta [38] presented a subset selection procedure aimed at selecting a subset of Normal populations with a common known variance that includes the best Normal population.

In the subset selection procedure, Gupta [38] designed an experiment to determine the common sample size n required per population, for specified numbers δ^*, P^* and k populations from Normal distributions which is overviewed in Section 1.2.1. The minimum required sample size n for each population is the smallest integer that is greater than or equal to the value of n calculated as

$$n = \left(\frac{\lambda\sigma}{\delta^*}\right)^2 \tag{1.5}$$

Here, λ is the critical cutoff constant to ensure that the subset contains the best population with probability at least P^*. The value of λ was obtained by the same method that are suggested for interpolation to find d in Equation (1.4). However, better results can be obtained using a different transformation as presented in [31].

To perform Gupta procedure, the k ordered sample means are obtained as follows

$$\bar{x}_{[1]} \leq \bar{x}_{[2]} \leq \ldots \leq \bar{x}_{[k]} \tag{1.6}$$

Then, the selection rule for any $i = 1, 2, \ldots, k$, the ith population is included in the selected subset if and only if its corresponding sample mean, $\bar{x}_{[i]}$, is at least as large as a certain quantity, which is

$$\bar{x}_{[i]} \geq \bar{x}_{[k]} - \frac{\sigma}{\sqrt{n}}\lambda \tag{1.7}$$

Subsequently, in Section 2.6, a comparative analysis is conducted between Bechhofer, Gupta and the method presented in Chapter 2.

1.3 Statistical decision theory

This section offers an overview of statistical decision theory and explores different types of loss function, as discussed in existing literature. Decision theory focuses on the problem of making decisions, while statistical decision theory specifically addresses decision making in the presence of statistical knowledge, which is the information derived from data that helps to shed light on uncertainties in decision problems.

Wald [73] introduced statistical decision theory to guide decision making in the presence of uncertainty. It is assumed that these uncertainties can be considered unknown numerical quantities, represented by θ, which can be any parameter of interest, such as a mean or variance. The aim is to estimate or make decisions about θ. This theory shifts the focus from statistical inference, drawing conclusions from data, to making decisions under uncertainty. Traditionally, classical statistics focuses mainly on using sample information for inferences about a parameter θ, often without considering how these inferences will be used, which is based on incomplete information. Incomplete information in this context means that inferences about

the parameter θ are derived from just a sample of the population, rather than the entire population. Because only a part of the total data is available, the information used to make inferences is incomplete. In contrast, in decision theory the goal is to combine sample information with other relevant aspects of the problem to make optimal decisions. Subsequently, statistical decision theory is centred around the problem of statistical decision rather than inference. It involves deciding on a sensible course of decision based on incomplete information.

An important aspect is understanding the possible consequences of decisions. This can be quantified by determining the loss incurred for each possible decision. The incorporation of loss functions into statistical analysis was extensively studied by Wald [73] . When making a decision, the parameter θ is unknown or uncertain at the time of decision making. To handle this uncertainty, one common use of decision theory is to evaluate decisions based on their expected loss (risk) and then select the best decision that minimises this expected loss.

The risk of making a wrong decision can often be reduced by taking enough observations. However, experiments usually incur costs. The risk function introduced by Wald [73] depends on both the cost of experimentation and the cost of making a wrong decision, with estimates of these costs provided by the experimenter. Calculating the expected total loss requires knowledge of the prior probabilities of the possible values of unknown quantities θ that are related to different decisions and the populations studied in the experiment. This type of information is hard to obtain.

A valuable source of non-sample information is prior information. Information about θ is obtained from sources outside of the statistical investigation, such as past experiences with similar situations, as refined by Savage [65].

As an example of making a decision, imagine a drug company deciding whether to launch a new pain reliever. Key factors in this decision include the drug's effectiveness for patients and its market potential, which are generally unknown but can be studied through experiments. This decision making process is part of decision theory, with the aim of determining whether to market the drug and quantifying the loss incurred from incorrect decisions. The losses involved in deciding whether to

market the drug include the risk of underestimating or overestimating its effectiveness. Underestimating the effectiveness of the drug could make the product appear less effective than it really is, while overestimating it could result in penalties for misleading advertising [5].

In the drug example, prior information about effectiveness and market share can be derived from other pain relievers. This prior information can enhance the decision making process by providing an informed basis for estimating the drug's effectiveness for patients and its market potential. However, this approach is not utilized in this thesis.

Commonly utilised loss functions include zero-one, linear, and quadratic losses, which are essential concepts in decision-making. Additionally, entropy serves as a valuable inference loss, offering quantitative assessments of uncertainty or errors when drawing conclusions from a statistical model. Good [34] found loss function called "Quasi-utilities", which provides a more flexible method for decision making under uncertainty than other methods, allowing decision makers to indicate their preferences when complete information is lacking. The different types of loss functions used in the selection and ranking problems are discussed in Chapters 3 and 4.

Shafer [67] explored how decision makers can combine various decision theory problems rather than considering them independently. However, the need to combine problems can present challenges for decision makers when constructing loss functions. Lehmann et al. [57] and Ferguson [28] illustrated that frequentist and Bayesian inference measures can also be formally interpreted as risk or Bayesian expected loss. However, these approaches are not utilized in this thesis. Rather than addressing general loss functions, the focus here is on those specifically related to subset selection, as discussed in Chapter 3.

1.4 Nonparametric Predictive Inference (NPI)

Imprecise probability, introduced by Boole [7], measures the uncertainties associated with events using intervals rather than single values, as seen in classical probability

[11]. In recent years, several alternative methods for quantifying uncertainty have emerged, such as Walley's imprecise probability theory [74] and Weichselberger's interval probability theory [75], which use lower and upper probabilities instead of fixed probabilities. These probabilities form the basis of a statistical framework called Nonparametric Predictive Inference (NPI).

Nonparametric Predictive Inference (NPI) [1, 12] is a statistical framework based on Hill's assumption $A_{(n)}$ [46, 47] which provides direct probabilities for future observation given n observations of related random quantities. Inferences derived from this assumption are predictive and nonparametric, and seem suitable if there is hardly any knowledge about the random quantities of interest, or it may be more realistic in situations where one does not want to utilize such information, e.g. when studying the effects of additional assumptions in other statistical methods.

To introduce the assumption $A_{(n)}$, suppose that $X_1, X_2, \ldots, X_{n+1}$ are continuous and exchangeable random quantities. The ordered observations of $X_1, X_2, \ldots, X_n$ are $x_1 < x_2 < \ldots < x_n$. Let $x_0 = -\infty$ and $x_{n+1} = \infty$. Note that x_{n+1} does not denote an observed value for X_{n+1}. These n observations divide the real line into $n+1$ intervals $I_j = (x_{j-1}, x_j)$, where $j = 1, \ldots, n+1$. The assumption $A_{(n)}$ is that the probability for the future observation X_{n+1} to fall in the open interval I_j is equal for all I_j, which is

$$P(X_{n+1} \in I_j) = \frac{1}{n+1} \text{ for each } j = 1, \ldots, n+1 \qquad (1.8)$$

Throughout this thesis, it is assumed that there are no tied observations. However, to deal with tied observations, one can either assign the probability mass to the closed intervals $[x_{j-1}, x_j]$ instead of the open interval, or tied observations can be dealt with by assuming that such observations differ by a very small amount [48].

$A_{(n)}$ does not assume anything else, and can be considered to be a post-data assumption related to exchangeability [22]. The assumption $A_{(n)}$ alone does not yield precise probabilities for various events of interest [47]. However, it provides bounds for probabilities for all events of interest involving X_{n+1}, by what is essentially an application of De Finetti's fundamental theorem of probability [22]. These bounds are called lower and upper probabilities in imprecise probability theory [74] and interval probability [76]. The NPI lower probability for an event A is denoted by

$\underline{P}(A)$, the NPI upper probability for the same event is denoted by $\overline{P}(A)$. The NPI lower probability can be interpreted as a maximum lower bound for the precise probability for A, and the NPI upper probability can be interpreted as a minimum upper bound for the precise probability for A [12]. Therefore, the uncertainty is quantified by the NPI lower and upper probabilities for the event of interest based on the assumption $A_{(n)}$. Augustin and Coolen [1] introduced the NPI lower and upper probabilities for the event $X_{n+1} \in B$ given the past n data observations where $B \subset \mathbb{R}$, as follows

$$\underline{P}(X_{n+1} \in B) = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbb{1}\{I_j \subseteq B\} \tag{1.9}$$

$$\overline{P}(X_{n+1} \in B) = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbb{1}\{I_j \cap B \neq \emptyset\} \tag{1.10}$$

where $\mathbb{1}\{A\}$ is an indicator function which is equal to 1 if event A occurs and 0 else. The NPI lower probability is defined by summing only the probability masses assigned to intervals I_j that are necessarily within B, while the NPI upper probability is defined by summing all the probability masses that can be in B, which is the case for the probability masses per interval I_j if the intersection of I_j and B is not empty.

Augustin and Coolen [1] provided properties for the NPI lower and upper probabilities. First, the conjugacy property $\overline{P}(A) = 1 - \underline{P}(A^c)$ holds, where A^c is the complement of an event A. Secondly, $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$, when $\underline{P}(A) = \overline{P}(A)$, this represents a precise classical probability. The case in which $P(A) = 0$ and $P(A) = 1$ represents a complete lack of information about the event A. Moreover, $\triangle(A) = \overline{P}(A) - \underline{P}(A)$ is called imprecision [11].

1.5   NPI for selection and ranking

This section outlines the NPI procedure as developed for selection and ranking problems in the literature. As these methods form the basis of this work, both are briefly reviewed, starting with selection and then ranking applications.

NPI for pairwise comparisons has been developed for real-valued observations

[9]. Coolen [9] proposed a comparison method within the NPI framework for two independent groups of real-valued data, involving a single future observation for each group, and then selecting the group based on the NPI lower and upper probabilities for the event that a group will have the largest future observation. Suppose that $X_1, \ldots, X_{n_x}, X_{n_x+1}$ and $Y_1, \ldots, Y_{n_y}, Y_{n_y+1}$ are real-valued, absolutely continuous, and exchangeable random quantities from X and Y, respectively. Let their ordered observed values be $x_1 < x_2 < \cdots < x_{n_x}$ and $y_1 < y_2 < \cdots < y_{n_y}$, with $x_0 = y_0 = -\infty$ and $x_{n_x+1} = y_{n_y+1} = \infty$ defined for ease of notation.

The derivation of the NPI lower and upper probabilities for the event $X^f > Y^f$ involves considering the extreme positions of the probability mass for each random quantity, X^f and Y^f, within their respective intervals, which minimise the NPI lower probability and maximise the NPI upper probability. The NPI lower and upper probabilities for the event $X^f > Y^f$ are

$$\underline{P}(X^f > Y^f) = \frac{1}{(n_x+1)(n_y+1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \mathbb{1}\{x_{i-1} > y_j\} \tag{1.11}$$

$$\overline{P}(X^f > Y^f) = \frac{1}{(n_x+1)(n_y+1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \mathbb{1}\{x_i > y_{j-1}\} \tag{1.12}$$

The method is applied to derive predictive inferences to compare two future observations, X^f and Y^f. The method for deriving the NPI lower and upper probabilities serves as a starting point for deriving the NPI lower and upper probabilities relevant to this thesis.

Subsequently, the NPI method has been extended to multiple comparisons for $k \geq 2$ independent groups and developed for different types of data, which have been applied to various inferences. These include group selection when dealing with real-valued observations for three different events of interest based on a single future observation for each group, which are the selection of the best group, the subset of the best groups, and the subset that includes the best group [18].

Suppose there are $k \geq 2$ groups and $n_j + 1$ random quantities from group j denoted by X_{j,i_j} where $j = 1, 2, \ldots, k$ and $i_j = 1, 2, \ldots, n_j + 1$ and their ordered observation values are $x_{j,1} < x_{j,2} < \ldots < x_{j,n_j}$ with $x_{j,0} = -\infty$ and $x_{j,n_j+1} =$

∞ defined for ease of notation. Then, an inference is considered based on Hill's assumption $A_{(n_j)}$ for each group.

First, select the best group with a specific future observation X_ℓ^f which is the maximum for all future observations X_j^f, $j = 1, 2, \ldots, k$, i.e., $X_\ell^f \geq \max_{j \neq \ell} X_j^f$. The NPI lower and upper probabilities for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ are

$$\underline{P}(X_\ell^f \geq \max_{j \neq \ell} X_j^f) = \frac{1}{\prod_{j=1}^k (n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^k \sum_{i_j=1}^{n_j+1} \mathbb{1}\{x_{\ell, i_\ell - 1} \geq \max_{\ell \neq j}(x_{j, i_j})\} \qquad (1.13)$$

$$\overline{P}(X_\ell^f \geq \max_{j \neq \ell} X_j^f) = \frac{1}{\prod_{j=1}^k (n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^k \sum_{i_j=1}^{n_j+1} \mathbb{1}\{x_{\ell, i_\ell} \geq \max_{\ell \neq j}(x_{j, i_j - 1})\} \qquad (1.14)$$

The second event is to select a subset of the best groups. Let $S = \{\ell_1, \ell_2, \ldots, \ell_w\} \subset \{1, 2, \ldots, k\}$ be the subset of w groups, and let $NS = \{1, 2, \ldots, k\} \backslash S$, so NS is the set of $k - w$ non-selected groups. All future observations belonging to the subset S exceed all future observations that belong to the non-selected groups NS, i.e. $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$. The NPI lower and upper probabilities for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ are

$$\underline{P}(\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f) = \frac{1}{\prod_{j=1}^k (n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^k \sum_{i_j=1}^{n_j+1} \mathbb{1}\{\min_{\ell \in S}(x_{\ell, i_\ell - 1}) \geq \max_{j \in NS}(x_{j, i_j})\}$$
$$(1.15)$$

$$\overline{P}(\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f) = \frac{1}{\prod_{j=1}^k (n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^k \sum_{i_j=1}^{n_j+1} \mathbb{1}\{\min_{\ell \in S}(x_{\ell, i_\ell}) \geq \max_{j \in NS}(x_{j, i_j - 1})\}$$
$$(1.16)$$

where $\sum_{i_\ell=1}^{n_\ell+1}$ is used for w sums $\sum_{i_{\ell_1}=1}^{n_\ell+1} \cdots \sum_{i_{\ell_w}=1}^{n_\ell+1}$.

Finally, the third event is to select the subset S that contains the best group. The future observation of at least one of the selected groups in S is greater than the future observation of each group in NS, that is, $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$.

$$\underline{P}(\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f) = \frac{1}{\prod_{j=1}^k (n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^k \sum_{i_j=1}^{n_j+1} \mathbb{1}\{\max_{\ell \in S}(x_{\ell, i_\ell - 1}) \geq \max_{j \in NS}(x_{j, i_j})\}$$
$$(1.17)$$

$$\overline{P}(\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f) = \frac{1}{\prod_{j=1}^{k}(n_j + 1)} \sum_{i_\ell=1}^{n_\ell+1} \prod_{j=1}^{k} \sum_{i_j=1}^{n_j+1} \mathbb{1}\{\max_{\ell \in S}(x_{\ell,i_\ell}) \geq \max_{j \in NS}(x_{j,i_j-1})\}$$

(1.18)

Then, NPI has also been developed for other types of data and has been applied to various inference problems. Maturi et al. [61] introduced the NPI method for comparing future observations from two groups of real-valued data, where the tails of the data may have been terminated. The selection of the optimal group in the Bernoulli trials was introduced by Coolen and Coolen-Schrijner [16]. Lastly, Coolen et al. [15] introduced the Nonparametric Predictive Inference (NPI) method for pairwise comparison that specifically focuses on future order statistics. However, these are not relevant to this thesis.

In addition, selecting the optimal group in the context of Bernoulli data was introduced by Coolen and Coolen-Schrijner [17], while Baker [2] explored selecting the best category within the multinomial data. Maturi [60] also introduced the application of multiple comparisons for right-censored observations. Furthermore, Coolen et al. [15] developed the NPI method for multiple comparisons, specifically focusing on future order statistics. In all these studies, multiple comparisons were applied for $k \geq 2$ populations, focusing on the three different events of interest described above.

In many situations, one may want not just to compare between groups, but rather to rank the groups from the best to the worst. For example, this occurs when evaluating the effectiveness of various treatments and determining if they follow a specific order in which they are effective. Ranking multiple groups based on the NPI method is introduced in the literature. Coolen-Maturi [20] presented NPI for the ranking of real-valued future observations for three and more than three groups, resulting in a more complex optimisation process.

To derive the NPI lower and upper probabilities for the event $X_{n_x+1} < Y_{n_y+1} < Z_{n_z+1}$, the probability masses for X_{n_x+1} and Z_{n_z+1} are assigned to the extreme positions within their respective intervals, which minimise the NPI lower probability and maximise the NPI upper probability. For the group in the middle, Y_{n_y+1}, an optimization is performed by assigning the probability mass for Y_{n_y+1} within each

interval (y_{j-1}, y_j) in a way that minimises the NPI lower probability and maximise the NPI upper probability.

To derive the NPI lower and upper probabilities for the event $X_{n_x+1} < W_{n_w+1} < Y_{n_y+1} < Z_{n_z+1}$, similarly to the method used for the event $X_{n_x+1} < Y_{n_y+1} < Z_{n_z+1}$, the probability masses for X_{n_x+1} and Z_{n_z+1} are again assigned at the extreme ends of their intervals. Then, the complexity arises from the need to simultaneously assign probability masses for the remaining groups, W_{n_w+1} and Y_{n_y+1}. The approximate NPI lower and upper probabilities calculation requires simultaneously considering all possible ways to assign probability masses for W_{n_w+1} and Y_{n_y+1}. This simultaneous consideration is necessary to minimise the NPI lower probability and maximise the NPI upper probability, especially with large and overlapping datasets, so that bounds have been introduced to avoid this computational complexity.

1.6 Outline of the thesis

This thesis presents the NPI method for a range of selection and ranking problems. It also quantifies the loss from making suboptimal decisions using various loss functions, which not only measure performance but also guide the selection and ranking of future observations within a chosen subset of the best groups. The structure of the thesis is as follows: Chapter 2 introduces the ranking of future observations of a subset of best groups, considering various scenarios. Within these subsets, the partial ranking of a subset of the best groups is investigated. Finally, a comparative analysis is conducted between the proposed method and the methods in the literature to investigate their performance.

In Chapter 3, the loss incurred in certain NPI selection problems is measured using three types of loss functions: zero-one, linear, and quadratic loss functions. Then, the NPI lower and upper expected zero-one, linear and quadratic losses quantify the uncertainty associated with the loss of making non-optimal selections and guide the selection of the best groups, reflecting the quality of the selection.

Chapter 4 develops NPI decision theory for ranking a subset of the best groups, using both the zero-one loss function and a general multi-level loss function. The

multi-level loss function assigns penalties based on the severity of ranking errors, allowing each ordering mistake to be weighted according to the investigator's priorities.

Chapter 5 extends the methodology to the general case of ranking buckets, where independent groups are assigned to buckets such that each group belongs to exactly one bucket and each bucket contains at least one group. Linear and quadratic loss functions are used to quantify the loss from incorrect bucket rankings and to guide the ranking process, reflecting the quality of the resulting decisions.

Finally, some remarks and conclusions are provided in Chapter 6. Calculations were performed using R, and the R code is available from the author upon request.

Chapter 2

NPI for selecting a ranked subset

2.1   Introduction

In many scenarios, the key objective is to select a subset of groups and rank the best groups within that subset. For example, in medical trials where multiple treatments are tested, it is crucial not only to identify the best treatments but also to rank them according to their effectiveness. This allows decision-makers to prioritize treatments for further investigation. Selecting subsets of the best groups and ranking the groups is an important problem, particularly in fields such as medicine, industry and economics. Although Coolen and van der Laan [18] introduced a method for selecting a subset containing multiple best groups, their approach did not account for ranking of groups within the subset. The method developed in this thesis differs by focusing on ranking the best groups within the subset.

This chapter presents a methodology for selecting a ranked subset of best groups, based on one future observation from each group. First, the general scenario is considered, in which multiple groups are included in the subset. The necessary notations and probability formulas for this case are defined and introduced. Then, two special cases are examined, focusing on subsets containing two and three groups. These cases are crucial as they demonstrate how the NPI lower and upper probabilities are derived differently depending on the number of groups in the subset. Finally, one may wish to rank future observations within a partially ranked subset, where the ranked groups are nested within a larger selected subset of the best groups.

This chapter is organised as follows. Section 2.2 introduces the methodology for selecting a ranked subset of the best groups, based on one future observation from each group. Sections 2.3 and 2.4 focus on special cases where the subset contains two and three groups, respectively. Section 2.5 provides a brief illustration of partial ranking within subset groups. A comparison with alternative methods from the literature is presented in Section 2.6. Finally, Section 2.7 presents concluding remarks.

2.2 Selecting a ranked subset of the best groups

This section introduces the NPI-based method for selecting a ranked subset of the best groups, based on one future observation from each group. This section is important because it presents the general event of selecting a ranked subset of the best groups, providing a foundation for the next two sections, which derive special cases when the ranked subset has two and three groups. First, the notation is introduced, followed by the NPI lower and upper probabilities for this event. In some cases, computing these probabilities can be demanding, especially when data overlap occurs, so lower and upper bounds on these NPI probabilities are introduced.

For simplicity of notation, the next future observation is referred to as X^f instead of X_{n_x+1}. Suppose that there are k independent groups, where 'independent' means that information about the random quantities for one group does not provide any information about the random quantities for another group. In addition, consider a subset of all groups, consisting of w groups, with $1 \leq w \leq k - 1$. Let $S = \{\ell_1, \ell_2, \ldots, \ell_w\} \subset \{1, 2, \ldots, k\}$ be the subset of w groups, and let $NS = \{1, 2, \ldots, k\} \backslash S$, so NS is the set of $k - w$ non-selected groups.

Suppose that the event of interest is $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$, meaning that all non-selected groups have smaller future observations than the selected groups, where the selected groups are ranked. Therefore, the objective is to select the ranked subset S of the best w groups.

2.2.1 Lower and upper probabilities

Coolen and van der Laan [18] introduced NPI lower and upper probabilities for the event $\max\limits_{j\in NS} X_j^f < \min\limits_{\ell\in S} X_\ell^f$, where the future observation of each group in S is greater than all future observations belonging to NS. An overview of how to derive the NPI lower and upper probabilities for the event $\max\limits_{j\in NS} X_j^f < \min\limits_{\ell\in S} X_\ell^f$ was provided in Section 1.5. The methodology for deriving the NPI lower and upper probabilities is now extended to the event $\max\limits_{j_z\in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$, drawing similarities to the earlier case of $\max\limits_{j_z\in NS} X_{j_z}^f$ and $X_{\ell_1}^f$, as discussed in Section 1.5. In this thesis, we assume that there are no tied observations. To introduce the NPI lower and upper probabilities for the event $\max\limits_{j_z\in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$, Hill's assumption $A_{(n)}$ is applied per group. Let the ordered observations of group ℓ_i, $\ell_i \in S$, be denoted by $x_{\ell_i,1} < x_{\ell_i,2} < \ldots < x_{\ell_i,n_{\ell_i}}$ for ease of notation $x_{\ell_i,0} = -\infty$ and $x_{\ell_i,n_{\ell_i}+1} = \infty$. Let n_{ℓ_i} be the number of observations from group ℓ_i, these divide the real line into $n_{\ell_i} + 1$ intervals denoted by $\{I_1, \ldots, I_{n_{\ell_i}+1}\}$. $A_{(n_{\ell_i})}$ partially specifies a probability distribution for the next future observation of group ℓ_i by $P(X_{\ell_i}^f \in (x_{\ell_i,u_{\ell_i}-1}, x_{\ell_i,u_{\ell_i}})) = \dfrac{1}{n_{\ell_i}+1}$, where $u_{\ell_i} = 1, 2, \ldots, n_{\ell_i} + 1$.

The ordered observations from group j_z, $j_z \in NS$, be denoted by $x_{j_z,1} < x_{j_z,2} < \ldots < x_{j_z,n_{j_z}}$ with ease of notation $x_{j_z,0} = -\infty$ and $x_{j_z,n_{j_z}+1} = \infty$. Let n_{j_z} be the number of observations related to non-selected groups j_z and these divide the real line into $n_{j_z} + 1$ intervals, denoted by $\{I_1, \ldots, I_{n_{j_z}+1}\}$. The $\max\limits_{j_z\in NS} x_{j_z}$ represents the maximum observation that belongs to NS. $A_{(n_{j_z})}$ partially specifies a probability distribution of the next future observation per group $X_{j_z}^f$ by $P(X_{j_z}^f \in (x_{j_z,v_{j_z}-1}, x_{j_z,v_{j_z}})) = \dfrac{1}{n_{j_z}+1}$, where $v_{j_z} = 1, 2, \ldots, n_{j_z} + 1$.

The NPI lower probability for the event $\max\limits_{j_z\in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$ can be derived as follows: the probability mass for group ℓ_1 assigns to the left endpoint of the intervals $(x_{\ell_1,u_{\ell_1}-1}, x_{\ell_1,u_{\ell_1}})$, while the probability masses for all $j_z \in NS$ groups assign to the right endpoint of the intervals $(x_{j_z,v_{j_z}-1}, x_{j_z,v_{j_z}})$, where $z = 1, \ldots, k-w$. In contrast, for the NPI upper probability, the probability mass for group ℓ_1 is assigned to the right endpoint per interval, while the probability masses for all groups $j_z \in NS$ are assigned to the left endpoint per interval. Hence, the NPI lower

probability for the event $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < X^f_{\ell_1}$ is

$$
\underline{P}\left(\max_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < X^f_{\ell_1}\right) =
$$

$$
\frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1} P\left(\max_{j_z \in NS} x_{j_z,v_{j_z}} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < x_{\ell_1,u_{\ell_1}-1}\Big| D\right)
$$

(2.1)

where $D = \{X^f_{\ell_i} \in (x_{\ell_i,u_{\ell_i}-1}, x_{\ell_i,u_{\ell_i}}), i = 2,\ldots,w\}$ and the notation $\sum\limits_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}$ is used for

$k - w$ sums $\sum\limits_{v_{j_1}=1}^{n_{j_1}+1} \sum\limits_{v_{j_2}=1}^{n_{j_2}+1} \ldots \sum\limits_{v_{j_{k-w}}=1}^{n_{j_{k-w}}+1}$.

The NPI upper probability for the event $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < X^f_{\ell_1}$ is

$$
\overline{P}\left(\max_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < X^f_{\ell_1}\right) =
$$

$$
\frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1} P\left(\max_{j_z \in NS} x_{j_z,v_{j_z}-1} < X^f_{\ell_w} < \ldots < X^f_{\ell_2} < x_{\ell_1,u_{\ell_1}}\Big| D\right)
$$

(2.2)

We have shown above how the probability masses are assigned to the groups ℓ_1 and $j_z \in NS$. However, for the remaining middle groups $\ell_2,\ldots,\ell_w$, it is not straightforward how to allocate their probability masses within each interval to derive the NPI lower and upper probabilities. Coolen-Maturi [20] presented a proposal that involves performing optimisation for one group at a time. For the remaining groups, their corresponding probability masses are allocated at either the left or right endpoint of each interval $(x_{\ell_i,u_{\ell_i}-1}, x_{\ell_i,u_{\ell_i}})$, where $i = 2,\ldots,w$. Following this optimisation, which will be briefly discussed in Sections 2.3 and 2.4, the NPI lower and upper probabilities are derived. Examples illustrating the optimisation process for selecting a ranked subset of the two best groups, as well as for three or more best groups, will also be provided.

2.2.2 Bounds for the NPI lower and upper probabilities

Since calculating the NPI lower and upper probabilities involves optimisation which requires going through different sub-intervals to assign the probability mass of each

group, the calculation process becomes cumbersome. Therefore, considering lower and upper bounds for the NPI lower and upper probabilities allows for easier calculation. First, consider the lower bound for the NPI lower probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$, which is denoted by $\underline{P}^L$. It can be derived by requiring the total separation for the intervals $(x_{\ell_i,u_{\ell_i}-1}, x_{\ell_i,u_{\ell_i}})$, where $i = 1, 2, \ldots, w$ and $u_{\ell_i} = 1, 2, \ldots, n_{\ell_i} + 1$, and $(x_{j_z,v_{j_z}-1}, x_{j_z,v_{j_z}})$, where $z = 1, 2, \ldots, k - w$ and $v_{j_z} = 1, 2, \ldots, n_{j_z} + 1$. In other words, the lower bound for the NPI lower probability counts all the intervals that have to be totally separated. $I(A)$ is an indicator function that is equal to one if the event A occurs and zero otherwise. The lower bound for the NPI lower probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$

is

$$\underline{P}^L \left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f \right) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}$$

$$I\left\{ \left(\max_{j_z \in NS} x_{j_z,v_{j_z}} < x_{\ell_w,u_{\ell_w}-1} \right) \wedge \bigcap_{i=1}^{w-1} \left\{ x_{\ell_{w-i+1}, u_{\ell_{w-i+1}}} < x_{\ell_{w-i}, u_{\ell_{w-i}}-1} \right) \right\} \quad (2.3)$$

where $\sum\limits_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1}$ is used for w sum $\sum\limits_{u_{\ell_1}=1}^{n_{\ell_1}+1} \sum\limits_{u_{\ell_2}=1}^{n_{\ell_2}+1} \cdots \sum\limits_{u_{\ell_w}=1}^{n_{\ell_w}+1}$. The lower bound for the NPI lower probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$ is derived as follows

$$P\left(\max_{j_z \in NS} X_{j_z}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f \right) = \sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} P\left(\max_{j_z \in NS} X_{j_z}^f < \ldots < X_{\ell_1}^f \middle| X_{\ell_1}^f \in (x_{\ell_1,u_{\ell_1}-1}, x_{\ell_1,u_{\ell_1}}) \right)$$

$$\times P\left(X_{\ell_1}^f \in (x_{\ell_1,u_{\ell_1}-1}, x_{\ell_1,u_{\ell_1}}) \right)$$

$$\geq \frac{1}{(n_{\ell_1}+1)} \sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} P\left(\max_{j_z \in NS} X_{j_z}^f < \ldots < X_{\ell_2}^f < x_{\ell_1,u_{\ell_1}-1} \right)$$

$$= \frac{1}{(n_{\ell_1}+1)} \sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} P\left(\bigcap_{j_z \in NS} \left\{ X_{j_z}^f < \ldots < X_{\ell_2}^f < x_{\ell_1,u_{\ell_1}-1} \right\} \middle| X_{j_z}^f \in (x_{j_z,v_{j_z}-1}, x_{j_z,v_{j_z}}) \right)$$

$$\times P\left(X_{j_z}^f \in (x_{j_z,v_{j_z}-1}, x_{j_z,v_{j_z}}) \right)$$

$$\geq \frac{1}{(n_{\ell_1}+1)\prod_{z=1}^{k-w}(n_{j_z}+1)}\sum_{\substack{u_{\ell_1}=1}}^{n_{\ell_1}+1}\sum_{\substack{v_{j_z}=1\\j_z\in NS}}^{n_{j_z}+1}P\left(\max_{j_z\in NS}x_{j_z,v_{j_z}}<\ldots<X_{\ell_2}^{f}<x_{\ell_1,u_{\ell_1}-1}\right)$$

$$\geq \frac{1}{\prod_{t=1}^{k}(n_t+1)}\sum_{\substack{u_{\ell_i}=1\\\ell_i\in S}}^{n_{\ell_i}+1}\sum_{\substack{v_{j_z}=1\\j_z\in NS}}^{n_{j_z}+1}$$

$$I\left\{\left(\max_{j_z\in NS}x_{j_z,v_{j_z}}<x_{\ell_w,u_{\ell_w}-1}\right)\wedge\bigcap_{i=1}^{w-1}\left\{x_{\ell_{w-i+1},\,u_{\ell_{w-i+1}}}<x_{\ell_{w-i},\,u_{\ell_{w-i}}-1}\right\}\right\}$$

The upper bound for the NPI lower probability can be derived by assigning the probability mass for group ℓ_1 to the left endpoint per interval and the probability masses for the non-selected groups j_z to the right endpoint per interval. The middle groups ℓ_i, $i=2,\ldots,w$, can be assigned to either the left or the right endpoint per interval, as they have no influence on the upper bound result. An explanation for the absence of an effect on how the probability masses of the middle groups are placed is that if the probability masses for the middle groups $\ell_2,\ldots,\ell_w$ are assigned to the left endpoints, this implies that the lower limit of the first interval would be $-\infty$. When comparing $-\infty$ with any other observations in group ℓ_1, the indicator function $\max_{j_z\in NS}x_{j_z,v_{j_z}}<\ldots<x_{\ell_1,u_{\ell_1}-1}$ is required to hold. It will not affect the calculation, as $-\infty$ represents a bound and no values in group ℓ_1 can be smaller than it. The same scenario can occur when the probability masses for the middle groups $\ell_2,\ldots,\ell_w$ are assigned to the right endpoints and the upper limit of the last interval ∞ added to the data. Since ∞ represents an upper bound, no observations in group ℓ_1 can be greater than it. The upper bound for the NPI lower probability for the event $\max_{j_z\in NS}X_{j_z}^{f}<X_{\ell_w}^{f}<\ldots<X_{\ell_2}^{f}<X_{\ell_1}^{f}$ is

$$\underline{P}^{U}\left(\max_{j_z\in NS}X_{j_z}^{f}<X_{\ell_w}^{f}<\ldots<X_{\ell_2}^{f}<X_{\ell_1}^{f}\right)=\frac{1}{\prod_{t=1}^{k}(n_t+1)}\sum_{\substack{u_{\ell_i}=1\\\ell_i\in S}}^{n_{\ell_i}+1}\sum_{\substack{v_{j_z}=1\\j_z\in NS}}^{n_{j_z}+1}$$

$$I\left(\max_{j_z\in NS}x_{j_z,v_{j_z}}<x_{\ell_w,u_{\ell_w}}<x_{\ell_{w-1},u_{\ell_{w-1}}}<\ldots<x_{\ell_1,u_{\ell_1}-1}\right) \qquad (2.4)$$

The lower bound for the NPI upper probability for the event $\max_{j_z\in NS}X_{j_z}^{f}<X_{\ell_w}^{f}<\ldots<X_{\ell_2}^{f}<X_{\ell_1}^{f}$ is derived in a manner similar to that of the upper bound for the NPI lower probability. However, there is a difference in how the probability masses are assigned for the lower bound for the NPI upper probability. The probability

mass associated with group ℓ_1 is assigned to the right endpoint of the interval. On the other hand, the probability masses for the groups j_z are assigned to the left endpoint of the interval. Thus, the lower bound for the NPI upper probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$ is

$$
\overline{P}^L \left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f \right) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}
$$

$$
I \left(\max_{j_z \in NS} x_{j_z, v_{j_z}-1} < x_{\ell_w, u_{\ell_w}} < x_{\ell_{w-1}, u_{\ell_{w-1}}} < \ldots < x_{\ell_1, u_{\ell_1}} \right) \tag{2.5}
$$

Finally, the upper bound for the NPI upper probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$ can be derived by checking all possible combinations of the intervals $(x_{\ell_i, u_{\ell_i}-1}, x_{\ell_i, u_{\ell_i}})$ and $(x_{j_z, v_{j_z}-1}, x_{j_z, v_{j_z}})$ and counting all the intervals that could overlap, or in other words, by finding all the intervals whose intersections have common points. Thus, the upper bound for the NPI upper probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$ is

$$
\overline{P}^U \left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f \right) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}
$$

$$
\bigcap_{i=1}^w I \left\{ \max_{j_z \in NS} x_{j_z, v_{j_z}-1} < x_{\ell_i, u_{\ell_i}} \right\} \wedge \bigcap_{i=2}^w \bigcap_{a=1}^{i-1} I \left\{ x_{\ell_i, u_{\ell_i}-1} < x_{\ell_a, u_{\ell_a}} \right\} \tag{2.6}
$$

These bounds will be used in applications later in this chapter and in Chapter 5.

2.2.3   Special case for perfectly ordered data

A special scenario arises when the data are perfectly ordered without any overlap between the data from different groups, which means that the groups are well separated. To clarify, all data observations related to ℓ_1 are larger than all data observations related to the second largest group ℓ_2. Similarly, all observations in the second largest group ℓ_2 are larger than those in the third largest group ℓ_3, and so on. The letter s here is added as a subscript for all the bounds of these probabilities to indicate this special case. Equations (2.3), (2.4), (2.5) and (2.6) are reduced to

the following equations respectively.

$$\underline{P}_s^L\big(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_1}^f\big) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \prod_{z=1}^{k-w} n_{j_z} n_{\ell_1} \prod_{i=2}^{w}(n_{\ell_i} - 1) \qquad (2.7)$$

$$\underline{P}_s^U\big(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_1}^f\big) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \prod_{z=1}^{k-w} n_{j_z} \prod_{i=1}^{w} n_{\ell_i} \qquad (2.8)$$

$$\overline{P}_s^L\big(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_1}^f\big) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \prod_{z=1}^{k-w}(n_{j_z} + 1)(n_{\ell_1} + 1) \prod_{i=2}^{w} n_{\ell_i}$$

$$(2.9)$$

$$\overline{P}_s^U\big(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_1}^f\big) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \prod_{z=1}^{k-w}(n_{j_z} + 1) \prod_{i=1}^{w}(n_{\ell_i} + 1) \quad (2.10)$$

These equations offer a straightforward way to calculate the bounds. While this is an extreme and rarely encountered case, it is included to illustrate how the bounds behave when data are perfectly ordered and fully separated.

The next two sections will present two special cases for subsets containing two and three groups. These cases are crucial as they demonstrate how the NPI lower and upper probabilities are derived differently depending on the number of groups in the subset.

2.3 Selecting a ranked subset of two best groups

This section presents the NPI method for selecting a ranked subset of the two best groups in terms of a single future observation from each group. Generally, in Section 2.2, the event of interest was $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$. A special case occurs when the subset is of size two, $w = 2$, thus the event is $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$. The first case to be derived is when the subset S contains two groups. The case where S consists of only one group is not of interest here, as it has already been considered in the literature [20]. Therefore, the focus is on scenarios where S includes at least two groups. Hill's assumption $A_{(n)}$ is applied to each group, as presented in Section 2.2, to derive the NPI lower and upper probabilities for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, where $\ell_i = \{\ell_1, \ell_2\} \subset \{1, \ldots, k\}$. Section 2.2.1 presents the derivation of the NPI lower (upper) probability. The probability mass of the best group ℓ_1 is assigned to the left (right) endpoint of the interval $(x_{\ell_1, u_{\ell_1} - 1}, x_{\ell_1, u_{\ell_1}})$,

while the probability masses of all $j_z \in NS$ groups are assigned to the right (left) endpoint of the interval $(x_{j_z, v_{j_z}-1}, x_{j_z, v_{j_z}})$. However, as mentioned in Section 2.2.1, the main question is how to assign the probability mass $\frac{1}{n_{\ell_2}+1}$ for the middle group ℓ_2 over each interval $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$, where $u_{\ell_2} = 1, 2, \ldots, n_{\ell_2}+1$. The answer will be presented by the following steps for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$. To obtain the exact NPI lower probability, the following optimisation procedure should be applied.

1. Determine the sub-intervals created by the observations from group $\ell_1 \in S$ and $j_z \in NS$, for $z = 1, \ldots, k-2$ within $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$. The number of sub-intervals created is $n_{\ell_1} + \sum_{z=1}^{k-2} n_{j_z} + 1$.

2. Let $S_{x_{\ell_1}}^{t_{\min}^{u_{\ell_2}}}$ represent the number of assigned probability mass $\frac{1}{n_{\ell_1}+1}$ to the left endpoints of each interval. Also, let $S_{x_{j_z}}^{t_{\min}^{u_{\ell_2}}}$ represent the number of assigned probability masses $\frac{1}{n_{j_z}+1}$ to the right endpoints of each interval. The probability mass for group ℓ_2 assigns to any value $t^{u_{\ell_2}}$, where $t^{u_{\ell_2}}$ is a single point (assumed to be the midpoint) belonging to a sub-interval within $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$.

3. The NPI lower probability can then be obtained by minimising $S_{x_{\ell_1}}^{t_{\min}^{u_{\ell_2}}} \times S_{x_{j_z}}^{t_{\min}^{u_{\ell_2}}}$ over all these sub-intervals.

To derive the exact NPI upper probability for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$

1. Determine the sub-intervals created by the observations from group ℓ_1 and j_z within $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$.

2. Let $S_{x_{\ell_1}}^{t_{\min}^{u_{\ell_2}}}$ represent the number of assigned probability mass $\frac{1}{n_{\ell_1}+1}$ to the right endpoints of each interval. Also, let $S_{x_{j_z}}^{t_{\min}^{u_{\ell_2}}}$ represent the number of assigned probability masses $\frac{1}{n_{j_z}+1}$ to the left endpoints of each interval. The probability mass for group ℓ_2 assigns to any value $t^{u_{\ell_2}}$, where $t^{u_{\ell_2}}$ is a single point (assumed to be the midpoint) belonging to a sub-interval within $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$. .

3. The NPI upper probability can then be obtained by maximising $S_{x_{\ell_1}}^{t_{\max}^{u_{\ell_2}}} \times S_{x_{j_z}}^{t_{\max}^{u_{\ell_2}}}$ over all these sub-intervals.
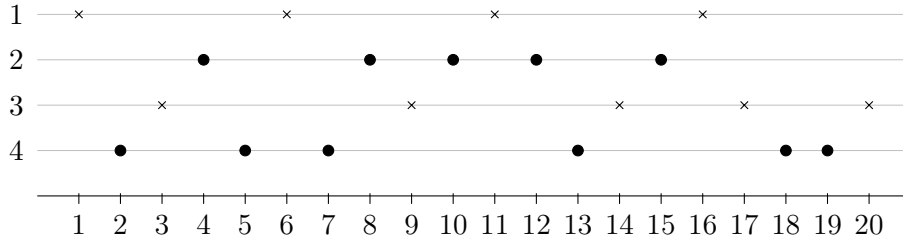
Figure 2.1: Data of Example 2.3.1

The exact NPI lower and upper probabilities for the event $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}$ are

$$\underline{P}(\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{u_{\ell_2}=1}^{n_{\ell_2}+1} S^{t^{u_{\ell_2}}_{\min}}_{x_{\ell_1}} \times S^{t^{u_{\ell_2}}_{\min}}_{x_{j_z}} \qquad (2.11)$$

$$\overline{P}(\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{u_{\ell_2}=1}^{n_{\ell_2}+1} S^{t^{u_{\ell_2}}_{\max}}_{x_{\ell_1}} \times S^{t^{u_{\ell_2}}_{\max}}_{x_{j_z}} \qquad (2.12)$$

In addition, one may wish to avoid the detailed optimisation described earlier, because the optimisation becomes more demanding, especially as the number of intervals for group ℓ_2 increases. Consequently, the number of sub-intervals within $(x_{\ell_2,u_{\ell_2}-1}, x_{\ell_2,u_{\ell_2}})$ also increases, each of which requires a detailed examination. Therefore, the lower and upper bounds for the NPI lower and upper probabilities in Equations (2.3), (2.4), (2.5) and (2.6) with $w = 2$, offer a straightforward alternative.

Example 2.3.1 illustrates the method presented in Section 2.3 for selecting a ranked subset of the two best groups out of four groups with small sample sizes.

Example 2.3.1 Consider a dataset with $k = 4$ groups and sample sizes $n_1 = 4$, $n_2 = 5$, $n_3 = 5$ and $n_4 = 6$. The dataset is provided in Figure 2.1. Let S consist of the dataset that belongs to ℓ_1 and ℓ_2, and the non-selected groups, NS consist of j_1 and j_2 [20]. The event of interest is $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}$. There are 12 different possible combinations for selecting a ranked subset of the two best groups out of four groups, as presented in Table 2.1. Starting with the first case, it is assumed that the groups in the subset are 1 as the best group and 2 as the second best group, and the groups 3 and 4 belong to NS.

To obtain the exact NPI lower probability for the event $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, the probability mass corresponding to group 1 should be assigned to the left endpoint per interval, while the probability masses corresponding to groups 3 and 4 should be assigned to the right endpoint per interval.

With regard to group 2, its corresponding probability mass can be assigned to any point $t^{u_{\ell_2}}$, where $u_{\ell_2} = 1, \ldots, 6$, that belongs to specific sub-intervals within each (x_{2,u_2-1}, x_{2,u_2}), in order to minimise the probability. Consider all possible combinations of the sub-intervals that lie within group 2 intervals. Then select the sub-intervals in each (x_{2,u_2-1}, x_{2,u_2}), interval that minimises the probability. There are six intervals for group 2: $(-\infty, 4)$, $(4, 8)$, $(8, 10)$, $(10, 12)$, $(12, 15)$ and $(15, \infty)$.

For the first interval $(-\infty, 4)$, there are four possible sub-intervals where the probability mass for group 2 can be assigned: $(-\infty, 1)$, $(1, 2)$, $(2, 3)$ and $(3, 4)$. These sub-intervals are created by the observations from group 1, which belongs to S, and groups 3 and 4, which belong to NS. Let $t_{\min}^1$ be any point belonging to the sub-interval $(-\infty, 3)$, which is the sub-interval that minimises the probability, implying that no probability mass to the left of this sub-interval is assigned from groups 3 and 4. Therefore, $S_{(3,4)}^{t_{\min}^1} = 0$. On the other hand, the number of probability mass corresponding to group 1 to the right of this sub-interval is equal to 4, denoted as $S_1^{t_{\min}^1} = 4$. Thus, $S_{(3,4)}^{t_{\min}^1} \times S_1^{t_{\min}^1} = 0$.

Moving on to the second interval $(4, 8)$, the sub-interval $(4, 5)$ is the best choice as it helps minimise the probability for the event of interest. Let $t_{\min}^2$ belong to $(4, 5)$, where the number of the probability mass of groups 3 and 4 assigned to the left of the sub-interval $(4, 5)$ is 1, and the number of the probability mass of group 1 assigned to the right of the same sub-interval is 3. Therefore, $S_{(3,4)}^{t_{\min}^2} \times S_1^{t_{\min}^2} = 3$.

For the third interval $(8, 10)$, there are two sub-intervals: $(8, 9)$ and $(9, 10)$. The sub-interval $(8, 9)$ is the best choice, therefore, $t_{\min}^3$ belonging to $(8, 9)$ results in 6 probability masses to groups 1, 3 and 4 observations. Similarly, the fourth and fifth sub-intervals, $(11, 12)$ and $(12, 13)$ respectively, yield the same number of probability masses as the third sub-interval.

Finally, for the last sub-interval $(16, \infty)$, the number of probability mass assigned by group 1 to the right of this sub-interval is 0. Hence, $(16, \infty)$ is the sub-interval

that minimises the probability with 0 probability mass. The exact NPI lower probability is

$$\underline{P}(\max(X_3^f, X_4^f) < X_2^f < X_1^f) = \frac{1}{\prod_{t=1}^4 (n_t + 1)} \sum_{u_{\ell_2}=1}^6 S_{(3,4)}^{t_{\min}^{u_{\ell_2}}} \times S_1^{t_{\min}^{u_{\ell_2}}}$$
$$= \frac{1}{1260}(21) = 0.0167$$

To derive the exact NPI upper probability, apply the same strategy as mentioned earlier. The selected sub-interval provides the exact NPI upper probability as follows: $(3,4)$ with 16 probability masses, $(7,8)$ with 24 probability masses, $(9,10)$ with 36 probability masses, $(10,11)$ with 36 probability masses, $(14,15)$ with 40 probability masses, and $(20,\infty)$ with 41 probability masses. The exact NPI upper probability is

$$\overline{P}(\max(X_3^f, X_4^f) < X_2^f < X_1^f) = \frac{1}{\prod_{t=1}^4 (n_t + 1)} \sum_{u_{\ell_2}=1}^6 S_{(3,4)}^{t_{\max}^{u_{\ell_2}}} \times S_1^{t_{\max}^{u_{\ell_2}}}$$
$$= \frac{1}{1260}(194) = 0.1539$$

The lower and upper bounds for the exact NPI lower and upper probabilities are given by Equations (2.3), (2.4), (2.5) and (2.6) for $w = 2$. Specifically, the lower and upper bounds for the NPI lower probability are given by

$$\left[\underline{P}^L, \underline{P}^U \right] (\max(X_3^f, X_4^f) < X_2^f < X_1^f) = \left[0.0158, 0.0309 \right]$$

The lower and upper bounds for the NPI upper probability is given by

$$\left[\overline{P}^L, \overline{P}^U \right] (\max(X_3^f, X_4^f) < X_2^f < X_1^f) = \left[0.1111, 0.1754 \right]$$

In this example, the exact NPI lower probability, which is 0.0167 in this case, lies between its corresponding lower and upper bounds. The exact NPI upper probability, which is 0.1539 in this example, lies between its lower and upper bounds. Furthermore, the exact NPI lower probability is closer to its corresponding lower bound, 0.0158, and the exact NPI upper probability is closer to its corresponding upper bound, 0.1754.

Table 2.1 displays the exact NPI lower and upper probabilities associated with selecting a ranked subset of the two best groups for all 12 possible combinations.

October 6, 2025

ℓ_1	ℓ_2	$\underline{P}$	$\overline{P}$
1	2	0.0167	0.1539
2	1	0.0063	0.1317
1	3	0.0071	0.1889
3	1	0.0278	0.2302
1	4	0.0032	0.1484
4	1	0.0151	0.1913
2	3	0.0103	0.1921
3	2	0.0508	0.2881
2	4	0.0079	0.1500
4	2	0.0269	0.2071
3	4	0.0413	0.2794
4	3	0.0286	0.2698

Table 2.1: The exact NPI lower and upper probabilities for all group selections categories for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$ for Example 2.3.1

Choosing group 3 as the best group and 2 as the second best group results in giving the event of interest, as it yields the highest exact NPI lower and upper probabilities. Additionally, selecting group 3 as the best and group 4 as the second best results in second highest exact NPI lower and upper probabilities. However, the exact NPI lower and upper probabilities are very small, with the imprecision ranging from 0.1254 to 0.2381.

The lowest exact NPI lower probability occurs when selecting group 1 as the best and 4 as the second best. In addition, the lowest exact NPI upper probability arises when choosing group 2 as the best and 1 as the second best.

$\diamond$

The following example illustrates the NPI method for selecting a ranked subset of the two best groups using simulated data from the Normal distribution for further investigation of the method's performance.

Example 2.3.2 This example investigates the method for selecting a ranked subset of the two best groups among four groups via simulation. A simulation was conducted using Normal distribution. Each group is represented as follows: $X_1 \sim N(0, 1)$, $X_2 \sim N(0.1, 1)$, $X_3 \sim N(0.2, 1)$ and $X_4 \sim N(0.3, 1)$, with a common sample size $n = 25$ for all groups.

In Table 2.2, the exact NPI lower and upper probabilities are presented for 12 different combinations of group selection involving four groups. Table 2.2 presents the results of simulation studies under different numbers of simulation replications 1, 1000 and 10,000, where the number of simulation replications is denoted by h. The objective is to examine how the number of simulation replications influences the selection of the two best groups.

The results in Table 2.2 show that increasing the number of simulation replications from 1 to 1000 and 10,000 leads to consistent conclusions. Selecting group 4 as the best and group 3 as the second best is supported by the highest exact NPI lower and upper probabilities across all replication levels. In contrast, choosing group 1 as the best and group 2 as the second , making it the least favorable combination.

The changes between the exact NPI lower and upper probability values are minimal, and most combinations differ only in the third decimal place when the number of replications is 1000 and 10000. However, some combinations, such as the exact NPI lower probability of selecting group 2 as the best and group 3 as the second best, and the exact NPI upper probability of selecting group 1 as the best and group 3 as the second best, show slightly larger variations. Overall, increasing the number of replications from 1 to 1,000 and then to 10,000 results in only minor changes in the exact NPI lower and upper probabilities.

Finally, the impression, the difference between the exact NPI lower and upper probabilities remains stable as the number of simulations increases from 1 to 10,000.

$\diamond$

Based on the preceding examples, some conclusions can be drawn. In order to improve the computation of the exact NPI lower and upper probabilities, the first sub-interval within the first interval is likely to provide the minimum exact NPI lower probability, as no probability mass is assigned to the right of such a

		1 simulation		1000 simulation		10000 simulation	
ℓ_1	ℓ_2	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
1	2	0.0323	0.0653	0.0456	0.0814	0.0454	0.0812
2	1	0.0333	0.0663	0.0474	0.0838	0.0488	0.0857
1	3	0.0345	0.0657	0.0521	0.0911	0.0504	0.0889
3	1	0.0369	0.0693	0.0589	0.1002	0.0580	0.0987
1	4	0.0550	0.0969	0.0565	0.0981	0.0557	0.0970
4	1	0.0666	0.1116	0.0697	0.1151	0.0695	0.1147
2	3	0.0591	0.1015	0.0585	0.1004	0.0600	0.1023
3	2	0.0520	0.0904	0.0651	0.1087	0.0645	0.1081
2	4	0.0851	0.1406	0.0645	0.1091	0.0665	0.1119
4	2	0.1079	0.1657	0.0765	0.1244	0.0773	0.1257
3	4	0.0854	0.1393	0.0798	0.1299	0.0789	0.1287
4	3	0.1097	0.1669	0.0859	0.1378	0.0856	0.1366

Table 2.2: The exact NPI lower and upper probabilities for all group selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$ for Example 2.3.2

sub-interval. Similarly, the last sub-interval within the last interval yields the same result. However, for the intervals in between, the minimum number of probability masses is placed to the right and left of the initial sub-interval in each interval.

Conversely, for the exact NPI upper probability, the last sub-interval within the first interval is likely to provide the maximum exact NPI upper probability. Similarly, the last sub-interval within the last interval also yields the same result. However, for the intervals in between, the maximum number of probability masses is placed to the right and left of the ending sub-intervals in each interval.

To derive the exact NPI lower and upper probabilities when the data are perfectly ordered without any overlap, Equations (2.7), (2.8), (2.9), and (2.10) are used to derive the lower and upper bounds for the exact NPI lower and upper probabilities for $w = 2$. This approach shows that the exact NPI lower probability matches its lower bound for the NPI lower probability, denoted by $\underline{P} = \underline{P}_s^L$. Similarly, the

exact NPI upper probability matches its upper bound for the NPI upper probability, denoted by $\overline{P} = \overline{P}_s^U$.

2.4 Selecting a ranked subset of three best groups

Section 2.3 presented the NPI lower and upper probabilities for a ranked subset of the two best groups. This section considers the selection and ranking of future observations within a subset of the three best groups. Therefore, the event of interest is $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f$. This representation implies that the future observation for each group that belongs to S is greater than that for each group not belonging to S, where the future observations for groups in S are ranked from the smallest $X_{\ell_3}^f$ to the largest $X_{\ell_1}^f$.

Now, with two or more groups in the middle, denoted by ℓ_2 and ℓ_3, it becomes challenging to find the exact NPI lower and upper probabilities especially for large data sets with much overlap between groups. An optimisation is proposed for each group independently [20]. The goal is to find optimal sub-intervals for each group independently, and then derive the approximate NPI lower and upper probabilities based on this optimisation process over the sub-intervals. To derive the approximation of the NPI lower and upper probabilities, Hill's assumption $A_{(n)}$ per group is applied. With regard to the best group and the groups in NS, it was explained in Section 2.2.1 where their corresponding probability masses in each interval should be assigned. For the middle groups ℓ_3 and ℓ_2, the main question is how to assign their corresponding probability masses $\dfrac{1}{n_{\ell_3} + 1}$ and $\dfrac{1}{n_{\ell_2} + 1}$ over each interval $(x_{\ell_3, u_{\ell_3}-1}, x_{\ell_3, u_{\ell_3}})$ and $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$, respectively. In order to solve this, a heuristic algorithm is proposed to provide approximations for the NPI lower and upper probabilities. To derive approximate NPI lower and upper probabilities for a ranked subset of the three best groups, optimisation using Equations (2.1) and (2.2) in Section 2.2.1, for $w = 3$ is required.

First, the optimisation will focus on the group represented by ℓ_3. For the other group, ℓ_2, it is irrelevant whether the probability mass is assigned to the left or right endpoint of each interval; therefore, it will be neglected for now.

$$\underline{P}^{\min \ell_3}\left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f\right) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}$$

$$P\left(\max_{j_z \in NS} x_{j_z,v_{j_z}} < X_{\ell_3}^f < x_{\ell_2,u_{\ell_2}} \Big| X_{\ell_3}^f \in \left(x_{\ell_3,u_{\ell_3}-1}, x_{\ell_3,u_{\ell_3}}\right)\right) \times I\left(x_{\ell_2,u_{\ell_2}} < x_{\ell_1,u_{\ell_1}-1}\right)$$

$$(2.13)$$

where $\displaystyle\sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1}$ is used for 3 sums $\displaystyle\sum_{u_{\ell_1}=1}^{n_{\ell_1}+1} \sum_{u_{\ell_2}=1}^{n_{\ell_2}+1} \sum_{u_{\ell_3}=1}^{n_{\ell_3}+1}$.

$$\overline{P}^{\max \ell_3}\left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f\right) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}$$

$$P\left(\max_{j_z \in NS} x_{j_z,v_{j_z}-1} < X_{\ell_3}^f < x_{\ell_2,u_{\ell_2}} \Big| X_{\ell_3}^f \in \left(x_{\ell_3,u_{\ell_3}-1}, x_{\ell_3,u_{\ell_3}}\right)\right) \times I\left(x_{\ell_2,u_{\ell_2}} < x_{\ell_1,u_{\ell_1}}\right)$$

$$(2.14)$$

To minimise Equation (2.13) for the lower probability over group ℓ_3, assign the probability mass $\frac{1}{n_{\ell_3}+1}$ to a sub-interval that minimises the probability. Then select the combination of sub-intervals within all $(x_{\ell_3,u_{\ell_3}-1}, x_{\ell_3,u_{\ell_3}})$ intervals that minimises the probability. In contrast, to maximise Equation (2.14) for the upper probability over group ℓ_3, assign the probability mass $\frac{1}{n_{\ell_3}+1}$ to a sub-interval that maximises the probability. Then select the combination of sub-intervals within all $(x_{\ell_3,u_{\ell_3}-1}, x_{\ell_3,u_{\ell_3}})$ intervals that maximise the probability.

Secondly, to optimise over group ℓ_2, apply the same procedure to group ℓ_3, but a modification to the intervals is needed. Instead of using the intervals $(x_{\ell_3,u_{\ell_3}-1}, x_{\ell_3,u_{\ell_3}})$ as before, the intervals $(x_{\ell_2,u_{\ell_2}-1}, x_{\ell_2,u_{\ell_2}})$ will be used now. Then

$$\underline{P}^{\min \ell_2}\left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f\right) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}$$

$$P\left(x_{\ell_3,u_{\ell_3}} < X_{\ell_2}^f < x_{\ell_1,u_{\ell_1}-1} \mid X_{\ell_2}^f \in \left(x_{\ell_2,u_{\ell_2}-1}, x_{\ell_2,u_{\ell_2}}\right)\right) \times I\left(\max_{j_z \in NS} x_{j_z,v_{j_z}} < x_{\ell_3,u_{\ell_3}}\right)$$

$$(2.15)$$
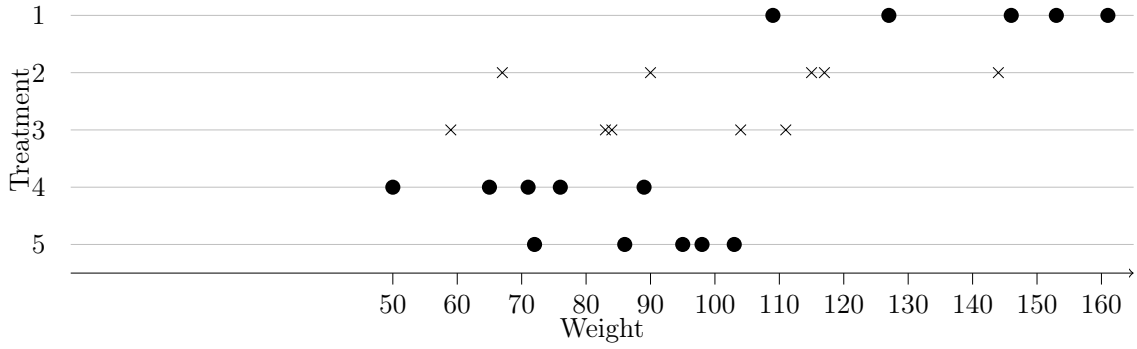
Figure 2.2: Fish weights across treatment levels for Example 2.4.1

$$
\overline{P}^{\max \ell_2}\left(\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f\right) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i \in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z \in NS}}^{n_{j_z}+1}
$$
$$
P\left(x_{\ell_3, u_{\ell_3}} < X_{\ell_2}^f < x_{\ell_1, u_{\ell_1}} \mid X_{\ell_2}^f \in (x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})\right) \times I\left(\max_{j_z \in NS} x_{j_z, v_{j_z}-1} < x_{\ell_3, u_{\ell_3}}\right)
$$
$$
(2.16)
$$

Finally, determine the approximation of the NPI lower probability by selecting the lowest of the two NPI lower probabilities over ℓ_2 and ℓ_3, denoted as $\underline{P}^{\min \ell_2}$ and $\underline{P}^{\min \ell_3}$ respectively. Similarly, determine the approximation of the NPI upper probability by selecting the largest of the two NPI upper probabilities over ℓ_2 and ℓ_3, denoted as $\overline{P}^{\max \ell_2}$ and $\overline{P}^{\max \ell_3}$ respectively.

By following the same reasoning and principles outlined in Section 2.2, lower and upper bounds for the NPI lower and upper probabilities in Equations (2.3), (2.4), (2.5) and (2.6) with $w = 3$, offer a straightforward alternative.

The following example illustrates the method for selecting a ranked subset of the three best groups.

Example 2.4.1 The lengthWeight dataset in the StatCharme Package in R contains variables which are: The chemical concentration, the fish were exposed to 6 levels of treatments $\{1, 2, 3, 4, 5, 6\}$ and weight of the fish. In this example, a small dataset is used instead of a large one, as the smaller dataset helps to explain each sub-interval and identify the optimal one for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f$. Therefore, 5 datasets are created for the weight variable corresponding to 5 lev-

els of treatments. The weight of the fish in three subset groups 1, 2 and 3 that correspond to treatment levels $1, 2$ and 3, respectively, are $\{109, 127, 146, 153, 161\}$, $\{67, 90, 115, 117, 144\}$ and $\{59, 83, 84, 104, 111\}$. The two groups that belong to NS are group 4, which consists of fish with treatment 4 and weights $\{50, 65, 71, 76, 89\}$, and group 5, which consists of fish with treatment 5 and weights $\{72, 86, 95, 98, 103\}$. The dataset is provided in Figure 2.2.

The optimisation over middle groups that are 2 and 3 should be performed. First, the process will start with group 2 and try to find the approximate NPI lower probability. With regard to group 3, its corresponding probability mass can be assigned to either the left or to the right endpoint per interval. However, the probability mass for group 2 should be assigned to certain sub-intervals. The combination of those sub-intervals in each (x_{2,u_2-1}, x_{2,u_2}) interval should minimise Equation (2.15). In this case, there are six intervals: $(-\infty, 67)$, $(67, 90)$, $(90, 115)$, $(115, 117)$, $(117, 144)$ and $(144, \infty)$.

For the first interval, which is $(-\infty, 67)$, there are no $\max(x_{4,v_4}, x_{5,v_5})$ values that satisfy the data. What the data mean here is the first indicator function $I(x_{3,u_3} < x_{2,1}^* < x_{1,u_1-1}) \times I(\max(x_{4,v_4}, x_{5,v_5}) < x_{3,u_3})$. In other words, the number of observations that satisfy this indicator function within this interval, $(-\infty, 67)$, is zero. Therefore, the probability mass corresponding to group 2, $\dfrac{1}{n_2 + 1}$ can be assigned to any point within this interval, suppose 66. For the second interval $(67, 90)$, there are also no values of $\max(x_{4,v_4}, x_{5,v_5})$ satisfying the indicator function $I(x_{3,u_3} < x_{2,2}^* < x_{1,u_1-1}) \times I(\max(x_{4,v_4}, x_{5,v_5}) < x_{3,u_3})$ in a specific sub-interval $(67, 83)$. The probability mass for group 2 is assigned to 75. For the third interval $(90, 115)$, $(90, 104)$ is the optimal sub-interval that gives the minimum number of assigned probability mass, which leads to minimise the probability. There are 40 observations for which the indicator function $I(x_{3,u_3} < x_{2,3}^* < x_{1,u_1-1}) \times I(\max(x_{4,v_4}, x_{5,v_5}) < x_{3,u_3})$ is required to hold. In this situation, the probability mass that belongs to group 2 is assigned to a chosen point, which is 97. For the fourth interval $(115, 117)$, the number of assigned probability mass for which the indicator function $I(x_{3,u_3} < x_{2,4}^* < x_{1,u_1-1}) \times I(\max(x_{4,v_4}, x_{5,v_5}) < x_{3,u_3})$ is required to hold within this interval is 232. Therefore, the probability mass belonging to

group 2 is assigned to 116. $(117, 144)$ is the fifth interval, the sub-interval that helps us to minimise the probability is $(127, 144)$ with the number of observations 174. The chosen point to which the probability mass can be assigned for group 2 is 135.5. Finally, for the last interval $(144, \infty)$, there are no observations satisfying the indicator $I(x_{3,u_3} < x_{2,6}^* < x_{1,u_1-1}) \times I(\max(x_{4,v_4}, x_{5,v_5}) < x_{3,u_3})$ within the sub-interval $(161, \infty)$. Therefore, the probability mass $\dfrac{1}{n_2 + 1}$ should be allocated to any point within this sub-interval, suppose 170.

$$\underline{P}^{\min 2}(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.0574$$

For group 2, its corresponding probability mass should be assigned to a certain sub-interval that maximises Equation (2.16). Likewise, the probability mass corresponding to group 3 can be assigned to the left or right endpoint per interval. The probability mass corresponding to group 2 for the intervals $(x_{2,0}, x_{2,1})$, $(x_{2,1}, x_{2,2})$, $(x_{2,2}, x_{2,3})$, $(x_{2,3}, x_{2,4})$, $(x_{2,4}, x_{2,5})$, $(x_{2,5}, x_{2,6})$ is allocated to the following subintervals, respectively: $(60, 67)$, $(85, 90)$, $(112, 115)$, $(115, 117)$, $(117, 126)$, and $(144, 145)$, with the corresponding points of interest being 66, 87.5, 113.5, 116, 121.5, and 144.5, respectively. The number of observations that satisfy the data for these sub-intervals was 12, 132, 470, 470, 470, and 376. Therefore, the NPI upper probability over group 2 is

$$\overline{P}^{\max 2}(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.2482$$

Secondly, the optimisation process over group 3, similar to the optimisation process over group 2, begins by allocating $\frac{1}{n_2+1}$ to either the left or right endpoint per interval. To minimise Equation (2.13) and maximise Equation (2.14), the probability mass of group 3 is assigned to specific sub-intervals within each (x_{3,u_3-1}, x_{3,u_3}) interval, then

$$\underline{P}^{\min 3}(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.0518$$

where $x_{3,1}^* = (-\infty, 59)$, $x_{3,2}^* = (59, 72)$, $x_{3,3}^* = (83, 84)$, $x_{3,4}^* = (84, 86)$, $x_{3,5}^* = (104, 111)$, and $x_{3,6}^* = (144, \infty)$.

$$\overline{P}^{\max 3}(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.2526$$

where $x_{3,1}^* = (51, 59)$, $x_{3,2}^* = (77, 83)$, $x_{3,3}^* = (83, 84)$, $x_{3,4}^* = (103, 104)$, $x_{3,5}^* = (104, 111)$, and $x_{3,6}^* = (111, 114)$.

Consequently, a heuristic algorithm is proposed to provide the approximate NPI lower and upper probabilities. The approximate NPI lower probability derived as the minimum value of $\underline{P}^{\min 2} = 0.0574$ and $\underline{P}^{\min 3} = 0.0518$. Thus, the approximate NPI lower probability is 0.0518. Similarly, the approximate NPI upper probability is derived as the maximum value of $\overline{P}^{\max 2} = 0.2482$ and $\overline{P}^{\max 3} = 0.2526$. Hence, the approximate NPI upper probability is 0.2526.

In this example, a notable improvement in the optimisation process can be achieved by observing that each sub-interval contributing to the minimisation of the approximate NPI lower probability is primarily positioned at the beginning of its corresponding interval. It is worth noting that, for the last interval, the sub-interval that minimises the approximate NPI lower probability is placed at the end of its corresponding interval.

However, for the maximisation of the approximate NPI upper probability, each sub-interval is strategically positioned at the end of its corresponding interval. For the last interval, the sub-interval that contributes to the maximisation of the upper probability is placed at the beginning of the interval.

For the same reasons mentioned earlier in Section 2.2, the lower and upper bounds for the NPI lower and upper probabilities can be calculated. The lower and upper bounds for the NPI lower probability are

$$\underline{P}^L(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.0318$$
$$\underline{P}^U(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.0872$$

The lower and upper bounds for the NPI upper probability are

$$\overline{P}^L(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.1876$$
$$\overline{P}^U(\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f) = 0.3645$$

The approximate NPI lower and upper probabilities should be nested between their corresponding bounds. For the approximate NPI lower probability, $\underline{P}^L \leq \underline{P} \leq \underline{P}^U$ equal to $0.0318 < 0.0518 < 0.0872$, it is obvious that the approximate NPI lower probability is closer to its corresponding lower bound, which is equal to 0.0318. However, for the approximate NPI upper probability, which is 0.2526, is between its corresponding lower bound 0.1876 and upper bound 0.3645 respectively.

Previously, a detailed explanation was provided on how the approximate NPI lower and upper probabilities were found for the event $\max(X_4, X_5)^f < X_3^f < X_2^f < X_1^f$. Next, the same calculations need to be performed for all possible group combinations that would result in 60 cases. Since this number is very large, the process is simplified by fixing the groups that belong to NS, which are 4 and 5, and focusing only on the different combinations of the three groups that belong to S, which are 1, 2 and 3. In Table 2.3, the approximate NPI lower and upper probabilities for different ranked subset of the three best groups are presented. There are a total of 6 combinations formed by these groups, with each combination representing different approximate NPI lower and upper probabilities.

Table 2.3 shows that predicting that group 1 to provide the largest future observation, followed by group 2 for the second largest future observation and group 3 for the smallest future observation, leads to make a good decision. The corresponding approximate NPI lower and upper probabilities are the highest compared to the rest, with values of 0.0518 and 0.2526, respectively. In contrast, predicting group 2 to provide the largest future observation, followed by group 3 for the second largest future observation and group 1 for the smallest future observation, leads to a poor decision. Its corresponding approximate NPI lower and upper probabilities are the lowest compared to the rest, with values of 0.0000 and 0.0417, respectively. This is because the data corresponding to groups 1 and 2 are slightly larger than the others, which is visualized in Figure 2.2. It is observed that combinations suggesting that groups 1 and 2 provide the smallest future observation generally result in low approximate NPI lower and upper probabilities.

$\diamond$

October 6, 2025

ℓ_1	ℓ_2	ℓ_3						
1	2	3	$\underline{P}^{\min 2}$	$\overline{P}^{\max 2}$	$\underline{P}^{\min 3}$	$\overline{P}^{\max 3}$	$\underline{P}$	$\overline{P}$
			0.0574	0.2482	0.0518	0.2526	0.0518	0.2526
2	1	3	$\underline{P}^{\min 1}$	$\overline{P}^{\max 1}$	$\underline{P}^{\min 3}$	$\overline{P}^{\max 3}$	$\underline{P}$	$\overline{P}$
			0.0075	0.1386	0.0073	0.1227	0.0073	0.1386
2	3	1	$\underline{P}^{\min 1}$	$\overline{P}^{\max 1}$	$\underline{P}^{\min 3}$	$\overline{P}^{\max 3}$	$\underline{P}$	$\overline{P}$
			0.0000	0.0309	0.0000	0.0417	0.0000	0.0417
3	1	2	$\underline{P}^{\min 1}$	$\overline{P}^{\max 1}$	$\underline{P}^{\min 2}$	$\overline{P}^{\max 2}$	$\underline{P}$	$\overline{P}$
			0.0000	0.0864	0.0000	0.0949	0.0000	0.0949
1	3	2	$\underline{P}^{\min 2}$	$\overline{P}^{\max 2}$	$\underline{P}^{\min 3}$	$\overline{P}^{\max 3}$	$\underline{P}$	$\overline{P}$
			0.0000	0.0809	0.0051	0.0968	0.0000	0.0968
3	2	1	$\underline{P}^{\min 1}$	$\overline{P}^{\max 1}$	$\underline{P}^{\min 2}$	$\overline{P}^{\max 2}$	$\underline{P}$	$\overline{P}$
			0.0000	0.0324	0.0000	0.0463	0.0000	0.0463

Table 2.3: Approximate NPI lower and upper probabilities for the event $\max(X_4^f, X_5^f) < X_{\ell_3}^f < X_{\ell_2}^f < X_{\ell_1}^f$ for Example 2.4.1

2.5 Selecting a partially ranked subset

Selecting a partially ranked subset may be of interest in statistics when there is a need to compare or rank groups based on certain criteria, but a complete ranking is not meaningful or necessary. The partially ranked subset is more practical than complete ranking in such cases. For example, in drug development, researchers may select a subset of the most effective treatments without ranking all of them. Instead, only the top treatments within this subset are ranked, while the rest remain unranked.

This section introduces the notation and formulas for this scenario. Let S be a selected subset, and let $S^* \subseteq S$ be a partially ranked subset, where some groups are ranked. Let $w = |S|$ be the total number of groups in S, and let $w^* = |S^*|$ be the number of ranked groups, so $1 \leq w^* \leq w$. $X_{\ell_{i*}}^f$ denotes all future observations that are in a partially ranked subset. The aim here is to compute the NPI lower and upper

probabilities for the event of interest $(\max\limits_{j_z\in NS} X^f_{j_z} < \min\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i}) \wedge (\max\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i} < X^f_{\ell_{w^*}} < \ldots < X^f_{\ell_{1^*}})$, where $i^* \subset \{1, 2, \ldots, w^*\}$. However, obtaining the NPI lower and upper probabilities for the above event is computationally demanding and cumbersome, as highlighted by Coolen-Maturi [20]. To this end, bounds for the NPI lower and upper probabilities are provided, with their derivations explained in Section 2.2.2.

The lower bound for the NPI lower probability for the event $(\max\limits_{j_z\in NS} X^f_{j_z} < \min\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i}) \wedge (\max\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i} < X^f_{\ell_{w^*}} < \ldots < X^f_{\ell_{1^*}})$ is

$$
\underline{P}^L\left(\left(\max_{j_z\in NS} X^f_{j_z} < \min_{\ell_i\in S\setminus S^*} X^f_{\ell_i}\right) \wedge \left(\max_{\ell_i\in S\setminus S^*} X^f_{\ell_i} < X^f_{\ell_{w^*}} < \ldots < X^f_{\ell_{1^*}}\right)\right) =
$$

$$
\frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{u_{\ell_{i^*}}=1 \\ \ell_{i^*}\in S^*}}^{n_{\ell_{i^*}}+1} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i\in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z\in NS}}^{n_{j_z}+1} I\left(\left(\max_{j_z\in NS} x_{j_z,v_{j_z}} < \min_{\ell_i\in S\setminus S^*} x_{\ell_i,u_{\ell_i}-1}\right)\right.
$$

$$
\left. \wedge \left(\max_{\ell_i\in S\setminus S^*} x_{\ell_i,u_{\ell_i}} < x_{\ell_{w^*},u_{\ell_{w^*}}-1}\right) \wedge \bigcap_{i=1}^{w^*-1}\left\{x_{\ell_{w^*-i+1},u_{\ell_{w^*-i+1}}} < x_{\ell_{w^*-i},u_{\ell_{w^*-i}}-1}\right\}\right)
$$

(2.17)

where the notations $\sum\limits_{\substack{v_{j_z}=1 \\ j_z\in NS}}^{n_{j_z}+1}$ is used for $k-w$ sums $\sum\limits_{v_{j_1}=1}^{n_{j_1}+1}\sum\limits_{v_{j_2}=1}^{n_{j_2}+1}\ldots\sum\limits_{v_{j_{k-w}}=1}^{n_{j_{k-w}}+1}$, $\sum\limits_{\substack{u_{\ell_i}=1 \\ \ell_i\in S\setminus S^*}}^{n_{\ell_i}+1}$ is used

for w sums $\sum\limits_{u_{\ell_1}=1}^{n_{\ell_1}+1}\sum\limits_{u_{\ell_2}=1}^{n_{\ell_2}+1}\ldots\sum\limits_{u_{\ell_w}=1}^{n_{\ell_w}+1}$, and $\sum\limits_{\substack{u_{\ell_{i^*}}=1 \\ \ell_{i^*}\in S^*}}^{n_{\ell_{i^*}}+1}$ is used for w^* sums $\sum\limits_{u_{\ell_{1^*}}=1}^{n_{\ell_{1^*}}+1}\sum\limits_{u_{\ell_{2^*}}=1}^{n_{\ell_{2^*}}+1}\ldots\sum\limits_{u_{\ell_{w^*}}=1}^{n_{\ell_{w^*}}+1}$.

The upper bound for the NPI upper probability for the event $(\max\limits_{j_z\in NS} X^f_{j_z} < \min\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i}) \wedge (\max\limits_{\ell_i\in S\setminus S^*} X^f_{\ell_i} < X^f_{\ell_{w^*}} < \ldots < X^f_{\ell_{1^*}})$ is

$$
\overline{P}^U\left(\left(\max_{j_z\in NS} X^f_{j_z} < \min_{\ell_i\in S\setminus S^*} X^f_{\ell_i}\right) \wedge \left(\max_{\ell_i\in S\setminus S^*} X^f_{\ell_i} < X^f_{\ell_{w^*}} < \ldots < X^f_{\ell_{1^*}}\right)\right) =
$$

$$
\frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{u_{\ell_{i^*}}=1 \\ \ell_{i^*}\in S^*}}^{n_{\ell_{i^*}}+1} \sum_{\substack{u_{\ell_i}=1 \\ \ell_i\in S}}^{n_{\ell_i}+1} \sum_{\substack{v_{j_z}=1 \\ j_z\in NS}}^{n_{j_z}+1} I\left(\left(\max_{j_z\in NS} x_{j_z,v_{j_z}-1} < \min_{\ell_i\in S\setminus S^*} x_{\ell_i,u_{\ell_i}}\right)\right.
$$

$$
\wedge \bigcap_{i^*=1}^{w^*}\left(\max_{j_z\in NS} x_{j_z,v_{j_z}-1} < x_{\ell_{i^*},u_{\ell_{i^*}}}\right) \quad \wedge \bigcap_{i^*=1}^{w^*}\left(\min_{\ell_i\in S\setminus S^*} x_{\ell_i,u_{\ell_i}-1} < x_{\ell_{i^*},u_{\ell_{i^*}}}\right)
$$

$$
\left. \wedge \bigcap_{i^*=2}^{w^*}\bigcap_{a=1}^{i^*-1}\left(x_{\ell_{i^*},u_{\ell_{i^*}}-1} < x_{\ell_a,u_{\ell_a}}\right)\right)
$$

(2.18)

means for $\delta^* = 0.2$	means for $\delta^* = 0.4$	means for $\delta^* = 0.6$
$\mu_1 = 0.0$	$\mu_1 = 0.0$	$\mu_1 = 0.0$
$\mu_2 = 0.2$	$\mu_2 = 0.4$	$\mu_2 = 0.6$
$\mu_3 = 0.4$	$\mu_3 = 0.8$	$\mu_3 = 1.2$
$\mu_4 = 0.6$	$\mu_4 = 1.2$	$\mu_4 = 1.8$

Table 2.4: Simulated mean values of a Normal distribution for $\delta^* = 0.2, 0.4$ and 0.6

δ^*	P^*	n	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
0.2	0.95	237	0.1947	0.2492	0.3387	0.5208
0.2	0.90	174	0.2155	0.2899	0.5458	0.4013
0.2	0.75	90	0.1737	0.1180	0.2311	0.5620
0.4	0.95	59	0.02373	0.1064	0.4203	0.5523
0.4	0.90	43	-0.0770	-0.0784	0.4915	0.4367
0.4	0.75	22	-0.0618	0.0437	0.4417	0.5927
0.6	0.95	26	0.1086	0.2106	1.0626	1.0225
0.6	0.90	19	0.5011	0.1545	1.2626	0.9839
0.6	0.75	10	0.0952	-0.3054	0.7606	1.1078

Table 2.5: Sample sizes and means for four populations, based on simulated data from Normal distributions with standard deviation $\sigma = 1$. Results are shown for varying values of δ^* and P^* according to Bechhofer's method

2.6 Application

This section compares three methods for selecting a ranked subset of the two best groups: the indifference zone procedure in Section 1.2.1, Gupta's subset selection procedure in Section 1.2.2, and the NPI method introduced in Section 2.3. The objective is to evaluate their performance in predictive inference and determine whether they select the same subset and the same ranking within that subset.

This section presents a simulated study with sampling from Normal distributions, assuming $k = 4$ populations with a common standard deviation $\sigma = 1$. In

δ^*	P^*	n	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$
0.2	0.95	98	0.1574	0.1362	0.3554	0.7858
0.2	0.90	59	-0.0446	0.5234	0.1905	0.9711
0.4	0.95	25	0.0494	0.2908	1.1190	1.4147
0.4	0.90	15	-0.1052	0.4655	0.5221	0.5386
0.6	0.95	11	0.2052	0.5526	1.2051	1.7232
0.6	0.90	7	-0.2244	0.6951	0.4595	1.6051

Table 2.6: Sample sizes and means for four populations, based on simulated data from Normal distributions with standard deviation $\sigma = 1$. Results are shown for varying values of δ^* and P^* according to Gupta's method

order to apply Bechhofer's method, the following steps are followed: Take n observations from each population. Then, compute the sample means for each population $\bar{x}_{[1]}, \ldots, \bar{x}_{[4]}$. Note that $\bar{x}_{[1]}$ denotes the smallest sample mean after ranking, while $\bar{x}_1$ refers to the sample mean of the first population. Rank the sample means as in Equation (1.2) from the smallest to the largest sample mean. Finally, the populations associated with $\bar{x}_{[4]}$ and $\bar{x}_{[3]}$ are the best and second best populations, respectively. The probability of making a correct selection of a ranked subset of the two best populations depends on n. Therefore, Bechhofer proposed designing an experiment in such a way that the distance between the best and second best populations means $\mu_{[4]} - \mu_{[3]} \geq \delta^*$ as in Equation (1.3). The probability of making a correct selection of a ranked subset of the two best populations is greater than or equal to a preassigned value P^*.

This section uses three values of P^* : $(0.95, 0.90, 0.75)$ representing the minimum probability of correctly selecting a ranked subset of two populations, and three corresponding values of δ^* : $(0.2, 0.4, 0.6)$ indicating the minimum difference between the means of the best and second best populations, as shown in Table 2.4.

Table 2.4 shows the mean population values corresponding to the three specified values of δ^*: 0.2, 0.4 and 0.6. The mean values in the table were chosen arbitrarily, with the primary consideration being to ensure that the distance between the largest and second largest means matches the assumed δ^*. The mean values were used in

$P^* = 0.95, \lambda = 2.9162$	R.H.S of Equation (1.7)	Selected sample means within subset
$n = 98 \quad \bar{x}_{[4]} = 0.7858$	0.4912	$\bar{x}_4$
$n = 25 \quad \bar{x}_{[4]} = 1.4147$	0.8315	$\bar{x}_3, \bar{x}_4$
$n = 11 \quad \bar{x}_{[4]} = 1.7232$	0.8439	$\bar{x}_3, \bar{x}_4$
$P^* = 0.90, \lambda = 2.4516$		
$n = 59 \quad \bar{x}_{[4]} = 0.9711$	0.6519	$\bar{x}_4$
$n = 15 \quad \bar{x}_{[4]} = 0.5386$	-0.0944	$\bar{x}_2, \bar{x}_3, \bar{x}_4$
$n = 7 \quad \bar{x}_{[4]} = 1.6051$	0.6788	$\bar{x}_4$

Table 2.7: λ values are given from tables provided by Gupta in [31], n values from and $\bar{x}_4$ from Table 2.6. The last column shows which sample means are greater than the corresponding R.H.S values of Equation (1.7)

the simulation study, where data were sampled from Normal distributions with these means and a standard deviation of $\sigma = 1$.

Using Equation (1.4) nine different sample sizes n are calculated as presented in Table 2.5. Table 2.5 provides the sample means and sample sizes of four populations calculated using simulated data generated from Normal distributions with nine different sample sizes n. Case 1 is defined by a case with $n = 237$, $P^* = 0.95$ and $\delta^* = 0.2$. Case 2 is defined by a case with $n = 174$, $P^* = 0.90$ and $\delta^* = 0.2$, and so on until Case 9 with $n = 10$, $P^* = 0.75$ and $\delta^* = 0.6$. According to Bechhofer's method, the sample means are ranked as in Equation (1.2) for all cases. First, in Case 1 the probability $P^* = 0.95$ was specified. This means that the probability of a correct selection of a ranked subset of the two best populations is at least 0.95 whenever the true difference between $\mu_{[4]} - \mu_{[3]}$ is at least 0.2, and in order to make a correct selection of the ranked subset of the two best populations, the required sample size is $n = 237$. Then, the sample means for this case are: 0.1947 for population 1, 0.2492 for population 2, 0.3387 for population 3, and 0.5208 for population 4. Using the Bechhofer method, the populations associated with $\bar{x}_{[4]}$ and $\bar{x}_{[3]}$ are the best and second best populations, respectively. Therefore, it can be concluded that there is 95% confidence in selecting population 4 as the best and population 3 as the second best in the ranked subset. For the rest of the cases, Table 2.5 shows

ℓ_1	ℓ_2	Case 1		Case 2		Case 3	
		$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
1	2	0.0555	0.0577	0.0590	0.0621	0.0542	0.0598
2	1	0.0592	0.0615	0.0563	0.0593	0.0515	0.0571
1	3	0.0611	0.0635	0.0739	0.0778	0.0569	0.0630
3	1	0.0677	0.0701	0.0893	0.0930	0.0647	0.0710
1	4	0.0705	0.0733	0.0695	0.0730	0.0750	0.0830
4	1	0.0916	0.0945	0.0706	0.0740	0.1158	0.1242
2	3	0.0759	0.0786	0.0783	0.0823	0.0542	0.0602
3	2	0.0786	0.0812	0.0989	0.1030	0.0617	0.0677
2	4	0.0876	0.0906	0.0729	0.0765	0.0712	0.0790
4	2	0.1067	0.1098	0.0791	0.0828	0.1118	0.1199
3	4	0.0998	0.1031	0.1152	0.1198	0.0868	0.0952
4	3	0.1174	0.1207	0.0985	0.1029	0.1232	0.1320

Table 2.8: The NPI lower and upper probabilities for all population selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, based on simulated data from Normal distributions with a common $\delta^* = 0.2$ and $\sigma = 1$. In Case 1, $P^* = 0.95$ and $n = 237$; in Case 2, $P^* = 0.90$ and $n = 174$; and in Case 3, $P^* = 0.75$ and $n = 90$.

that the population 4 is associated with the largest sample mean and the population 3 is associated with the second largest sample mean for cases 1, 3, 4, 6 and 9. In addition, population 3 is associated with the largest sample mean and population 4 is associated with the second largest sample mean for cases 2, 5, 7 and 8. The conclusion is not consistent across cases because random variation in the generated samples can lead to different sample means. Also, small sample sizes usually result in more variability in sample means, which may not match the true ranking of the population means.

Gupta [38] introduced a methodology for selecting a subset that contains the best populations. The goal is to select a subset that includes the populations with the largest sample means, regardless of their specific order. Gupta's method is applied to the same scenarios mentioned earlier by specifying the same values of P^*, δ^*, and

ℓ_1	ℓ_2	Case 4		Case 5		Case 6	
		$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
1	2	0.0526	0.0603	0.0228	0.0306	0.0259	0.0536
2	1	0.0285	0.0353	0.0337	0.0417	0.0368	0.0537
1	3	0.0640	0.0739	0.0431	0.0567	0.0326	0.0546
3	1	0.0541	0.0629	0.0817	0.0959	0.0642	0.0873
1	4	0.0602	0.0708	0.0391	0.0516	0.0479	0.0523
4	1	0.0730	0.0829	0.0672	0.0802	0.0949	0.1223
2	3	0.0687	0.0794	0.0666	0.0820	0.0459	0.0708
3	2	0.0847	0.0955	0.0915	0.1064	0.0623	0.0861
2	4	0.0606	0.0718	0.0621	0.0766	0.0403	0.0671
4	2	0.1045	0.1164	0.0674	0.0805	0.0970	0.1261
3	4	0.1058	0.1198	0.1457	0.1675	0.0786	0.1120
4	3	0.1343	0.1489	0.1319	0.1532	0.1154	0.1514

Table 2.9: The NPI lower and upper probabilities for all population selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, based on simulated data from Normal distributions with a common $\delta^* = 0.4$ and $\sigma = 1$. In Case 4, $P^* = 0.95$ and $n = 59$; in Case 5, $P^* = 0.90$ and $n = 43$; and in Case 6, $P^* = 0.75$ and $n = 22$.

σ as considered in Bechhofer's method. However, the tables provided in [31] do not include information for $P^* = 0.75$, which was considered in Bechhofer's method; thus, no conclusions can be drawn for this case.

Using Equation (1.5) six different sample sizes n are calculated as presented in Table 2.6. According to Gupta's procedure, populations with sample means $\bar{x}_{[i]}$ greater than or equal to $\bar{x}_{[4]} - \frac{\lambda}{\sqrt{n}}$ are selected, as stated in Equation (1.7). Table 2.7 shows that populations 3 and 4 are included in the selected subset when the sample sizes are 25 and 11. Furthermore, populations 2, 3, and 4 are included in the selected subset when the sample size is 15. The selected subset contains only one population, which is 4 with the largest mean when the sample sizes are 98, 59, and 7. In conclusion, Gupta's method does not always yield the same selected subset as Bechhofer's method.

ℓ_1	ℓ_2	Case 7		Case 8		Case 9	
		$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
1	2	0.0109	0.0184	0.0089	0.0189	0.0022	0.0155
2	1	0.0088	0.0162	0.0079	0.0198	0.0064	0.0228
1	3	0.0370	0.0588	0.0546	0.0881	0.0265	0.0669
3	1	0.0536	0.0722	0.0750	0.1093	0.0265	0.0663
1	4	0.0343	0.0547	0.0370	0.0636	0.0212	0.0771
4	1	0.0504	0.0690	0.0425	0.0697	0.0766	0.1359
2	3	0.0238	0.0445	0.0269	0.0541	0.0141	0.0455
3	2	0.0783	0.1016	0.0606	0.0872	0.0152	0.0445
2	4	0.0316	0.0522	0.0205	0.0419	0.0113	0.0589
4	2	0.0660	0.0882	0.0283	0.0482	0.0367	0.0751
3	4	0.2026	0.2463	0.1779	0.2309	0.1073	0.1948
4	3	0.1726	0.2135	0.1580	0.2104	0.1546	0.2457

Table 2.10: The NPI lower and upper probabilities for all population selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, based on simulated data from Normal distributions with a common $\delta^* = 0.6$ and $\sigma = 1$. In Case 7, $P^* = 0.95$ and $n = 26$; in Case 8, $P^* = 0.90$ and $n = 19$; and in Case 9, $P^* = 0.75$ and $n = 10$.

Now, the NPI method for selecting a ranked subset of the best two populations applies for the same sample sizes presented in Table 2.5 and mean values presented in Table 2.4. Also, the data were generated from Normal distributions and nine different cases are considered.

Tables 2.8, 2.9 and 2.10 show the NPI lower and upper probabilities for twelve combinations of selecting a ranked subset of two populations out of four. Generally, the NPI lower and upper probabilities were found to be small. When the sample size is large, $n = 237$, the lowest NPI lower probability corresponding to the selection of population 1 as the best and population 2 as the second best is 0.0555. Meanwhile, the highest NPI upper probability associated with selecting population 4 as the best and population 3 as the second best is 0.1207. In contrast, when the sample size is small, $n = 10$, the lowest NPI lower probability, related to selecting population 1 as

ℓ_1	ℓ_2	Case 1		Case 2		Case 3	
		$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$
1	2	0	0	0	0	0	0
2	1	0	0	0	0	0	0
1	3	0	0	0	0	0	0
3	1	0	0	0	0	0	0
1	4	0	0	0	0	0	0
4	1	0	0	0	2	9	12
2	3	0	0	0	0	0	0
3	2	0	0	0	0	1	4
2	4	0	0	0	0	8	4
4	2	39	37	72	85	127	116
3	4	102	113	133	129	176	170
4	3	859	850	795	784	679	694

Table 2.11: Frequency of the NPI lower and upper probabilities for all population selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$ after repeating the simulation 1000 times with a common $\delta^* = 0.2$ and $\sigma = 1$. In Case 1, $P^* = 0.95$ and $n = 237$; in Case 2, $P^* = 0.90$ and $n = 174$; and in Case 3, $P^* = 0.75$ and $n = 90$

the best and population 2 as the second best, is 0.0022, and the highest NPI upper probability, related to selecting population 4 as the best and population 3 as the second best, is 0.2457. This shows that imprecision, which is defined as the difference between the NPI lower and upper probabilities across all tables, is influenced by the sample size. When the sample size is small, the imprecision is larger. In contrast, when the sample size is large, the imprecision is small.

Generally, selecting either population 3 as the best and population 4 as the second best, or population 4 as the best and population 3 as the second best, results in a good selection. Across all cases, these combinations consistently have the largest or second largest NPI lower and upper probabilities. In contrast, selecting either population 1 as the best and population 2 as the second best, or population 2 as the best and population 1 as the second best, results in a poor selection. Across all

ℓ_1	ℓ_2	Case 4		Case 5		Case 6	
		$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$
1	2	0	0	0	0	0	0
2	1	0	0	0	0	0	0
1	3	0	0	0	0	0	0
3	1	0	0	0	0	0	0
1	4	0	0	0	0	0	1
4	1	0	0	3	0	2	5
2	3	0	0	0	0	0	0
3	2	0	0	0	0	0	0
2	4	0	0	1	0	0	1
4	2	37	36	62	54	99	102
3	4	59	72	102	119	155	176
4	3	904	892	832	827	744	715

Table 2.12: Frequency of the lower and upper probabilities for all population selections combinations for the event $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}$ after repeating the simulation 1000 times with a common $\delta^* = 0.4$ and $\sigma = 1$. In Case 4, $P^* = 0.95$ and $n = 59$; in Case 5, $P^* = 0.90$ and $n = 43$; and in Case 6, $P^* = 0.75$ and $n = 22$

cases, these combinations consistently have the lowest or second lowest NPI lower and upper probabilities. This is because, by the design of the simulation study, populations 1 and 2 have smaller means compared to populations 3 and 4, as shown in Table 2.4.

It should be noted that in Bechhofer's method, P^* is the desired probability of correct selection, this value is specified by the experiment, and it determines the value of d. The constant d is then used to calculate the required sample size. In contrast, NPI does not require specifying P^* in advance. Instead, NPI provides lower and upper probabilities for the events of interest, the uncertainty found directly from the data.

Tables 2.11, 2.12 and 2.13 display the frequencies of the NPI lower and upper probabilities for all possible combinations across 1000 simulation repetitions. Re-

ℓ_1	ℓ_2	Case 7		Case 8		Case 9	
		$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$	$\#\underline{P}$	$\#\overline{P}$
1	2	0	0	0	0	0	0
2	1	0	0	0	0	0	0
1	3	0	0	0	0	0	0
3	1	0	0	0	0	0	0
1	4	0	0	0	0	0	0
4	1	0	1	0	0	4	4
2	3	0	0	0	0	0	0
3	2	0	0	0	0	1	0
2	4	0	0	0	1	5	5
4	2	26	23	49	46	95	81
3	4	44	44	66	71	128	124
4	3	930	932	885	882	767	786

Table 2.13: Frequency of the lower and upper probabilities for all population selections combinations for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$ after repeating the simulation 1000 times with a common $\delta^* = 0.6$ and $\sigma = 1$. In Case 7, $P^* = 0.95$ and $n = 26$; in Case 8, $P^* = 0.90$ and $n = 19$; and in Case 9, $P^* = 0.75$ and $n = 10$

peating this simulation thousands of times for all combinations ensures that the method works well in different situations presented in Tables 2.11, 2.12 and 2.13. Once again, the findings confirm the previous conclusion, clearly indicating that population 4 is selected as the best, followed by population 3 as the second best across all tables. For example, for Case 1 in Table 2.11, these two populations appear with frequencies of 859 out of 1000 for the NPI lower probability, and 850 out of 1000 for the NPI upper probability. In addition, it was observed that as the sample sizes increased, the variability in the selected combinations decreased within each table. Specifically, Case 3 shows a greater variation compared to Cases 1 and 2 in Table 2.11; similarly, Case 6 has more variation than Cases 4 and 5 in Table 2.12, and this pattern continues in Table 2.13. Larger sample sizes result in more consistent selection, reducing variation among less frequently selected combinations.

Most of these combinations have frequencies of zero, indicating that they are never selected.

In conclusion, the indifference zone approach introduced by Bechhofer and the NPI method for selecting a ranked subset of the two best populations yield both the same selected subset and the same ranking within the subset. However, Gupta's method does not always yields the same selected subset as Bechhofer's method or the NPI method when selecting a ranked subset of the two best populations.

2.7 Concluding remarks

This chapter has presented a method for selecting a ranked subset of future observations, resulting in exact and approximate NPI lower and upper probabilities. There are many scenarios of interest. In the first scenario, when selecting a ranked subset of the two best groups, an optimisation is performed, and the exact NPI lower and upper probabilities are derived. In this scenario, determining the exact NPI lower and upper probabilities becomes easy as there is only one group in the middle. Even with large and overlapping observations, obtaining exact NPI lower and upper probabilities remains possible.

In addition, lower and upper bounds for the NPI lower and upper probabilities are derived to simplify the calculations. A real-valued dataset demonstrates that the exact NPI lower and upper probabilities lie within their respective lower and upper bounds. The example shows that the exact NPI lower probability is close to its corresponding lower bound, and the exact NPI upper probability is close to its upper bound.

For further investigation of the NPI method for the selection of a ranked subset of the two best groups, a simulation study was conducted where the means were assumed and arranged from smallest to largest. The results from the simulation study conclude that selecting the group with the largest mean as the best and selecting the groups with the smallest means as the worst is a good decision.

Some findings were made to help improve this method and reduce the computational burden of calculating the exact NPI lower and upper probabilities. It is

evident that the first sub-interval within the first interval is most likely to yield the minimum NPI lower probability, as it often has zero assigned probability mass. Similarly, the last sub-interval within the last interval tends to produce the same result. However, for the intervals in between, the minimum number of assigned probability mas typically comes from the initial sub-interval of each interval. On the other hand, regarding the NPI upper probability, the last sub-interval within the first interval is most likely to yield the maximum NPI upper probability, as it contains the highest number of assigned probability mass. Similarly, the first sub-interval within the last interval tends to produce the same result. However, for the intervals in between, the maximum number of assigned probability mass mostly comes from the ending sub-intervals of each interval.

The second scenario arises when the number of best groups within a subset is three or more. In this case, it is computationally expensive to optimise over all groups at the same time. Therefore, an optimisation process and a heuristic algorithm should be performed for each group in the middle independently, and then approximate NPI lower and upper probabilities are obtained. The approximate NPI lower probability is calculated as the smallest value among the NPI lower probabilities for the middle groups in the ranked subset. Likewise, the approximate NPI upper probability is obtained as the largest value among the NPI upper probabilities for those same groups. In addition, the lower and upper bounds for the NPI lower and upper probabilities are derived to avoid the calculation being cumbersome.

Moreover, a partial ranking within subset groups is presented in this chapter. Partial ranking is a simpler and more practical approach when comparing or ranking groups based on multiple criteria, especially when a complete ranking is difficult or unnecessary. In this section, a brief overview is provided, focusing on the formulas of the bounds for the NPI lower and upper probabilities.

Finally, a comparison was made between the NPI method for selecting a ranked subset of the two best groups, the Bechhofer and Gupta methods from the literature. The Bechhofer and NPI methods for selecting a ranked subset of the two best populations yield both the same selected subset and the same ranking within the subset. However, Gupta's method does not always yield the same selected subset as

Bechhofer's method or the NPI method when selecting a ranked subset of the two best populations.

Chapter 3

NPI for selection using loss functions

3.1 Introduction

Decision theory provides a systematic approach to guide decision makers when faced with uncertainty by quantifying potential losses based on the outcomes of the selected groups. A loss function represents the penalty incurred when selecting sub-optimal groups instead of the best possible ones. The worse the decision, the higher the loss. A decision should aim to minimise the expected loss to improve the quality of the decision. A brief overview of statistical decision theory is provided in Section 1.3.

In classical statistics, a loss function is used for parameter estimation by specifying the penalty as a function of the difference between the estimated and true parameters [72]. Loss functions can be applied in many areas, each with its own specific interpretation. In the context of classification, a loss function measures misclassification, for example, assigning a value of 0 for correct classifications and 1 for incorrect ones. In economics, loss functions are applied to evaluate point forecasts of financial returns [49]. In actuarial science, loss functions are used to quantify risk and guide decisions about premiums and reserves in insurance [21]. In optimal control, loss functions are used to quantify the penalty for failing to achieve a desired value. Finally, in process safety assessment, the loss functions describe the relationship between process deviations and system loss [45]. Whether it is minimising shipping costs, making medical diagnoses, or optimising financial investments, understanding

the potential losses allows informed decision-making. The use of loss functions is fundamental in decision-making and statistical modeling, providing a quantitative measure of the consequences associated with different choices or predictions.

In the literature, many loss functions have been used. A commonly used loss function in classification is the zero-one loss function. Although simple and intuitive, this loss function treats all misclassifications equally, regardless of the degree of error. In some situations, the zero-one loss function is replaced with alternatives such as hinge loss, which provides a smoother transition between correct and incorrect predictions. More complex loss functions, such as the quadratic loss function, measure the squared difference between predicted and actual values, penalising larger errors more severely [72].

This chapter considers the use of loss functions in NPI for selection. Section 3.2 reviews the loss functions used in subset selection as presented in the literature. Section 3.3 introduces NPI decision theory, with the NPI lower and upper expected losses as the key components for making decisions. Section 3.4 presents pairwise comparisons based on loss functions. Section 3.5 introduces multiple comparisons using loss functions and includes two illustrative examples, one with real-valued and another with simulated data to demonstrate their application. Finally, Section 3.6 presents some concluding remarks.

3.2   Loss functions in selection

In the context of selection, a loss function measures the penalty associated with selecting suboptimal groups in the selection process. The goal is to minimise the expected loss associated with the selection of suboptimal groups. This section provides an overview of loss functions commonly used in the subset selection literature, while the primary focus of this thesis is developing the NPI approach for selection and ranking problems using such loss functions.

Various methodologies have explored the application of loss functions in subset selection. Gupta's subset selection rule, originally proposed by Gupta and Panchapakesan [41], serves as a foundational approach. It defines decision rules and eval-

uates their performance using loss functions, with an emphasis on minimising losses associated with incorrect selection. Another consideration in subset selection is the minimax criterion, investigated by Berger and Gupta [6]. Berger and Gupta propose a selection rule that minimises the maximum risk, which corresponds to maximising the minimum probability that the selected subset includes the best population. Hsu [50] introduced an alternative approach specifically tailored for location parameter families. This method not only provides confidence intervals, but also simultaneously addresses all distances from the best population. It compares each population with the best and uses confidence intervals for the differences between the mean of each population and the best mean. These intervals help determine which populations are not significantly worse than the best and which ones can be excluded from the set of the best population. The works of Hsu [51, 52] and Finner and Giani [29] further support this alternative perspective by establishing connections between multiple comparisons and simultaneous confidence intervals and by extending multiple comparison procedures to more complex settings such as block designs. The loss functions mentioned above are not relevant to this thesis, but could be considered for future work.

In addition, various loss functions can be employed in the selection process, including zero-one loss, linear and quadratic loss functions. The choice of a particular loss function depends on the nature of the decision problem and the available information. Below is a brief overview of several loss functions discussed in the literature.

3.2.1 Zero-one loss function

In the context of subset selection procedures for comparing k populations, each assumed to follow a specific parametric distribution, the zero-one loss function is denoted as $L_{\{0,1\}}(S, \mu^*)$. Here, μ^* represents the true vector of parameter values associated with the k populations, and S is the selected subset of the best populations, where $S \subseteq \{1, \ldots, k\}$.

For simplicity, we assume that each population is characterized by a one-dimensional parameter μ_i, such that for Normally distributed populations $X_i \sim N(\mu_i, \sigma^2)$, the variance σ^2 is not considered in the loss function. The true vector of the parameters

is given by $\mu^* = \{\mu_{[1]}, \ldots, \mu_{[k]}\}$, where $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}$. Specifically, $\mu_{[k]}$ is the largest parameter. A subset S is selected based on a predefined selection rule. The parameter $\mu_{[q]}$ is the parameter of the q-th population included in S, where q indexes the position of that population in the ordered list, and q refers to any population in S. Since the true ranking of populations is unknown in practice, selection methods rely on observed data to approximate the ranking and determine S.

Deely and Gupta [23] introduced the zero-one loss function for selecting a subset that contains the best population, which is defined as

$$L_{\{0,1\}}(S, \mu^*) = \begin{cases} 0 & \text{if } \mu_{[k]} \text{ is in } S \\ 1 & \text{if otherwise} \end{cases} \tag{3.1}$$

This loss function quantifies the penalty incurred when the selected subset S does not include the population with the largest parameter $\mu_{[k]}$, given the true parameter vector μ^*.

3.2.2 Linear loss function

The linear loss function is a common choice in decision theory. It assumes that the cost of selecting suboptimal populations is represented by a linear function of the difference between the largest population parameter $\mu_{[k]}$ and the parameter of each selected population in S $\mu_{[q]}$ [23]. The linear loss function is given by

$$L_l(S, \mu^*) = \sum_{q \in S} c_q(\mu_{[k]} - \mu_{[q]}) \tag{3.2}$$

where $c_q > 0$ serves as a constant that indicates the penalty weight to include population q in the selected subset. This loss function is used to quantify the loss associated with selecting a subset S that does not contain the largest population parameter $\mu_{[k]}$.

3.2.3 Quadratic loss function

The quadratic loss function [32], also known as the squared loss, is defined by

$$L_Q(S, \mu^*) = \sum_{q \in S} c_q(\mu_{[k]} - \mu_{[q]})^2 \tag{3.3}$$

This loss function is employed to measure the quadratic loss incurred by selecting a subset S that does not contain the largest population parameter, $\mu_{[k]}$.

The quadratic loss function can be seen as an extension of the linear loss function, as it introduces a squared term. Linear and quadratic loss functions often lead to similar conclusions [5]. Therefore, the aim is to investigate whether all three loss functions zero-one, linear and quadratic lead to the same or different conclusions.

3.2.4 Other loss functions

The average loss function quantifies the impact of selecting suboptimal populations and helps guide the selection of the subset of the best populations [23].

$$L_a(S, \mu^*) = \frac{1}{|S|} \sum_{q \in S} (\mu_{[k]} - \mu_{[q]}) \tag{3.4}$$

where $|S|$ denotes the number of populations in the subset S. This loss function represents the average loss for each population in the subset S. Specifically, it calculates the difference between the largest parameter $\mu_{[k]}$ and each selected parameter of the q-th population and then divides this difference by the size of the subset $|S|$.

Goel and Rubin [33] introduced a linear loss function that combines two essential components. First, it considers the size of the selected subset $|S|$, denoted as

$$L_{c_1}(S, \mu^*) = c|S|$$

where c is a fixed number, $c > 0$. Second, it incorporates the difference between the largest true parameter value $\mu_{[k]}$ and the largest parameter value within the selected subset S. This component is expressed as

$$L_{c_2}(S, \mu^*) = \mu_{[k]} - \max_{q \in S}(\mu_{[q]})$$

This loss ensures that only positive differences contribute to the loss, penalising the selection of a subset that does not include the best possible result represented by $\mu_{[k]}$. The general linear loss function is given by

$$L_{c_3}(S, \mu^*) = c|S| + \mu_{[k]} - \max_{q \in S}(\mu_{[q]})$$

October 6, 2025

This formulation captures the trade-off between the size of the selected subset and the difference between the best population and the best within the selected subset.

Gupta and Miescke [40] utilized another combined loss function with two variations, denoted L_1 and L_2, each designed to address different considerations. $L_1(S, \mu^*)$ encourages the inclusion of all populations with the largest parameters. It is defined as

$$L_1(S, \mu^*) = c|S| - \sum_{q \in S} I_{\{\mu_{[k]}\}}(\mu_{[q]}),$$

where the summation ensures that the indicator function is evaluated for all selected populations in S. The indicator function $I_{\{\mu_{[k]}\}}(\mu_{[q]})$ checks whether the parameter $\mu_{[q]}$ of a selected population q is equal to the largest population parameter $\mu_{[k]}$. If it is, it contributes 1 to the sum; otherwise, it contributes 0.

In contrast, $L_2(S, \mu^*)$ rewards the inclusion of at least one population with the largest parameter. It is defined as

$$L_2(S, \mu^*) = c|S| - I_{\{\mu_{[k]}\}}\left(\max_{q \in S}(\mu_{[q]}) \right),$$

where the maximum is taken over the selected subset S.

L_2 is a modified version of L_1 that accounts for the number of populations that are tied for the best. The loss L_2 combines the zero-one loss for including the best population with the loss of including other populations. Note that the inclusion of a single best population is penalised under L_1, when there is more than one best population, but not under L_2.

Lehmann [55, 56] introduced an additive loss function. In subset selection, populations are classified into selected populations in S and non-selected populations in NS, and the loss is defined as

$$L_d(S, \mu^*) = \sum_{q \in S} L_A(S, \mu^*) I_{NS}(q) + \sum_{q \in S} L_B(S, \mu^*) I_S(q) \tag{3.5}$$

where $I_{NS}(q)$ is an indicator function that equals 1 if population q belongs to the non-selected population NS, and 0 otherwise, while $I_S(q)$ is an indicator function that equals 1 if population q belongs to the selected population S, and 0 otherwise.

$L_A(S, \mu^*)$ represents the loss of not selecting population q when it has one of the largest parameter values, while $L_B(S, \mu^*)$ represents the loss of selecting population q when it does not have a large parameter value

3.2.5 Expected loss function

The risk function of a decision, also known as the expected loss function [73], is introduced to evaluate the quality of decisions under uncertainty. Loss functions quantify the consequences of a decision in a given situation, providing a measure of its effectiveness. Risk functions serve multiple purposes. Primarily, they guide the minimisation of expected loss, aiming to identify decision rules that, on average, lead to the least negative outcomes. Additionally, risk functions enable the comparison of different decision strategies, which is crucial since certain strategies may perform well in some contexts but poorly in others.

Consider the scenario where the aim is to predict an outcome based on available data. Assume that $L(\mu^*, d) \in \mathbb{R}$ is the loss function, measuring the penalty for choosing the decision d when the true vector of parameters is μ^*. The expected loss is given by

$$R(\mu^*, d) = E\left[L(\mu^*, d(X))\right] = \sum_X L(\mu^*, d(X)) \, P(X \mid \mu^*) \tag{3.6}$$

here, R denotes a risk function, $L(\mu^*, d(X))$ The loss incurred by applying the decision d to the data X, and $P(X \mid \mu^*)$ is the probability density function of observing X given μ^*.

3.3 NPI decision theory

After introducing expected loss functions from the literature, it is important to explain how these ideas are specifically applied within the NPI framework. Understanding decision theory in the context of NPI provides valuable insights into evaluating decisions based on future observations.

NPI decision theory is introduced using the NPI lower and upper expected losses, which quantify the impact of selecting suboptimal groups and guide the selection of

the best groups.

Assume that $L(X^f, d) \in \mathbb{R}$ represents the loss associated with a decision d and a future observation X^f. Let X^f denote a single future observation from group X. Suppose that $X_1, \ldots, X_{n_x}, X_{n_x+1}$ are real-valued, continuous, and exchangeable random quantities from X. Let its ordered observed values be $x_1 < x_2 < \cdots < x_{n_x}$, with $x_0 = -\infty$ and $x_{n_x+1} = \infty$ defined for ease of notation. These n_x observations divide the real line into n_x+1 intervals $I_i = (x_{i-1}, x_i)$, where $i = 1, 2, \ldots, n_x+1$. The NPI lower expected loss $\underline{E}(L(X^f, d))$ for the future observation X^f and a decision d is defined as

$$\underline{E}(L(X^f, d)) = \inf_{X^f \in I_i} \sum_{i=1}^{n_x+1} L(x_i, d) P(X^f \in I_i) \tag{3.7}$$

where $X^f \in I_i$ is defined by assigning the probability mass $\frac{1}{n_x+1}$ to each interval I_i. Then, $\inf_{X^f \in I_i}$ is taken over each interval I_i. The use of the infimum guarantees that the calculation accounts for the minimum expected loss across all intervals I_i.

The NPI upper expected loss for the future observation X^f and a decision d is defined as

$$\overline{E}(L(X^f, d)) = \sup_{X^f \in I_i} \sum_{i=1}^{n_x+1} L(x_i, d) P(X^f \in I_i) \tag{3.8}$$

This equation is determined by assigning the probability mass $\dfrac{1}{n_x + 1}$ to each interval I_i. Then, the supremum takes over each interval I_i. The use of the supremum guarantees that the calculation accounts for the maximum expected loss across all intervals I_i.

The NPI lower and upper expected losses are calculated for pairwise and multiple comparisons in Sections 3.4 and 3.5, respectively.

3.4 Pairwise comparisons based on loss functions

Suppose there are two groups X and Y and the goal is to select the best group and quantify the impact of choosing a suboptimal group. Let group X have n_x observations, and group Y have n_y observations. Let X^f denote a single future observation from group X, and let Y^f denote a single future observation from group

Y. Assume that groups X and Y are independent, which means that the data from group X does not provide any information about the data from group Y. Suppose the event of interest is $X^f \geq Y^f$, where the objective is to compare X^f and Y^f using loss functions and to select either group X or Y based on which yields the minimum expected loss.

In the following sections, pairwise comparisons are conducted using three loss functions—zero-one, linear, and quadratic—to examine whether they lead to the same decision during the selection process.

3.4.1 Zero-one loss function

In the literature, the common loss function for selection has been the zero–one loss function [58]. The zero-one loss function for the pairwise comparison for the event $X^f \geq Y^f$ is considered as follows:

$$
L_{\{0,1\}}(X^f, Y^f) = \begin{cases} 0 & \text{if } X^f \geq Y^f \\ 1 & \text{if } X^f < Y^f \end{cases} \tag{3.9}
$$

Suppose $X_1, \ldots, X_{n_x+1}$ and $Y_1, \ldots, Y_{n_y+1}$ are real-valued, exchangeable, and continuous random quantities of X and Y, respectively. Let group X have n_x observations and group Y have n_y observations. Their ordered observed values are given by $l_x < x_1 < \cdots < x_{n_x} < r_x$ and $l_y < y_1 < \cdots < y_{n_y} < r_y$, where $x_0 = l_x$, $x_{n_x+1} = r_x$ are defined for ease of notation, and $y_0 = l_y$, $y_{n_y+1} = r_y$. Inference is based on the $A_{(n_x)}$ assumption for X^f and on the $A_{(n_y)}$ assumption for Y^f. This is combined with the zero-one loss function to quantify the loss by comparing the future observation from group X with that from group Y. In other words, the zero-one loss function only penalises when $X^f < Y^f$.

The expected loss is defined as the sum of all possible losses, each multiplied by its corresponding probability mass. The expected value of the loss function used to compare groups X and Y is denoted by $E(L(X^f, Y^f))$. This expected loss is

$$
E(L(X^f, Y^f)) = \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L(X^f, Y^f) P(X^f \in (x_{i-1}, x_i)) P(Y^f \in (y_{j-1}, y_j)) \tag{3.10}
$$

$A_{(n)}$ is not sufficient to derive a precise expected loss for many events of interest, but optimal bounds based on $A_{(n)}$ can be derived for all events of interest. Therefore, the NPI lower and upper expected losses should be derived. The maximum lower bound for the expected value of the zero-one loss, as given by Equation (3.11), is derived by summing the probabilities based on $A_{(n_x)}$ and $A_{(n_y)}$ for which $x_i < y_{j-1}$. This results from assigning the probability mass of X to the right endpoint per interval and of Y to the left endpoint per interval to minimise the NPI lower expected zero-one loss. The NPI lower expected zero-one loss is

$$\underline{E}(L_{\{0,1\}}(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_{\{0,1\}}(x_i, y_{j-1}) \qquad (3.11)$$

where $L_{\{0,1\}}(x_i, y_{j-1})$ is obtained from Equation (3.9).

The minimum upper bound for the expected value of the zero-one loss, as given by Equation (3.12), is derived by summing the probabilities based on $A_{(n_x)}$ and $A_{(n_y)}$ for which $x_{i-1} < y_j$. This results from assigning the probability mass of X to the left endpoint per interval and of Y to the right endpoint per interval, to maximise the NPI expected zero-one loss. The NPI upper expected zero-one loss is

$$\overline{E}(L_{\{0,1\}}(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_{\{0,1\}}(x_{i-1}, y_j) \qquad (3.12)$$

A direct relationship exists between the NPI lower and upper expected zero-one losses and the NPI lower and upper probabilities. Specifically, the NPI lower expected zero-one loss for the event $X^f \geq Y^f$ is equal to the NPI lower probability of the complementary event $X^f < Y^f$.

$$\underline{E}(L_{\{0,1\}}(X^f, Y^f)) = \underline{P}(X^f < Y^f) = 1 - \overline{P}(X^f \geq Y^f) \qquad (3.13)$$

Also, the NPI upper expected zero-one loss for the event $X^f \geq Y^f$ is equal to the NPI upper probability of the complementary event $X^f < Y^f$.

$$\overline{E}(L_{\{0,1\}}(X^f, Y^f)) = \overline{P}(X^f < Y^f) = 1 - \underline{P}(X^f \geq Y^f) \qquad (3.14)$$

Proof: Using the conjugacy property, $\underline{P}(A^c) = 1 - \overline{P}(A)$, for an event A and its complement event A^c. The proof showing that the NPI lower expected zero-one

loss for the event $X^f \geq Y^f$ is equal to the NPI lower probability of the complement event $X^f < Y^f$ is as follows.

By applying the definition of the zero-one loss function, we have

$$
\begin{aligned}
\underline{E}(L_{\{0,1\}}(X^f,Y^f)) = &(L_{\{0,1\}}(X^f,Y^f) \mid X^f \geq Y^f)\underline{P}(X^f \geq Y^f) \\
&+ (L_{\{0,1\}}(X^f,Y^f) \mid X^f < Y^f)\underline{P}(X^f < Y^f)
\end{aligned}
$$

Since the zero-one loss function takes the value 0 when $X^f \geq Y^f$, and 1 when $X^f < Y^f$, the expected loss simplifies to the probability that $X^f < Y^f$:

$$
\underline{E}(L_{\{0,1\}}(X^f,Y^f)) = \underline{P}(X^f < Y^f)
$$

The proof of the NPI upper expected zero-one loss follows the same reasoning as that of the NPI lower expected zero-one loss, with the NPI lower expected zero-one loss and NPI lower probability replaced by their upper counterparts. To avoid repetition, the proof is omitted. □

3.4.2   Linear loss function

In this section, the linear loss function is considered for the event $X^f \geq Y^f$, as discussed in Section 3.4. The linear loss function in pairwise comparison is defined as

$$
L_l(X^f,Y^f) = \begin{cases} 0 & \text{if } X^f \geq Y^f \\ Y^f - X^f & \text{if } X^f < Y^f \end{cases} \tag{3.15}
$$

The NPI lower expected linear loss is derived by assigning the probability mass of X^f to the right endpoint per interval, and the probability mass of Y^f to the left endpoint per interval. For each pair, the NPI lower expected linear loss is then computed as the difference $y_{j-1} - x_i$ whenever $x_i < y_{j-1}$, and zero otherwise. Similarly, the NPI upper expected linear loss is obtained by assigning the probability mass of X^f to the left endpoint per interval, while the probability mass of Y^f is assigned to the right endpoint per interval. For each pair, the NPI upper expected linear loss is then computed as the difference $y_j - x_{i-1}$ whenever $x_{i-1} < y_j$. The

NPI lower and upper expected linear losses are given by

$$\underline{E}(L_l(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_l(x_i, y_{j-1}) \qquad (3.16)$$

$$\overline{E}(L_l(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_l(x_{i-1}, y_j) \qquad (3.17)$$

When considering the linear loss function, the assumption is made that X^f and Y^f have a finite range. The presence of infinite lower and upper bounds, such as $x_0 = -\infty$ or $x_{n+1} = \infty$, can affect the calculation of the NPI lower and upper expected linear losses. For clarity, the presence of an infinite value will overpower the sum, leading to an overall expected loss that is infinite and thus not meaningful for decision-making.

3.4.3 Quadratic loss function

Another commonly used loss function is the quadratic loss function [73]. The quadratic loss function is defined as a measure of the deviation between future observations from groups X and Y. The definition of the quadratic loss function in pairwise comparison is

$$L_q(X^f, Y^f) = \begin{cases} 0 & \text{if } X^f \geq Y^f \\ (Y^f - X^f)^2 & \text{if } X^f < Y^f \end{cases} \qquad (3.18)$$

Based on assumptions $A_{(n_x)}$ for X^f and $A_{(n_y)}$ for Y^f, the NPI lower and upper expected quadratic losses are

$$\underline{E}(L_q(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_q(x_i, y_{j-1}) \qquad (3.19)$$

$$\overline{E}(L_q(X^f, Y^f)) = \frac{1}{(n_x + 1)} \frac{1}{(n_y + 1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} L_q(x_{i-1}, y_j) \qquad (3.20)$$

Regarding the earlier assumption that the future observations are bounded in the linear loss function, this consideration is extended to the quadratic loss function as well.

Group	Data					
B	2625	2628	2795	2847	2925	2968
	2975	3163	3176	3292	3421	3473
G	2412	2539	2729	2754	2817	2875
	2935	2991	3126	3210	3231	3317

Table 3.1: Birth weights (g) for Example 3.4.1.

Example 3.4.1 To illustrate the NPI lower and upper expected losses for pairwise comparison based on different loss functions, the dataset on the birth weights of 12 males (B) and 12 females (G) is used [26]. The original data also included estimated gestational ages, a known influential factor in birth weight. The dataset is presented in Table 3.1.

When making decisions about the selection of the best group, the loss incurred is measured in the event that the weight of the next baby boy, B^f, is greater than the weight of the next baby girl, G^f. When considering the zero-one loss function, the NPI lower and upper expected zero-one losses can be calculated directly from Equations (3.11) and (3.12), which yield $\underline{E}(L_{\{0,1\}}(B^f, G^f)) = 0.3432$ and $\overline{E}(L_{\{0,1\}}(B^f, G^f)) = 0.4911$. For the complement event $(B^f \geq G^f)^c$, the weight of the next baby boy is less than the weight of the next baby girl. Then, the NPI lower and upper expected zero-one losses for the complement event $(B^f \geq G^f)^c$ are $\underline{E}(L_{\{0,1\}}(B^f, G^f)) = 0.5090$ and $\overline{E}(L_{\{0,1\}}(B^f, G^f)) = 0.6570$.

The results show that the NPI lower and upper expected zero-one losses for the complement event $(B^f \geq G^f)^c$ are higher than those for the event $B^f \geq G^f$. Therefore, claiming that $B^f < G^f$ is more likely to result in an incorrect pairwise comparison.

For the linear and quadratic loss functions, it is assumed that future observations are bounded by l and r. Since the observed values range from 2412 to 3473 grams, we set $l = 2400$ and $r = 3500$ to define the lower and upper bounds for future observations.

The NPI lower and upper expected linear losses associated with predicting that the weight of the next baby boy B^f is greater than that of the next baby girl G^f are

89.6864 and 166.3018, respectively. In contrast, for the prediction of $B^f < G^f$, the NPI lower and upper expected linear losses are 185.6864 and 278.3018, respectively. Since the smallest loss values indicate better decisions, favouring the prediction that the weight of the next baby boy is greater than the weight of the next baby girl leads to a more optimal decision.

Similarly, the NPI lower and upper expected quadratic losses can be calculated using Equations (3.19) and (3.20) leading to the following results: The NPI lower and upper expected quadratic losses for the event $B^f \geq G^f$ are $\underline{E}(L_q(B^f, G^f)) = 35213.07$ and $\overline{E}(L_q(B^f, G^f)) = 87403.39$. The NPI lower and upper expected quadratic losses for the complement event $(B^f \geq G^f)^c$ are 100126.8 and 169917.1 respectively. The NPI lower and upper expected quadratic losses for $B^f \geq G^f$ implies a more optimistic scenario with less deviation compared to the complement event, because its corresponding NPI lower and upper expected quadratic losses are low compared to the complement event. Therefore, the weight of the next baby boy is expected to be higher than that of the next baby girl.

In conclusion, after employing zero-one, linear, and quadratic loss functions to evaluate the prediction of the weight of the next baby boy relative to the next baby girl, all methods consistently support the same optimal decision. Regardless of the loss function used, the prediction that the weight of the next baby boy will be larger than that of the next baby girl emerges as the optimal choice.

$\diamond$

3.5 Multiple comparisons based on loss functions

In this section, the comparison is extended to multiple comparisons based on loss functions for $k \geq 2$ independent groups. Multiple comparisons involve three different types of events, namely, selecting the best group, selecting the subset of the best groups, and selecting the subset that includes the best group as presented by Coolen and van der Laan [18], also reviewed in Section 1.5.

Suppose there are $k \geq 2$ groups and $n_j + 1$ random quantities from group j denoted by X_{j,i_j} where $j = 1, 2, \ldots, k$ and $i_j = 1, 2, \ldots, n_j + 1$ and their ordered

observation values are $x_{j,1} < x_{j,2} < \ldots < x_{j,n_j}$ where $x_{j,0} = l_j$ and $x_{j,n_j+1} = r_j$ are defined for ease of notation. Then, inference is based on Hill's assumption $A_{(n_j)}$ for each group. In the following three sections, the impact of selecting a suboptimal groups is measured using zero-one, linear, and quadratic loss functions.

3.5.1 Zero-one loss function

First, selecting the best group is considered. This means that the focus is on the event that a specific future observation X_ℓ^f is the maximum of all future observations X_j^f, where $j = 1, 2, \ldots, k$. The definition of the zero-one loss for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ is

$$L_{\{0,1\}}(X_\ell^f, \max_{j \neq \ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j \neq \ell} X_j^f \\ 1 & \text{if } X_\ell^f < \max_{j \neq \ell} X_j^f \end{cases} \tag{3.21}$$

The penalty is zero for making the correct selection, selecting the best future observation X_ℓ^f, and one otherwise.

The NPI lower (upper) expected zero-one loss for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ is equal to the NPI lower (upper) probability for the complementary event, which is $X_\ell^f < \max_{j \neq \ell} X_j^f$, as discussed in Section 3.4.1. Hence, the NPI lower and upper expected zero-one losses, respectively, are

$$\underline{E}(L_{\{0,1\}}(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \underline{P}(X_\ell^f < \max_{j \neq \ell} X_j^f) = 1 - \overline{P}(X_\ell^f \geq \max_{j \neq \ell} X_j^f) \tag{3.22}$$

$$\overline{E}(L_{\{0,1\}}(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \overline{P}(X_\ell^f < \max_{j \neq \ell} X_j^f) = 1 - \underline{P}(X_\ell^f \geq \max_{j \neq \ell} X_j^f) \tag{3.23}$$

Second, the focus is on the selection of a subset of the best groups. In other words, all future observations belonging to the subset S are considered greater than all future observations belonging to the non-selected groups, NS, $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$. We consider a subset of the k independent groups containing the w best groups, where $1 \leq w \leq k - 1$, noting that $w = 1$ corresponds to selecting the best group case presented earlier in this section. Let $S = \{\ell_1, \ldots, \ell_w\} \subset \{1, \ldots, k\}$ denote the indices of these w groups, and let $NS = \{1, \ldots, k\} \setminus S$ denote the indices of

the remaining $k - w$ groups not included S. Let $L_{\{0,1\}}$ be a zero-one loss function designed to penalise incorrect selections, which occur when choosing the non-selected groups, NS, as the best.

$$L_{\{0,1\}}(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f) = \begin{cases} 0 & \text{if } \min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ 1 & \text{if } \min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \quad (3.24)$$

The NPI lower (upper) expected zero-one loss for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ gives the same result as the NPI lower (upper) probability for the complement event $\min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f$. Therefore, the NPI lower and upper expected zero-one losses, respectively, are

$$\underline{E}(L_{\{0,1\}}(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \underline{P}(\min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f) \quad (3.25)$$

$$\overline{E}(L_{\{0,1\}}(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \overline{P}(\min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f) \quad (3.26)$$

Finally, the selection of at least one future observation that belongs to S is greater than all future observations that belong to NS. The zero-one loss function $L_{\{0,1\}}$ is used here to select a subset to include the best groups, $\ell \in S$. It assigns one for an incorrect selection, selecting $j \in NS$.

$$L_{\{0,1\}}(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f) = \begin{cases} 0 & \text{if } \max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ 1 & \text{if } \max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \quad (3.27)$$

The NPI lower (upper) expected zero-one loss for the event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ gives the same result as the NPI lower (upper) probability but for the complement event $\max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f$. Consequently, the NPI lower and upper expected zero-one losses, respectively, are

$$\underline{E}(L_{\{0,1\}}(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \underline{P}(\max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f) \quad (3.28)$$

$$\overline{E}(L_{\{0,1\}}(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \overline{P}(\max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f) \quad (3.29)$$

3.5.2 Linear loss function

In this section, the linear loss function is considered, focusing on the same events previously discussed under the zero-one loss function. The penalty incurred is calculated based on the deviation of the selected groups from the non-selected groups across three different events. The first scenario involves selecting the best group. The definition of the linear loss for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ is

$$L_l(X_\ell^f, \max_{j \neq \ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j \neq \ell} X_j^f \\ X_\ell^f - \max_{j \neq \ell} X_j^f & \text{if } X_\ell^f < \max_{j \neq \ell} X_j^f \end{cases} \qquad (3.30)$$

This loss function quantifies the loss in one direction, specifically when $X_\ell^f < \max_{j \neq \ell} X_j^f$, as the linear deviation between the selected group ℓ and all other non-selected groups j, where $\ell \neq j$.

Then, the NPI lower expected linear loss is derived by assigning the probability mass for group ℓ to the right endpoint per interval and for each group $j \neq \ell$ to the left endpoint per interval. Then, for each pair, the NPI lower expected linear loss is computed as the difference $(x_{\ell,i_\ell} - \max_{j \neq \ell} x_{j,i_j-1})$ whenever $x_{\ell,i_\ell} < \max_{j \neq \ell} x_{j,i_j-1}$, and zero otherwise. Similarly, the NPI upper expected linear loss is obtained by assigning the probability mass for group ℓ to the left endpoint per interval and for each group $j \neq \ell$ to the right endpoint per interval. Then, for each pair, the NPI upper expected linear loss is computed as the difference $(x_{\ell,i_\ell-1} - \max_{j \neq \ell} x_{j,i_j})$ whenever $x_{\ell,i_\ell-1} < \max_{j \neq \ell} x_{j,i_j}$, and zero otherwise. Therefore, the NPI lower and upper expected linear losses are

$$\underline{E}(L_l(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(x_{\ell,i_\ell}, \max_{j \neq \ell} x_{j,i_j-1}) \qquad (3.31)$$

$$\overline{E}(L_l(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(x_{\ell,i_\ell-1}, \max_{j \neq \ell} x_{j,i_j}) \qquad (3.32)$$

The $\sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1}$ represents multiple summations over a set of k groups, $j = 1, 2, \ldots, k$, excluding ℓ and over all i_j from 1 to $n_j + 1$.

The second scenario involves selecting the subset of the best groups. The definition of this loss for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ is as follows.

$$L_l\left(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f\right) = \begin{cases} 0 & \text{if } \min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ \min_{\ell \in S} X_\ell^f - \max_{j \in NS} X_j^f & \text{if } \min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \quad (3.33)$$

To derive the NPI lower (upper) expected linear loss for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, the probability mass that corresponds to groups $\ell \in S$ will be assigned to the right (left) endpoint per interval, while the probability mass that corresponds to groups $j \in NS$ will be assigned to the left (right) endpoint per interval. For each pair, the NPI lower (upper) expected linear loss is computed as the difference $\left(\min_{\ell \in S} x_{\ell,i_\ell} - \max_{j \in NS} x_{j,i_j-1}\right) \left(\left(\min_{\ell \in S} x_{\ell,i_\ell-1} - \max_{j \in NS} x_{j,i_j}\right)\right)$ whenever $\min_{\ell \in S} x_{\ell,i_\ell} < \max_{j \in NS} x_{j,i_j-1}$ $\left(\min_{\ell \in S} x_{\ell,i_\ell-1} < \max_{j \in NS} x_{j,i_j}\right)$, and zero otherwise. Then, the NPI lower and upper expected linear losses are

$$\underline{E}\left(L_l\left(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f\right)\right) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l\left(\min_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}\right) \quad (3.34)$$

$$\overline{E}\left(L_l\left(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f\right)\right) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l\left(\min_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}\right) \quad (3.35)$$

where $\sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1}$ represents multiple summations over a set of w groups and over all i_ℓ from 1 to $n_\ell+1$, with $\ell \in S$. The $\sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1}$ is defined under Equations (3.31) and (3.32).

The last scenario appears when selecting the subset that includes the best group, for the event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$. The definition of the linear loss function for the event mentioned above is

$$L_l\left(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f\right) = \begin{cases} 0 & \text{if } \max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ \max_{\ell \in S} X_\ell^f - \max_{j \in NS} X_j^f & \text{if } \max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \quad (3.36)$$

To derive the NPI lower (upper) expected linear loss for the event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, the probability mass that corresponds to groups $\ell \in S$ will be assigned to

the right (left) endpoint per interval, while the probability mass that corresponds to groups $j \in NS$ will be assigned to the left (right) endpoint per interval. For each pair, the NPI lower (upper) expected linear loss is computed as the difference $(\max_{\ell \in S} x_{\ell,i_\ell} - \max_{j \in NS} x_{j,i_j-1}) \left((\max_{\ell \in S} x_{\ell,i_\ell-1} - \max_{j \in NS} x_{j,i_j}) \right)$ whenever $\max_{\ell \in S} x_{\ell,i_\ell} < \max_{j \in NS} x_{j,i_j-1}$ $(\max_{\ell \in S} x_{\ell,i_\ell-1} < \max_{j \in NS} x_{j,i_j})$, and zero otherwise. Then, the NPI lower and upper expected linear losses are

$$\underline{E}(L_l(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\max_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}) \quad (3.37)$$

$$\overline{E}(L_l(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\max_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}) \quad (3.38)$$

These multiple summations $\sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1}$ and $\sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1}$ are defined as above.

3.5.3 Quadratic loss function

Similarly, the same events mentioned above are considered; however, the focus here is on quantifying the impact of selecting a suboptimal group based on future observations from the three different events, using the quadratic loss function. The first case is selecting the best group. The definition of the quadratic loss for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ is

$$L_q(X_\ell^f, \max_{j \neq \ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j \neq \ell} X_j^f \\ (X_\ell^f - \max_{j \neq \ell} X_j^f)^2 & \text{if } X_\ell^f < \max_{j \neq \ell} X_j^f \end{cases} \quad (3.39)$$

The NPI lower (upper) expected quadratic loss for the event $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ is derived by putting the probability mass per interval at the right (left) endpoint per interval for group ℓ and at the left (right) endpoint per interval for all other groups. Therefore, the NPI lower and upper expected quadratic losses, respectively, are

$$\underline{E}(L_q(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \frac{1}{\prod_{j=1}^{k}(n_j+1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(x_{\ell,i_\ell}, \max_{j \neq \ell} x_{j,i_j-1}) \quad (3.40)$$

$$\overline{E}(L_q(X_\ell^f, \max_{j \neq \ell} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(x_{\ell,i_\ell-1}, \max_{j \neq \ell} x_{j,i_j}) \qquad (3.41)$$

The second case is to select the subset that contains the best groups. Again, the selection of the subset of best groups will be evaluated using the quadratic loss function. The definition of quadratic loss for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ is

$$L_q(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f) = \begin{cases} 0 & \text{if } \min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ (\min_{\ell \in S} X_\ell^f - \max_{j \in NS} X_j^f)^2 & \text{if } \min_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \qquad (3.42)$$

For this case, the NPI lower (upper) expected loss for the quadratic loss function is derived by assigning the probability mass per interval to the right (left) endpoint for all w groups in the subset and to the left (right) endpoint for the other $k - w$ groups. The NPI lower and upper expected quadratic losses, respectively, for the event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ are

$$\underline{E}(L_q(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(\min_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}) \quad (3.43)$$

$$\overline{E}(L_q(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(\min_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}) \quad (3.44)$$

The last case is to select the subset including the best group. The definition of quadratic loss for the event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ is

$$L_q(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f) = \begin{cases} 0 & \text{if } \max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f \\ (\max_{\ell \in S} X_\ell^f - \max_{j \in NS} X_j^f)^2 & \text{if } \max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f \end{cases} \qquad (3.45)$$

In the final case, which involves selecting a subset that includes the best group, deriving the NPI lower (upper) expected quadratic loss requires assigning the probability mass per interval to the right (left) endpoints for all the groups that belong to the subset S and to the left (right) endpoints for the other groups that belong to NS.

The NPI lower and upper expected quadratic losses for the event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, respectively, are

$$\underline{E}(L_q(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(\max_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}) \quad (3.46)$$

$$\overline{E}(L_q(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_q(\max_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}) \quad (3.47)$$

Note: As in pairwise comparisons with linear and quadratic loss, future observations are bounded by l and r, set below the minimum and above the maximum observed data points, respectively.

3.5.4   Strong and weak preference

This section extends the analysis by introducing strong and weak preference criteria based on comparisons of NPI lower and upper expected losses across group combinations. A strong preference occurs when the NPI upper expected loss for event A, $\overline{E}(L(A))$, is less than the NPI lower expected loss for event B, $\underline{E}(L(B))$; that is, $\overline{E}(L(A)) < \underline{E}(L(B))$, indicating that A is clearly preferred over B due to a smaller NPI upper expected loss. When this condition is not met, a weak preference may still be established if both the lower and upper expected losses of A are less than those of B; that is, $\underline{E}(L(A)) < \underline{E}(L(B))$ and $\overline{E}(L(A)) < \overline{E}(L(B))$, suggesting that A tends to be better than B, albeit with less certainty.

In this section, the events A and B correspond to either selecting the best group, selecting a subset of the best groups or selecting the subset that includes the best group. By incorporating strong and weak preference criteria, we not only identify the best groups, subsets of the best groups and the subset that includes the best group, but we also determine which groups or subsets are clearly inferior and which might still be an acceptable alternative.

Additionally, it may be useful to consider all events whose NPI lower and upper expected losses overlap with the lowest NPI lower and upper expected losses of the best group, subset of the best groups or the subset that includes the best group.

This allows for decision-making where a non-optimal but acceptable group or subset could still be considered a reasonable alternative.

The following example quantifies the loss of making incorrect NPI multiple comparisons using zero-one, linear, and quadratic loss functions. Additionally, the results under each loss function are compared to assess whether they lead to the selection of the same groups and support consistent conclusions.

Example 3.5.1 Simpson and Margolin [68] presented the dataset shown in Table 3.2. It contains data from an experiment in which six different doses of a chemical, Acid Red 114, were tested using the Ames test. This test uses the Salmonella strain TA98 to detect whether the chemical induces mutations. For each dose, the table lists the number of mutation colonies—an indicator of mutagenic activity. Each row corresponds to a dose level (Dose 1 to Dose 6), and the values represent replicate measurements. Higher counts indicate more mutation colonies and may suggest a stronger mutagenic effect of the chemical.

The aim is to illustrate the NPI multiple comparisons method using the loss functions introduced in this section, applying it to measure the penalty of selecting suboptimal groups based on NPI lower and upper expected zero-one, linear, and quadratic losses across all dose levels. In this example, three events of interest are considered: selecting the best dose $X_\ell^f \geq \max\limits_{j \neq \ell} X_j^f$, selecting a subset of the best doses $\min\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$, and selecting a subset that contains the best dose $\max\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$. It is worth noting that the dataset contains tied observations. To simplify inference, the ties between observations are broken by adding a very small amount to each tied observation. This adjustment does not impact their rankings within each dose or across other doses.

The lower and upper bounds for future observations are defined as $l = 7$ and $r = 99$, where l is smaller than the minimum observed value (8) and r is greater than the maximum (98). These bounds are used to calculate the NPI lower and upper expected linear and quadratic losses.

Table 3.3, in the first and second columns , presents the NPI lower and upper expected zero-one losses for selecting the best dose. Using Equations (3.22) and (3.23), the best dose in terms of selection is Dose 4, as it has the lowest NPI lower

Dose	Data								
1	14.0	16.0	17.0	19.0	22.0	22.1	23.0	23.1	35.0
2	15.0	21.0	23.2	24.0	25.0	27.0	54.0	59.1	60.0
3	17.1	26.0	28.0	31.0	35.1	37.0	50.0	78.0	98.0
4	30.0	37.1	39.0	41.0	43.0	44.0	59.0	60.1	82.0
5	21.1	22.2	23.3	26.1	28.1	30.1	33.0	33.1	44.1
6	8.0	10.0	13.0	16.1	19.1	21.2	23.4	25.1	

Table 3.2: Dataset from Ames tests showing the number of mutation colonies of Salmonella strain TA98 exposed to six different doses of Acid Red 114, used in Example 3.5.1. These are not the original data, and ties have been broken by adding 0.1

and upper expected zero-one losses $\underline{E}(L_{\{0,1\}}) = 0.4304$ and $\overline{E}(L_{\{0,1\}}) = 0.7459$. This is because the observations for Dose 4 are generally higher than those for the other doses. Values such as 82, 60.1, and 59 are among the highest in the dataset, indicating better performance compared to the other doses.

On the other hand, the worst doses appear to be Doses 1 and 6. Dose 1 has the highest NPI lower expected zero-one loss $\underline{E}(L_{\{0,1\}}) = 0.8932$, while Dose 6 has the highest NPI upper expected zero-one loss $\overline{E}(L_{\{0,1\}}) = 1.0000$. This is because observations from these doses are generally smaller than other doses. For example, Dose 1 includes smaller values such as 14 and 16, while Dose 6 includes values such as 8 and 10, which contribute to their higher losses. Therefore, according to the zero-one loss function, selecting these doses would be poor decisions.

Table 3.4 presents the NPI lower and upper expected zero-one losses for selecting a subset of the best groups. The subset $\{2, 3, 4, 5\}$ has the lowest NPI lower and upper expected zero-one losses, with $\underline{E}(L_{\{0,1\}}) = 0.3948$ and $\overline{E}(L_{\{0,1\}}) = 0.7155$, indicating that it is the best selected subset.

Selecting a non-optimal subset means choosing a subset with higher NPI lower and upper expected zero-one losses, which increases the risk of making a suboptimal selection decision. However, it may still be reasonable to select a subset that is not the best but remains an acceptable alternative. For example, consider the

ℓ	Zero-one Loss		Linear Loss		Quadratic Loss	
	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$
{1}	0.8932	0.9983	34.1748	57.1891	1673.339	3879.282
{2}	0.7129	0.8992	22.7565	45.3154	1033.496	2869.519
{3}	0.6216	0.8537	14.0939	35.7977	457.8021	2130.9197
{4}	0.4304	0.7459	10.6018	32.3091	403.8369	1812.2227
{5}	0.8428	0.9780	27.2109	50.1899	1202.542	3162.278
{6}	0.8880	1.0000	37.5337	60.8572	1957.901	4305.504

Table 3.3: NPI lower and upper expected zero-one, linear, and quadratic loss values for selecting the best group, where the event is defined as $X_\ell^f \geq \max_{j \neq \ell} X_j^f$, for Example 3.5.1.

subset $\{3, 4, 5\}$ with $\underline{E}(L_{\{0,1\}}) = 0.5527$ and $\overline{E}(L_{\{0,1\}}) = 0.8003$. Although it does not perform as well as $\{2, 3, 4, 5\}$, it still performs better than many other subsets because its corresponding NPI lower and upper expected zero-one losses are lower in comparison. This indicates a smaller expected loss and, therefore, a lower risk of incorrect selection than that associated with other subsets whose losses are higher.

The range between $\underline{E}(L_{\{0,1\}})$ and $\overline{E}(L_{\{0,1\}})$ represents the uncertainty interval for each subset. In Table 3.4, the range between $\underline{E}(L_{\{0,1\}})$ and $\overline{E}(L_{\{0,1\}})$ varies across subsets. For subsets $\{2, 3, 4, 5\}$, the range is 0.3207, with low NPI lower and upper expected zero-one losses. However, for subsets $\{1, 2, 6\}$, the range is 0.0056, with high NPI lower and upper expected zero-one losses.

Tables 3.3, in the third and fourth columns, 3.6, and 3.7 show the NPI lower and upper expected linear losses for the three events considered above, respectively. The linear loss function differs from the zero-one loss function because it does not classify correct and incorrect selections, but measures the distance between them. Table 3.3, in the third and fourth columns, shows that Dose 6 is the worst selection, with the highest NPI lower and upper expected linear losses $\underline{E}(L_l) = 37.5337$, $\overline{E}(L_l) = 60.8572$. The second worst selection is Dose 1, with the NPI lower and upper expected linear losses of $\underline{E}(L_l) = 34.1748$ and $\overline{E}(L_l) = 57.1891$. This differs from the zero-one loss function, where Dose 1 had the highest NPI lower expected

zero-one loss and Dose 6 had the highest NPI upper expected zero-one loss. Similarly to the zero-one loss function, the lowest NPI lower and upper expected linear losses are associated with the selection of Dose 4 as the best.

To illustrate subset selection, we begin by considering the selection of a subset containing all the best doses. Table 3.6 presents three cases for selecting a subset of the best doses: a subset containing two out of six doses, a subset containing three out of six doses and a subset containing four out of six doses.

In the first case, the NPI lower and upper expected linear losses suggest that selecting $\{3, 4\}$, followed by $\{2, 4\}$, as a subset containing the best doses, leads to good selections. However, the NPI lower and upper expected linear losses indicate that selecting $\{1, 6\}$ and $\{5, 6\}$ would be poor decisions, making it unlikely to yield the largest future observations.

Moving on to subsets containing three doses out of six, the subset $\{2, 3, 4\}$ shows the smallest NPI lower and upper expected linear losses followed by $\{3, 4, 5\}$, with the NPI lower expected linear loss of 5.6885 and 10.3902 respectively, and an NPI upper expected linear loss of 29.8954 and 33.7647 respectively. The subset of best groups $\{2, 3, 4\}$ has the lowest NPI lower and upper expected linear losses, while the complementary subset $\{1, 5, 6\}$ has the highest NPI lower and upper expected linear losses, which would be a poor decision. Based on the results of this example, a possible conclusion is that when a subset of the best doses leads to a good selection, its complementary subset is more likely to lead to a poor selection. Moreover, most of the remaining NPI lower and upper expected linear losses were relatively high, suggesting an unfavourable decision.

Considering subsets containing four out of six doses, the values $\underline{E}(L_l) = 2.2401$ and $\overline{E}(L_l) = 22.0169$ corresponding to the subset $\{2, 3, 4, 5\}$ represent the smallest NPI lower and upper expected linear losses among all subsets in Table 3.6. This strongly suggests that selecting this subset would be a good decision. The complement of the subset of $\{2, 3, 4, 5\}$, which is $\{1, 6\}$, has the highest NPI lower and upper expected linear losses. Generally, it is evident that all selected subsets including doses 6 and 1 lead to high NPI lower and upper expected linear losses, while subsets including doses 4 and 3 result in low NPI lower and upper expected linear

losses.

Table 3.7 shows the NPI lower and upper expected linear losses for selecting a subset that includes the best group. Most of the NPI lower and upper expected linear losses in this table are lower than those in Table 3.6. In this table, the lowest NPI lower expected linear loss is almost zero and corresponds to the subset $\{2, 3, 4, 5\}$.

The NPI lower and upper expected quadratic losses for the three different selection cases are presented in Tables 3.3, in the fifth and sixth columns, 3.8 and 3.9, respectively. The quadratic loss function incurs a penalty of the quadratic deviation between the selected doses and non-selected doses. A large difference between the selected doses and non-selected doses is penalised more than a smaller one. For example, in Table 3.3, in the fifth and sixth columns, the NPI lower and upper expected quadratic losses for selecting Dose 6 as the best are $\underline{E}(L_q(X_6^f, \max_{j \neq 6} X_j^f)) = 1957.901$ and $\overline{E}(L_q(X_6^f, \max_{j \neq 6} X_j^f)) = 4305.504$, respectively, while penalised less for selecting Dose 1 as the best with $\underline{E}(L_q(X_1^f, \max_{j \neq 1} X_j^f)) = 1673.339$ and $\overline{E}(L_q(X_1^f, \max_{j \neq 1} X_j^f)) = 3879.282$. Selecting Dose 6 as the best is penalised more than selecting Dose 1 and this is due to the fact that the observations that belong to Dose 6 are smaller than those of Dose 1. For the rest doses, the results in this table confirm those in Table 3.3, in the third and fourth columns. In conclusion, the selected subsets based on linear and quadratic loss functions are similar. Specifically, the selected subsets in Table 3.8 (quadratic loss for a subset of best groups) and Table 3.9 (quadratic loss for a subset containing the best group) are similar to those in Table 3.6 and Table 3.7 (linear loss function). Furthermore, under the quadratic loss function, selecting a subset of the best groups in Table 3.8 is similar to selecting a subset including the best group Table 3.9.

In conclusion, the results of the zero-one loss function differ slightly from those of the linear and quadratic loss functions. However, the results of the linear and quadratic loss functions are fairly similar to each other and lead to the same conclusions.

This discussion can now be extended by introducing strong and weak preference criteria, based on comparisons of NPI lower and upper expected losses across different group combinations. Since the findings in Table 3.3 for the zero-one, linear, and

quadratic losses are similar, the discussion of strong and weak preferences will focus on the first and second columns, which present the NPI lower and upper expected zero-one losses.

For strong preference, selecting Dose 4 as the best is strongly preferred over selecting Doses 1, 5, and 6 as the best because the NPI upper expected zero-one loss for selecting Dose 4 as the best, which is $\overline{E}(L_{\{0,1\}}) = 0.7459$, is less than the NPI lower expected zero-one loss for selecting Doses 1, 5, and 6, which are 0.8932, 0.8428, and 0.8880, respectively. Therefore, they are considered inferior alternatives.

For weak preference, for example, a comparison between selecting Dose 4 and Dose 3 shows that the NPI lower and upper expected zero-one losses for selecting Dose 4 as the best are both smaller than those for selecting Dose 3 as the best. Specifically, $\underline{E}(L_{\{0,1\}}) = 0.4304$ for Dose 4 compared with 0.6216 for Dose 3, and $\overline{E}(L_{\{0,1\}}) = 0.7459$ for Dose 4 compared with 0.8537 for Dose 3. However, since the NPI lower and upper expected zero-one losses intervals overlap between Doses 4 and 3, Dose 4 is weakly preferred over Dose 3, which means that while Dose 4 is the better choice, Dose 3 remains a reasonable alternative.

It may be of interest to consider all doses which have the NPI lower and upper expected zero-one losses overlap with NPI lower and upper expected zero-one losses of the best dose. Comparing doses 3 and 2 shows that the NPI lower expected loss is $\underline{E}(L_{\{0,1\}}) = 0.6216$ for dose 3 and $\underline{E}(L_{\{0,1\}}) = 0.7129$ for dose 2, so dose 3 has the smaller value. Similarly, the NPI upper expected loss is $\overline{E}(L_{\{0,1\}}) = 0.8537$ for dose 3 and $\overline{E}(L_{\{0,1\}}) = 0.8992$ for dose 2. This again suggests a weak preference for Dose 3 over Dose 2, meaning that Dose 3 is more favorable, but Dose 2 is still relatively competitive. There is a partial overlap between the NPI lower and upper expected zero-one losses of Dose 4 with NPI lower expected zero-one losses of doses 3 and 2. Therefore, Doses 3 and 2 not strongly worse and still be considered a reasonable alternative.

To illustrate the concept of strong preference for selecting a subset of the best doses, we examine a few subsets from Table 3.4, rather than the entire table due to its length. The lowest NPI lower and upper expected zero-one losses are related to a select subset $\{2, 3, 4, 5\}$, with 0.3948 and 0.7155, respectively. Since the NPI

upper expected zero-one loss of the subset $\{2, 3, 4, 5\}$ is less than the NPI lower expected zero-one loss of multiple other subsets, this indicates a strong preference for $\{2, 3, 4, 5\}$ over these subsets. This suggests that selecting $\{2, 3, 4, 5\}$ results in a smaller loss compared to these alternatives. Furthermore, the NPI upper expected zero-one loss for $\{4, 5, 6\}$ was less than the NPI lower expected zero-one loss for $\{1, 5, 6\}$. Also, the NPI upper expected zero-one loss for $\{2, 3, 5\}$ was less than the NPI lower expected zero-one loss for $\{1, 3, 6\}$, $\{1, 2, 6\}$, and $\{1, 5, 6\}$. Therefore, the subset $\{4, 5, 6\}$ is strongly preferred over $\{2, 3, 5\}$ and the subset $\{2, 3, 5\}$ is strongly preferred over $\{1, 3, 6\}$, $\{1, 2, 6\}$, and $\{1, 5, 6\}$. These subsets should be excluded from consideration.

For weak preference, the NPI lower expected zero-one loss for subset $\{2, 3, 4, 5\}$ is less than the NPI lower expected zero-one loss for subsets $\{2, 4\}$, $\{3, 4\}$, $\{3, 4, 5\}$, and $\{2, 3, 4\}$, and the NPI upper expected zero-one loss for subset $\{2, 3, 4, 5\}$ is less than the NPI upper expected zero-one loss for subsets $\{2, 4\}$, $\{3, 4\}$, $\{3, 4, 5\}$, and $\{2, 3, 4\}$. Therefore, $\{2, 3, 4, 5\}$ demonstrates a weak preference over several other subsets. This suggests that $\{2, 3, 4, 5\}$ is a preferable selection compared to these subset alternatives. Unlike strong preference, weak preference allows for some overlap in NPI lower and upper expected zero-one losses. The NPI lower and upper expected zero-one losses for $\{2, 4\}$, $\{3, 4\}$, $\{3, 4, 5\}$, and $\{2, 3, 4\}$ partially overlap with the NPI lower and upper expected zero-one losses for the subset of the best groups $\{2, 3, 4, 5\}$, meaning they might still be considered in decision-making.

Similar methods are applied to all three events described earlier, using the three different loss functions. To avoid repetition, only key results and interpretations are presented.

S	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$	S	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$
$\{1,2\}$	0.9657	0.9991	$\{2,3,4\}$	0.6070	0.8366
$\{1,3\}$	0.9504	0.9975	$\{2,3,5\}$	0.9090	0.9911
$\{1,4\}$	0.9077	0.9848	$\{2,3,6\}$	0.9741	1.0000
$\{1,5\}$	0.9777	0.9992	$\{2,4,5\}$	0.7860	0.9312
$\{1,6\}$	0.9876	1.0000	$\{2,4,6\}$	0.9504	0.9965
$\{2,3\}$	0.8391	0.9531	$\{2,5,6\}$	0.9862	1.0000
$\{2,4\}$	0.7031	0.8844	$\{3,4,5\}$	0.5527	0.8003
$\{2,5\}$	0.9310	0.9890	$\{3,4,6\}$	0.9029	0.9840
$\{2,6\}$	0.9664	1.0000	$\{3,5,6\}$	0.9760	1.0000
$\{3,4\}$	0.4989	0.7738	$\{4,5,6\}$	0.9485	0.9942
$\{3,5\}$	0.8903	0.9823	$\{1,2,3,4\}$	0.8340	0.9607
$\{3,6\}$	0.9538	1.0000	$\{1,2,3,5\}$	0.9316	0.9994
$\{4,5\}$	0.7538	0.9138	$\{1,2,3,6\}$	0.9848	1.0000
$\{4,6\}$	0.9277	0.9980	$\{1,2,4,5\}$	0.8709	0.9720
$\{5,6\}$	0.9801	1.0000	$\{1,2,4,6\}$	0.9749	0.9984
$\{1,2,3\}$	0.9687	0.9988	$\{1,2,5,6\}$	0.9857	1.0000
$\{1,2,4\}$	0.9334	0.9891	$\{1,3,4,5\}$	0.7761	0.9302
$\{1,2,5\}$	0.9813	0.9996	$\{1,3,4,6\}$	0.9572	0.9954
$\{1,2,6\}$	0.9944	1.0000	$\{1,3,5,6\}$	0.9794	1.0000
$\{1,3,4\}$	0.8761	0.9699	$\{1,4,5,6\}$	0.9653	0.9962
$\{1,3,5\}$	0.9685	0.9987	$\{2,3,4,5\}$	0.3948	0.7155
$\{1,3,6\}$	0.9916	1.0000	$\{2,3,4,6\}$	0.8756	0.9722
$\{1,4,5\}$	0.9278	0.9857	$\{2,3,5,6\}$	0.9562	1.0000
$\{1,4,6\}$	0.9844	0.9994	$\{2,4,5,6\}$	0.9146	0.9843
$\{1,5,6\}$	0.9949	1.0000	$\{3,4,5,6\}$	0.8322	0.9509

Table 3.4: NPI lower and upper expected zero-one losses of selecting the subset of best groups, where the event is defined as $\min\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$ for Example 3.5.1.

S	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$	S	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$
$\{1,2\}$	0.6359	0.8860	$\{2,3,4\}$	0.0297	0.3312
$\{1,3\}$	0.5543	0.8353	$\{2,3,5\}$	0.3219	0.6583
$\{1,4\}$	0.3844	0.7141	$\{2,3,6\}$	0.3557	0.6852
$\{1,5\}$	0.7530	0.9736	$\{2,4,5\}$	0.1853	0.5076
$\{1,6\}$	0.7932	0.9981	$\{2,4,6\}$	0.2125	0.5417
$\{2,3\}$	0.4005	0.7202	$\{2,5,6\}$	0.5269	0.8452
$\{2,4\}$	0.2390	0.5928	$\{3,4,5\}$	0.1282	0.4353
$\{2,5\}$	0.5933	0.8624	$\{3,4,6\}$	0.1552	0.4708
$\{2,6\}$	0.6330	0.8866	$\{3,5,6\}$	0.4566	0.7874
$\{3,4\}$	0.1747	0.5300	$\{4,5,6\}$	0.3134	0.6438
$\{3,5\}$	0.5140	0.8110	$\{1,2,3,4\}$	0.0247	0.2514
$\{3,6\}$	0.5521	0.8354	$\{1,2,3,5\}$	0.2861	0.6174
$\{4,5\}$	0.3526	0.6837	$\{1,2,3,6\}$	0.3163	0.6474
$\{4,6\}$	0.3826	0.7139	$\{1,2,4,5\}$	0.1646	0.4479
$\{5,6\}$	0.7486	0.9753	$\{1,2,4,6\}$	0.1890	0.4860
$\{1,2,3\}$	0.3562	0.6866	$\{1,2,5,6\}$	0.4700	0.8253
$\{1,2,4\}$	0.2126	0.5434	$\{1,3,4,5\}$	0.1134	0.3670
$\{1,2,5\}$	0.5292	0.8448	$\{1,3,4,6\}$	0.1376	0.4067
$\{1,2,6\}$	0.5647	0.8718	$\{1,3,5,6\}$	0.4072	0.7610
$\{1,3,4\}$	0.1548	0.4731	$\{1,4,5,6\}$	0.2798	0.5995
$\{1,3,5\}$	0.4583	0.7875	$\{2,3,4,5\}$	0.0019	0.2068
$\{1,3,6\}$	0.4924	0.8147	$\{2,3,4,6\}$	0.0264	0.2470
$\{1,4,5\}$	0.3148	0.6443	$\{2,3,5,6\}$	0.2859	0.6156
$\{1,4,6\}$	0.3417	0.6781	$\{2,4,5,6\}$	0.1647	0.4457
$\{1,5,6\}$	0.6688	0.9703	$\{3,4,5,6\}$	0.1140	0.3641

Table 3.5: NPI lower and upper expected zero-one losses of selecting a subset including the best group, where the event is defined as $\max\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$ for Example 3.5.1.

S	$\underline{E}(L_l)$	$\overline{E}(L_l)$	S	$\underline{E}(L_l)$	$\overline{E}(L_l)$
$\{1,2\}$	34.9477	55.3738	$\{2,3,4\}$	5.68850	29.8954
$\{1,3\}$	28.7743	51.1647	$\{2,3,5\}$	22.3451	43.9112
$\{1,4\}$	27.8148	50.7094	$\{2,3,6\}$	29.3693	48.8225
$\{1,5\}$	37.3066	56.6131	$\{2,4,5\}$	19.2199	41.2513
$\{1,6\}$	42.4880	60.4514	$\{2,4,6\}$	27.5179	47.5186
$\{2,3\}$	19.8424	43.7958	$\{2,5,6\}$	38.4210	55.5623
$\{2,4\}$	17.3581	41.3344	$\{3,4,5\}$	10.3902	33.7647
$\{2,5\}$	30.3934	51.6082	$\{3,4,6\}$	19.9305	41.6327
$\{2,6\}$	37.8822	57.8642	$\{3,5,6\}$	32.5862	51.2700
$\{3,4\}$	8.46740	33.7454	$\{4,5,6\}$	31.4056	50.5844
$\{3,5\}$	23.0669	46.2162	$\{1,2,3,4\}$	8.73640	28.1915
$\{3,6\}$	32.0150	53.9602	$\{1,2,3,5\}$	25.9832	43.5788
$\{4,5\}$	21.1008	44.5300	$\{1,2,3,6\}$	30.8707	46.1092
$\{4,6\}$	31.3320	53.8619	$\{1,2,4,5\}$	22.6324	40.5829
$\{5,6\}$	40.4324	59.2961	$\{1,2,4,6\}$	28.9439	44.5852
$\{1,2,3\}$	26.5165	46.8793	$\{1,2,5,6\}$	39.8022	53.6459
$\{1,2,4\}$	24.4091	45.3344	$\{1,3,4,5\}$	13.5886	32.4965
$\{1,2,5\}$	35.6140	53.5713	$\{1,3,4,6\}$	21.3434	37.9693
$\{1,2,6\}$	40.1856	56.5326	$\{1,3,5,6\}$	34.1252	48.9638
$\{1,3,4\}$	16.5288	39.1716	$\{1,4,5,6\}$	32.8671	48.1050
$\{1,3,5\}$	29.5769	49.1493	$\{2,3,4,5\}$	2.24010	22.0169
$\{1,3,6\}$	34.6561	52.3227	$\{2,3,4,6\}$	12.0004	30.0951
$\{1,4,5\}$	28.1343	48.2002	$\{2,3,5,6\}$	28.7625	45.1369
$\{1,4,6\}$	34.2322	52.2573	$\{2,4,5,6\}$	25.7232	42.4603
$\{1,5,6\}$	42.5934	58.2843	$\{3,4,5,6\}$	16.9837	34.6531

Table 3.6: NPI lower and upper expected linear losses of selecting the subset of best groups, where the event is defined as $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ for Example 3.5.1.

S	$\underline{E}(L_l)$	$\overline{E}(L_l)$	S	$\underline{E}(L_l)$	$\overline{E}(L_l)$
$\{1,2\}$	19.3017	40.3889	$\{2,3,4\}$	0.12450	11.5390
$\{1,3\}$	12.0711	31.4007	$\{2,3,5\}$	5.53170	20.3509
$\{1,4\}$	9.49290	28.4785	$\{2,3,6\}$	6.35750	21.5724
$\{1,5\}$	23.7877	45.7575	$\{2,4,5\}$	5.55900	20.1295
$\{1,6\}$	29.1889	52.3636	$\{2,4,6\}$	5.69460	20.5044
$\{2,3\}$	7.19010	24.9083	$\{2,5,6\}$	14.7935	33.7774
$\{2,4\}$	6.40640	23.7837	$\{3,4,5\}$	1.86360	14.3990
$\{2,5\}$	16.6857	37.0411	$\{3,4,6\}$	2.01780	14.5894
$\{2,6\}$	19.7847	41.2150	$\{3,5,6\}$	9.43860	25.9697
$\{3,4\}$	2.27000	17.8944	$\{4,5,6\}$	8.16920	24.2211
$\{3,5\}$	10.6339	29.2327	$\{1,2,3,4\}$	0.10150	8.28640
$\{3,6\}$	12.3014	31.8729	$\{1,2,3,5\}$	4.87080	17.0155
$\{4,5\}$	9.19030	27.4469	$\{1,2,3,6\}$	5.53990	18.0082
$\{4,6\}$	9.42380	28.6125	$\{1,2,4,5\}$	4.98380	16.9224
$\{5,6\}$	24.0498	46.4143	$\{1,2,4,6\}$	5.09950	17.1635
$\{1,2,3\}$	6.25170	21.5808	$\{1,2,5,6\}$	13.0028	30.1196
$\{1,2,4\}$	5.73690	20.7035	$\{1,3,4,5\}$	1.66520	11.1756
$\{1,2,5\}$	14.6588	33.6260	$\{1,3,4,6\}$	1.80000	11.2582
$\{1,2,6\}$	16.9696	36.8722	$\{1,3,5,6\}$	8.32760	22.4753
$\{1,3,4\}$	2.02500	14.8396	$\{1,4,5,6\}$	7.32320	20.9291
$\{1,3,5\}$	9.37900	26.0007	$\{2,3,4,5\}$	0.00780	7.99790
$\{1,3,6\}$	10.6326	27.9503	$\{2,3,4,6\}$	0.11060	7.94240
$\{1,4,5\}$	8.23860	24.4311	$\{2,3,5,6\}$	4.91330	16.8239
$\{1,4,6\}$	8.43820	25.1496	$\{2,4,5,6\}$	4.94130	16.6126
$\{1,5,6\}$	21.0516	42.3873	$\{3,4,5,6\}$	1.65650	10.8340

Table 3.7: NPI lower and upper expected linear losses of selecting a subset including the best group, where the event is defined as $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ for Example 3.5.1.

S	$\underline{E}(L_q)$	$\overline{E}(L_q)$	S	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1,2\}$	1708.1740	3762.1080	$\{2,3,4\}$	80.715900	1911.2216
$\{1,3\}$	1098.1070	3280.4380	$\{2,3,5\}$	775.34130	2673.2810
$\{1,4\}$	1298.6850	3424.5130	$\{2,3,6\}$	1139.7230	3075.3430
$\{1,5\}$	1827.8860	3819.7430	$\{2,4,5\}$	931.67640	2758.7008
$\{1,6\}$	2236.2680	4248.0460	$\{2,4,6\}$	1321.0080	3221.2300
$\{2,3\}$	673.93840	2733.3166	$\{2,5,6\}$	1976.0380	3774.8407
$\{2,4\}$	789.77480	2721.4021	$\{3,4,5\}$	303.27680	2127.8704
$\{2,5\}$	1417.5300	3396.9990	$\{3,4,6\}$	630.49180	2616.2290
$\{2,6\}$	1941.7330	4039.7240	$\{3,5,6\}$	1343.3060	3243.9700
$\{3,4\}$	210.24360	2076.2528	$\{4,5,6\}$	1561.9900	3427.5630
$\{3,5\}$	796.11510	2841.8422	$\{1,2,3,4\}$	133.44520	1647.4998
$\{3,6\}$	1316.8800	3561.8270	$\{1,2,3,5\}$	961.78160	2532.0795
$\{4,5\}$	940.62030	2884.1461	$\{1,2,3,6\}$	1218.7610	2717.4820
$\{4,6\}$	1532.1110	3740.1370	$\{1,2,4,5\}$	1122.1620	2647.9050
$\{5,6\}$	2091.5310	4127.7950	$\{1,2,4,6\}$	1403.0410	2879.0640
$\{1,2,3\}$	965.71960	2904.1811	$\{1,2,5,6\}$	2078.9120	3520.3440
$\{1,2,4\}$	1145.8120	3037.9920	$\{1,3,4,5\}$	438.50690	1959.0988
$\{1,2,5\}$	1753.1950	3564.3350	$\{1,3,4,6\}$	687.20900	2191.3048
$\{1,2,6\}$	2088.5720	3863.8690	$\{1,3,5,6\}$	1439.8400	2926.5830
$\{1,3,4\}$	486.43670	2446.0625	$\{1,4,5,6\}$	1664.0700	3133.3480
$\{1,3,5\}$	1140.0590	3044.1850	$\{2,3,4,5\}$	22.213900	1433.1765
$\{1,3,6\}$	1453.7900	3332.7300	$\{2,3,4,6\}$	221.38460	1713.0703
$\{1,4,5\}$	1351.8360	3210.3620	$\{2,3,5,6\}$	1126.3560	2649.5620
$\{1,4,6\}$	1699.6140	3542.5310	$\{2,4,5,6\}$	1281.5220	2771.5500
$\{1,5,6\}$	2245.6750	3986.6030	$\{3,4,5,6\}$	559.99190	2058.4025

Table 3.8: NPI lower and upper expected quadratic losses of selecting the subset of best groups, where the event is defined as $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ for Example 3.5.1.

S	$\underline{E}(L_q)$	$\overline{E}(L_q)$	S	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1,2\}$	855.11350	2385.5194	$\{2,3,4\}$	0.932500	567.97400
$\{1,3\}$	379.96460	1720.7849	$\{2,3,5\}$	160.6063	996.80800
$\{1,4\}$	361.44270	1498.0677	$\{2,3,6\}$	188.5693	1075.2337
$\{1,5\}$	1039.6090	2709.2280	$\{2,4,5\}$	228.0368	991.80170
$\{1,6\}$	1394.0880	3337.8350	$\{2,4,6\}$	231.2735	1018.7411
$\{2,3\}$	214.09660	1312.9873	$\{2,5,6\}$	626.4363	1827.0093
$\{2,4\}$	260.18260	1211.3913	$\{3,4,5\}$	33.90700	686.22080
$\{2,5\}$	707.53340	2087.2030	$\{3,4,6\}$	36.55530	694.95610
$\{2,6\}$	886.43770	2467.8004	$\{3,5,6\}$	282.3014	1291.5134
$\{3,4\}$	41.124700	888.76560	$\{4,5,6\}$	311.6912	1207.1413
$\{3,5\}$	318.43500	1547.6210	$\{1,2,3,4\}$	0.768500	404.03420
$\{3,6\}$	392.94940	1765.9009	$\{1,2,3,5\}$	141.0007	779.25280
$\{4,5\}$	350.65260	1415.0635	$\{1,2,3,6\}$	163.0488	829.58280
$\{4,6\}$	358.96610	1521.0518	$\{1,2,4,5\}$	204.6445	814.22470
$\{5,6\}$	1059.0750	2774.1920	$\{1,2,4,6\}$	207.3289	826.50180
$\{1,2,3\}$	184.38720	1075.7943	$\{1,2,5,6\}$	547.3704	1543.1618
$\{1,2,4\}$	233.24500	1027.8210	$\{1,3,4,5\}$	30.28680	506.04670
$\{1,2,5\}$	617.76140	1813.8157	$\{1,3,4,6\}$	32.57200	503.55590
$\{1,2,6\}$	747.07410	2078.7775	$\{1,3,5,6\}$	247.1745	1025.0956
$\{1,3,4\}$	36.643500	708.69460	$\{1,4,5,6\}$	279.3515	998.23220
$\{1,3,5\}$	278.63650	1295.3316	$\{2,3,4,5\}$	0.034400	394.91390
$\{1,3,6\}$	331.89250	1431.6910	$\{2,3,4,6\}$	0.828900	385.85250
$\{1,4,5\}$	314.27040	1218.5646	$\{2,3,5,6\}$	142.5876	764.77990
$\{1,4,6\}$	321.28240	1273.2049	$\{2,4,5,6\}$	202.6994	797.09820
$\{1,5,6\}$	917.48620	2392.2870	$\{3,4,5,6\}$	30.13950	486.28550

Table 3.9: NPI lower and upper expected quadratic losses of selecting a subset including the best group, where the event is defined as $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ for Example 3.5.1.

$\diamond$

In the next example, simulated data from Normal distributions are used to quantify the loss incurred when suboptimal groups are selected instead of the best possible ones, based on the NPI lower and upper expected zero-one, linear and quadratic losses. The objective is to demonstrate how the method performs across various scenarios including different sample sizes and loss functions and to determine whether all three loss functions yield consistent conclusions.

Example 3.5.2 This example illustrates the NPI multiple comparisons method using zero-one, linear, and quadratic loss functions for four independent groups. Data are generated from Normal distributions with sample sizes $n = 10, 20, 50$, and 100, each with a standard deviation of 1. The group distributions are as follows: $X_1 \sim N(0, 1)$, $X_2 \sim N(0.2, 1)$, $X_3 \sim N(0.4, 1)$, and $X_4 \sim N(0.6, 1)$. To calculate the NPI lower and upper expected linear and quadratic losses, the lower and upper bounds for future observations are defined as $l = -5$ and $r = 5$.

The NPI lower and upper expected zero-one losses for the same simulated data mentioned above are not included here to avoid repetition, as they can be directly derived using the property, known as the NPI lower loss for an event is equal to the NPI lower probability for the complement event. However, the conclusions drawn from the zero-one loss function align with those of the linear and quadratic loss functions.

Tables 3.10, 3.11, and 3.12 present the NPI lower and upper expected linear losses for three different events: $X_\ell^f \geq \max_{j \neq \ell} X_j^f$, $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, and $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, while Tables 3.13, 3.14, and 3.15 show the corresponding to the NPI lower and upper expected quadratic losses for the same three events. Table 3.10 shows that the smallest NPI lower and upper expected linear losses are consistently associated with the selection of group 4 as the best across different sample sizes. Moreover, for sample sizes of $n = 10$ and $n = 50$, the second smallest NPI lower and upper expected linear losses are associated with selecting group 2 as the best, while for $n = 20$ and $n = 100$, the second smallest NPI lower and upper expected linear losses are associated with selecting group 3 as the best. Interestingly, the deviation between the NPI lower and upper expected linear losses for each selected group decreases as the sample size increases; this occurs because larger sample sizes provide more

information, which may lead to reduced uncertainty. For example, the difference between the NPI lower and upper expected linear losses for selecting group 4 as the best is about 1.5 when $n = 10$, while it is 0.2 when $n = 100$. However, selecting group 1 as the best leads to a poor decision because its associated NPI lower and upper expected linear losses are the largest.

In the context of selecting the subset of the best groups, as presented in Table 3.11, choosing $\{2, 3, 4\}$ as the best subset consistently leads to a good selection across all sample sizes. In particular, bad choices are influenced by the varying sample sizes. In this particular case, the subset $\{1, 3\}$ is the worst selection when the sample sizes are 10 and 50, while $\{1, 2\}$ is the worst selection when the sample sizes are 20 and 100. Furthermore, it is essential to note that the subset $\{1, 3\}$ results in the largest NPI upper expected linear loss, which is almost 3 for the smallest sample size $n = 10$.

Similarly, for Table 3.12, selecting the subset that includes the best group leads to the same conclusion as in Table 3.11, even though the NPI lower and upper expected linear losses in this table differ. Moreover, it is evident that for sample sizes of 50 and 100, all the NPI lower and upper expected linear losses in this table are less than 1.

Regarding the quadratic loss function, Table 3.13 shows similar results to Table 3.10, where the smallest NPI lower and upper expected quadratic losses are consistently associated with the selection of group 4 as the best across different sample sizes. Tables 3.14 and 3.15 present the NPI lower and upper expected quadratic losses for two different cases: selecting the subset of the best groups, and selecting the subset that includes the best group, respectively. The conclusions drawn from these tables are identical to those from the corresponding tables using the linear loss function, although the NPI lower and upper expected quadratic losses differ. In other words, for the case of selecting the best group, Table 3.10 corresponds to Table 3.13 when comparing the linear and quadratic loss functions, and the conclusions derived from these tables are similar. This pattern holds for all other cases as well.

From Table 3.10, no strong preference among groups is observed for sample sizes of 10 and 20, since the NPI upper expected linear loss of any group is not less than the NPI lower expected linear loss of another group. However, when the sample

ℓ	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$
$\{1\}$	0.9770	2.5165	1.1696	1.9923	1.5728	1.9181	1.2968	1.4727
$\{2\}$	0.6992	2.2214	1.0189	1.8331	0.9898	1.3299	1.2472	1.4217
$\{3\}$	0.8581	2.3926	0.8132	1.6237	1.1060	1.4486	0.9449	1.1185
$\{4\}$	0.4587	1.9691	0.6309	1.4369	0.6251	0.9591	0.9281	1.1007

Table 3.10: NPI lower and upper expected linear losses of selecting best group, where the event is defined as $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ for Example 3.5.2.

size increases to $n = 50$, the NPI upper expected linear loss of group 4, is less than the NPI lower expected linear loss of groups $1, 2$, and 3, meaning that group 4 is strongly preferred over these groups. For $n = 100$, the NPI upper expected linear loss of group 4 is less than the NPI lower expected linear loss of groups 1 and 2, but not group 3. Therefore, group 4 is strongly preferred over groups 1 and 2 in this case. Similar findings are observed in Table 3.13.

Table 3.11 shows that for sample sizes $n = 10$ and $n = 20$, no strong preference is observed among the subsets of the best groups. However, when the sample size increases to $n = 50$, the NPI upper expected linear loss of the subset $\{2, 3, 4\}$ is less than the NPI lower expected linear loss of all other subsets except $\{2, 4\}$, indicating that $\{2, 3, 4\}$ is strongly preferred over the remaining subsets. Furthermore, for $n = 100$, the NPI upper expected linear loss of the subset $\{2, 3, 4\}$ is less than the NPI lower expected linear loss of all subsets except $\{3, 4\}$ and $\{1, 3, 4\}$.

To avoid repetition, the same approach is followed for Tables 3.12, 3.14, and 3.15. In conclusion, the results consistently show that group 4 is the best selection across all sample sizes, with the subset $\{2, 3, 4\}$ performing best in subset selection scenarios. The observed reduction in uncertainty with increasing sample size under both linear and quadratic loss functions.

S	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$
$\{1,2\}$	1.1285	2.7800	1.3530	2.2240	1.6230	1.9828	1.5832	1.7658
$\{1,3\}$	1.2486	2.8850	1.2065	2.0869	1.7215	2.0789	1.3497	1.5347
$\{1,4\}$	0.9011	2.5687	1.0527	1.9403	1.3444	1.7110	1.3308	1.5161
$\{2,3\}$	1.0394	2.6943	1.0797	1.9662	1.2679	1.6356	1.3046	1.4901
$\{2,4\}$	0.6676	2.3429	0.9153	1.8067	0.8221	1.1935	1.2824	1.4681
$\{3,4\}$	0.8164	2.4894	0.7324	1.6255	0.9352	1.3053	1.0232	1.2102
$\{1,2,3\}$	1.1306	2.7934	1.1210	2.0213	1.5192	1.8883	1.2689	1.4531
$\{1,2,4\}$	0.7333	2.3680	0.9404	1.8329	1.0362	1.4008	1.2513	1.4361
$\{1,3,4\}$	0.8961	2.5361	0.7367	1.6218	1.1531	1.5189	0.9500	1.132
$\{2,3,4\}$	0.6139	2.2446	0.5823	1.4660	0.5725	0.9284	0.8994	1.0819

Table 3.11: NPI lower and upper expected linear losses of selecting the subset of best groups, where the event is defined as $\min\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$ for Example 3.5.2.

S	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_l)$	$\overline{E}(L_l)$
$\{1,2\}$	0.6387	1.5557	0.6603	1.1418	0.6147	0.8099	0.7480	0.8487
$\{1,3\}$	0.4310	1.3045	0.5703	1.0415	0.4597	0.6467	0.5694	0.6649
$\{1,4\}$	0.5245	1.4204	0.2733	0.6921	0.5043	0.6924	0.5065	0.6002
$\{2,3\}$	0.3519	1.1948	0.5826	1.0573	0.3988	0.5811	0.4619	0.5535
$\{2,4\}$	0.4414	1.3068	0.2838	0.7062	0.4365	0.6199	0.4081	0.4979
$\{3,4\}$	0.2852	1.1071	0.2076	0.6197	0.3162	0.4916	0.2786	0.3631
$\{1,2,3\}$	0.1909	0.6480	0.3484	0.6182	0.2036	0.2994	0.2917	0.3431
$\{1,2,4\}$	0.2575	0.7482	0.1316	0.3474	0.2317	0.3303	0.2544	0.3039
$\{1,3,4\}$	0.1011	0.5202	0.0734	0.2745	0.1472	0.2367	0.1478	0.1920
$\{2,3,4\}$	0.0723	0.4665	0.0797	0.2828	0.1157	0.2014	0.0859	0.1267

Table 3.12: NPI lower and upper expected linear losses of selecting a subset including the best group, where the event is defined as $\max\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$ for Example 3.5.2.

ℓ	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1\}$	1.6189	11.6665	2.2958	7.5086	3.6725	5.9530	2.9658	4.0525
$\{2\}$	1.3855	10.8252	1.9332	6.9806	1.9716	3.9797	2.7934	3.8706
$\{3\}$	1.6131	11.4092	1.5053	6.2983	2.3729	4.4363	1.9031	2.8982
$\{4\}$	0.5717	9.4787	0.9202	5.5084	1.0385	2.8588	1.8822	2.8819

Table 3.13: NPI lower and upper expected quadratic losses of selecting best group, where the event is defined as $X_\ell^f \geq \max_{j \neq \ell} X_j^f$ for Example 3.5.2.

S	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1, 2\}$	2.1833	13.9570	2.6994	8.8263	3.7238	6.2609	3.7977	5.0579
$\{1, 3\}$	2.4094	14.3368	2.4340	8.4552	4.1976	6.7605	3.1130	4.3350
$\{1, 4\}$	1.3558	12.8326	1.9087	7.8176	2.8229	5.2756	3.0179	4.2371
$\{2, 3\}$	2.1765	13.7849	2.0479	7.9678	2.7929	5.2200	2.8549	4.0699
$\{2, 4\}$	1.1661	12.1987	1.5124	7.2965	1.4458	3.6976	2.7390	3.9501
$\{3, 4\}$	1.4037	12.7230	1.2427	6.8381	1.8701	4.1643	2.0716	3.2323
$\{1, 2, 3\}$	2.3397	13.8047	2.1440	8.1053	3.5733	6.0721	2.9244	4.0950
$\{1, 2, 4\}$	1.2253	11.7027	1.6444	7.3624	1.9534	4.1984	2.7357	3.9012
$\{1, 3, 4\}$	1.5648	12.4727	1.3524	6.7892	2.4675	4.7732	2.0574	3.1407
$\{2, 3, 4\}$	1.0848	11.2143	0.8684	6.0858	1.0255	2.9947	1.6915	2.7611

Table 3.14: NPI lower and upper expected quadratic losses of selecting the subset of best groups, where the event is defined as $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$ for Example 3.5.2.

S	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1,2\}$	0.3938	4.1715	0.8661	2.8738	1.0957	1.9320	1.1206	1.5471
$\{1,3\}$	0.5509	4.5313	0.7630	2.6564	1.2808	2.1484	0.9194	1.3161
$\{1,4\}$	0.2559	3.8141	0.5623	2.3708	0.6008	1.3381	0.8360	1.2262
$\{2,3\}$	0.3383	4.0115	0.5270	2.3215	0.7454	1.4979	0.7213	1.1079
$\{2,4\}$	0.1145	3.4067	0.3496	2.0669	0.2813	0.9274	0.6430	1.0233
$\{3,4\}$	0.1967	3.6460	0.3438	1.9788	0.4032	1.0758	0.5251	0.8803
$\{1,2,3\}$	0.1640	1.8271	0.2876	1.1253	0.4873	0.8583	0.3712	0.5498
$\{1,2,4\}$	0.0404	1.5218	0.1826	0.9797	0.1380	0.4448	0.2893	0.4616
$\{1,3,4\}$	0.0937	1.6636	0.2052	0.9778	0.2432	0.5675	0.2717	0.4364
$\{2,3,4\}$	0.0303	1.4731	0.0638	0.7754	0.0737	0.3589	0.1127	0.2678

Table 3.15: NPI lower and upper expected quadratic losses of selecting a subset including the best group, where the event is defined as $\max\limits_{\ell \in S} X_\ell^f \geq \max\limits_{j \in NS} X_j^f$ for Example 3.5.2.

$\diamond$

3.6 Concluding remarks

A general overview of loss functions was presented, highlighting how they quantify the impact of selecting suboptimal groups and guide the identification of the best ones, particularly in selection processes. The discussion also covered how different types of loss functions evaluate the consequences of suboptimal choices, and introduced the NPI lower and upper expected losses.

Two main inferences were considered: pairwise comparison and multiple comparisons based on three types of loss functions. For pairwise comparisons, the impact of selecting suboptimal groups was evaluated using three loss functions: zero-one, linear, and quadratic loss functions. Through various examples, the conclusions derived from the zero-one, linear, and quadratic loss functions are consistent. The linear and quadratic loss functions are more accurate and informative than the zero-

one loss function. While the zero-one loss simply indicates whether a selection is correct or not, it does not quantify the exact loss. However, the linear loss function provides a measure of the amount of the loss, capturing how far off the group that provides the largest future observation is from the rest. The quadratic loss function measures the squared difference between correct and incorrect choices. This function penalises larger and more heavily as it squares the deviation.

The zero-one loss function offers several advantages, primarily its direct relationship with the NPI lower and upper probabilities, where the NPI lower expected zero-one loss for an event equals the NPI lower probability of the complement event. This property simplifies computation, enabling its application to a range of problems involving NPI lower and upper probabilities. However, the zero-one loss function only distinguishes between correct and incorrect selections.

Extending the evaluation to multiple comparisons involves three different events: selecting the best group, selecting a subset of the best groups, and selecting a subset that includes the best group, all based on the loss functions mentioned above. The examples provided in this chapter cover both real-valued and simulated data, providing an investigation of how the NPI method performs in various situations. Once again, our findings show that the zero-one, linear, and quadratic loss functions generally yield consistent results, though minor variations in selection may occur depending on the loss function used.

The NPI lower and upper expected losses provide a framework for selecting the best group, selecting a subset of the best groups, or selecting a subset that includes the best group. These decisions can be made in terms of strong and weak preference or by identifying the lowest NPI lower and upper expected losses to determine the best selection. The event associated with the lowest NPI lower and upper expected losses is considered the best choice.

Strong preference occurs when the NPI upper expected loss of one event is less than the NPI lower expected loss of another. In this case, the selected group, a subset of the best groups, or a subset that include the best group is strongly preferred over the others. Weak preference occurs when both the NPI lower and upper expected losses of selecting the best group, the subset of the best groups, or the subset that

includes the best group are lower than the corresponding losses for another such selection. This indicates that while one selection is preferred, alternative choices may still be reasonable.

The zero-one loss function allows unbounded ranges for future observations. In contrast, linear and quadratic loss functions require bounded ranges, as unbounded values would affect the calculation of NPI lower and upper expected losses.

Although this work focuses on zero-one, linear, and quadratic loss functions, future research could extend the analysis by exploring other types of loss function, which may offer additional insights into the decision-making process. The choice of loss function may influence which groups are selected, potentially resulting in different subsets being identified as optimal.

Chapter 4

NPI for ranking subset using loss functions

4.1 Introduction

In decision-making, especially when comparing multiple populations, classical statistical methods like hypothesis testing have limitations. While useful for assessing population equivalence, they fall short when the objective is to rank populations such as identifying the largest, second-largest, and so on. In such cases, selecting a ranked subset of the best groups becomes essential to support informed decisions. Additionally, concepts such as loss functions can be used to quantify the loss incurred when the ranking of selected groups deviates from the optimal ranking.

This chapter considers a similar scenario to that in Section 2.2, but extends the focus beyond selecting a ranked subset of the best groups. The objective is to use a loss function to guide the ranking of future observations within the selected subset and to evaluate the quality of that ranking.

Two different loss functions, the zero-one loss function and the general multi-level loss function, will be applied to scenarios involving the ranking of the future observations from a ranked subset of three or more best groups. Additionally, a special case involving the selection of a ranked subset of the two best groups will be considered. The multi-level loss function will be used to guide their ranking and assess its quality.

For the zero-one loss function, the focus is on evaluating whether the groups are ranked correctly. In contrast, the general multi-level loss function is used to guide the selection of a ranked subset of the best groups and to assess the quality of the ranking. The general multi-level loss function provides a fine-grained evaluation by assigning different penalties for different types of errors, where ranking mistakes can have different levels of impact. The zero-one loss function can be considered to be a special case of the general multi-level loss function where there is a single level of loss, i.e., a binary distinction between correct and incorrect rankings.

4.2 Ranked subset of multiple best groups using zero-one loss function

The zero-one loss function in this section evaluates whether the groups in the subset S are correctly ranked based on their future observations in a binary manner: the loss is $L_{\{0,1\}} = 0$ if the ranking is correct, and $L_{\{0,1\}} = 1$ if it is not. It assigns a fixed penalty for any incorrect ranking, without distinguishing between the type of the error.

Consider a subset of all groups, consisting of w groups out of k independent groups, with $1 \leq w \leq k-1$. Let $S = \{\ell_1, \ell_2, \ldots, \ell_w\} \subset \{1, 2, \ldots, k\}$, and let $NS = \{1, 2, \ldots, k\} \backslash S$, so NS is the set of $k-w$ non-selected groups. Suppose that the event of interest is $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \cdots < X_{\ell_1}^f$. The zero-one loss function for a ranked subset of w groups is

$$
L_{\{0,1\}}\big(\max_{j_z \in NS} X_{j_z}^f, X_{\ell_w}^f, X_{\ell_{w-1}}^f, \ldots, X_{\ell_1}^f\big) = \begin{cases} 0 & \text{if } \max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < X_{\ell_{w-1}}^f < \cdots < X_{\ell_1}^f \\ 1 & \text{if otherwise} \end{cases}
$$

(4.1)

The term 'otherwise' implies that the zero-one loss function assigns a value of one for all events involving incorrect ranking of future observations within selected subset, except when $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < X_{\ell_{w-1}}^f < \cdots < X_{\ell_1}^f$. This exception corresponds to correct ranking of future observations within the selected subset. The total number of events for incorrect ranking of the w best groups within subset is $(w+1)! - 1$.

This accounts for all permutations of the groups that belong to S, denoted by ℓ_w to ℓ_1 and the maximum future observation that belongs to NS.

As discussed in Section 3.4.1, the NPI lower and upper expected zero-one losses are directly linked to the NPI lower and upper probabilities. Specifically, the NPI lower expected zero-one loss for an event is equal to the NPI lower probability of the complement event. The same holds for the NPI upper expected loss and the NPI upper probability. This relationship provides a straightforward interpretation and allows the expected losses to be determined directly from the corresponding probabilities.

Sections 4.3 and 4.4 present the multi-level loss function for two cases: when the ranked subset contains two and three groups. These cases are important, as they demonstrate how the NPI lower and upper expected multi-level losses are derived differently depending on the subset size. For $w = 2$, the NPI lower and upper expected losses are derived, while for $w = 3$, NPI lower and upper expected losses are derived.

4.3 Ranked subset of two best groups using multi-level loss function

The objective of this section is to apply a multi-level loss function to assign graded penalties for different levels of error in selecting a ranked subset of the two best groups. The primary difference between the zero-one loss function and the multi-level loss function is how they penalise incorrect ranking of subset groups. The number of possible ways to rank two groups that belong to S, denoted by ℓ_1 and ℓ_2, together with the maximum future observation among groups in NS is 3!.

4.3. Ranked subset of two best groups using multi-level loss function 100

The multi-level loss function for the event of interest, $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$ is

$$L_{ML}(\max_{j_z \in NS} X_{j_z}^f, X_{\ell_2}^f, X_{\ell_1}^f) = \begin{cases} L_0 & \text{if } \max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f \\ L_1 & \text{if } \max_{j_z \in NS} X_{j_z}^f < X_{\ell_1}^f < X_{\ell_2}^f \\ L_2 & \text{if } X_{\ell_2}^f < \max_{j_z \in NS} X_{j_z}^f < X_{\ell_1}^f \\ L_3 & \text{if } X_{\ell_1}^f < \max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f \\ L_4 & \text{if } X_{\ell_2}^f < X_{\ell_1}^f < \max_{j_z \in NS} X_{j_z}^f \\ L_5 & \text{if } X_{\ell_1}^f < X_{\ell_2}^f < \max_{j_z \in NS} X_{j_z}^f \end{cases} \quad (4.2)$$

The multi-level loss function L_{ML} is more nuanced than the zero-one loss function, assigning different loss values to various ranking scenarios. It provides graded penalties for different levels of errors, with zero loss for L_0 and specific loss values from L_1 to L_5 based on the nature of the error. For example, if $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_1}^f < X_{\ell_2}^f$, the loss is L_1, which could be a smaller penalty compared to $X_{\ell_1}^f < X_{\ell_2}^f < \max_{j_z \in NS} X_{j_z}^f$, which faces a loss of L_5. This allows the evaluation to be more sensitive to the severity of errors in the ranking of the groups within the subset than the zero-one loss function. This can be particularly useful if certain selection and ranking errors are more costly than others in a specific application. For example, if the consequences of predicting $X_{\ell_1}^f < X_{\ell_2}^f < \max_{j_z \in NS} X_{j_z}^f$ are worse than those of predicting $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_1}^f < X_{\ell_2}^f$, one would aim to avoid the mistake associated with the largest penalty, denoted as L_5.

To introduce the NPI lower and upper expected multi-level losses for the event of interest, $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$, NPI is applied to each group using Hill's assumption $A_{(n)}$. Note that all the notation used in this section was provided earlier in Section 2.3. To derive the NPI lower expected multi-level loss, the probability mass corresponding to ℓ_1 is assigned to the right endpoint per interval. The probability masses corresponding to $j_z \in NS$ are assigned to the left endpoint per interval.

For the middle group ℓ_2, the corresponding probability mass is assigned to a chosen sub-interval within the ℓ_2 intervals, denoted as $(x_{\ell_2, u_{\ell_2} - 1}, x_{\ell_2, u_{\ell_2}})$, where $u_{\ell_2} = 1, 2, \ldots, n_{\ell_2} + 1$, which minimise the NPI lower expected multi-level loss.

There is a total of $\prod_{t=1}^{k}(n_t + 1)$ combinations of orderings (intervals) of future

observations from all k groups for the event $\max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}$. For each combination of orderings, it is necessary to determine which of $L_0, \ldots, L_5$ is actually possible, which is the minimum possible loss. The minimum possible loss occurs when group ℓ_1 is as large as possible and each group belonging to NS is as small as possible. Then, the probability $\dfrac{1}{\prod_{t=1}^{k}(n_t + 1)}$ is assigned to the actual possible loss L_c, where $c = 0, \ldots, 5$, the value that minimises the NPI lower expected multi-level loss. The process iterates over all orderings to determine the minimum possible losses and, therefore, the NPI lower expected multi-level loss denoted by

$$\underline{E}(L_{ML}(\max_{j_z \in NS} X^f_{j_z}, X^f_{\ell_2}, X^f_{\ell_1})) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t+1)} \left[\min_c L_c \,|O_i \right] \qquad (4.3)$$

where O_i is a specific ordering for which L_c is the minimum possible loss and $c = 0, \ldots, 5$ and $i = 1, \ldots, \prod_{t=1}^{k}(n_t + 1)$.

To derive the NPI upper expected multi-level loss, the probability mass corresponding to group ℓ_1 is assigned to the left endpoint per interval $(x_{\ell_1, u_{\ell_1}-1}, x_{\ell_1, u_{\ell_1}})$. The probability mass corresponding to each $j_z \in NS$ is assigned to the right endpoint per interval. For the group in the middle ℓ_2, the corresponding probability mass is assigned to a chosen sub-interval within ℓ_2 intervals, denoted as $(x_{\ell_2, u_{\ell_2}-1}, x_{\ell_2, u_{\ell_2}})$, $u_{\ell_2} = 1, 2, \ldots, n_{\ell_2} + 1$ that maximises the NPI upper expected multi-level loss.

There is a total of $\prod_{t=1}^{k}(n_t + 1)$ combinations of orderings of future observations from all k groups. For each combination of orderings, it is necessary to determine which of $L_0, \ldots, L_5$ is actually possible, which is the maximum possible loss. The maximum possible loss occurs when the group ℓ_1 is as small as possible and each group $j_z \in NS$ is as large as possible. Then, the probability $\dfrac{1}{\prod_{t=1}^{k}(n_t + 1)}$ is assigned to the actual possible loss L_c value that maximises the NPI upper expected loss. Again, the process repeated over all orderings to determine the maximum possible loss and the NPI upper expected multi-level loss is denoted by

$$\overline{E}(L_{ML}(\max_{j_z \in NS} X^f_{j_z}, X^f_{\ell_2}, X^f_{\ell_1}) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t+1)} \left[\max_c L_c \,|O_i \right] \qquad (4.4)$$

Example 4.3.1 illustrates the method using data from the literature, while Example 4.3.2 applies the same approach to a large dataset simulated from Normal distributions.

ℓ_1	ℓ_2	$\underline{E}(L_{ML})$ and $\overline{E}(L_{ML})$
1	2	$\underline{E}(L) : 0.1540L_0 + 0.0587L_1 + 0.1730L_2 + 0.1682L_3 + 0.2397L_4 + 0.2063L_5$
		$\overline{E}(L) : 0.0167L_0 + 0.0302L_1 + 0.0484L_2 + 0.0960L_3 + 0.2683L_4 + 0.5405L_5$
2	1	$\underline{E}(L) : 0.1317L_0 + 0.0857L_1 + 0.1714L_2 + 0.1682L_3 + 0.2968L_4 + 0.1460L_5$
		$\overline{E}(L) : 0.0063L_0 + 0.0452L_1 + 0.0595L_2 + 0.0722L_3 + 0.4008L_4 + 0.4159L_5$
1	3	$\underline{E}(L) : 0.1889L_0 + 0.1270L_1 + 0.1365L_2 + 0.2913L_3 + 0.1079L_4 + 0.1484L_5$
		$\overline{E}(L) : 0.0071L_0 + 0.0825L_1 + 0.0579L_2 + 0.2238L_3 + 0.2016L_4 + 0.4270L_5$
3	1	$\underline{E}(L) : 0.2302L_0 + 0.0548L_1 + 0.3222L_2 + 0.0865L_3 + 0.1810L_4 + 0.1254L_5$
		$\overline{E}(L) : 0.0278L_0 + 0.0619L_1 + 0.1960L_2 + 0.1056L_3 + 0.3095L_4 + 0.2992L_5$
1	4	$\underline{E}(L) : 0.1484L_0 + 0.0841L_1 + 0.1952L_2 + 0.2079L_3 + 0.2278L_4 + 0.1365L_5$
		$\overline{E}(L) : 0.0032L_0 + 0.0540L_1 + 0.0651L_2 + 0.1690L_3 + 0.2460L_4 + 0.4627L_5$
4	1	$\underline{E}(L) : 0.1913L_0 + 0.0460L_1 + 0.2389L_2 + 0.1571L_3 + 0.2270L_4 + 0.1397L_5$
		$\overline{E}(L) : 0.0151L_0 + 0.0492L_1 + 0.1341L_2 + 0.0968L_3 + 0.3079L_4 + 0.3968L_5$
2	3	$\underline{E}(L) : 0.1921L_0 + 0.1524L_1 + 0.1270L_2 + 0.2175L_3 + 0.2087L_4 + 0.1024L_5$
		$\overline{E}(L) : 0.0103L_0 + 0.0992L_1 + 0.0524L_2 + 0.1857L_3 + 0.2151L_4 + 0.4373L_5$
3	2	$\underline{E}(L) : 0.2881L_0 + 0.0603L_1 + 0.2754L_2 + 0.0754L_3 + 0.1865L_4 + 0.1143L_5$
		$\overline{E}(L) : 0.0508L_0 + 0.0738L_1 + 0.1563L_2 + 0.1087L_3 + 0.2651L_4 + 0.3452L_5$
2	4	$\underline{E}(L) : 0.1500L_0 + 0.1103L_1 + 0.1675L_2 + 0.1778L_3 + 0.2778L_4 + 0.1167L_5$
		$\overline{E}(L) : 0.0079L_0 + 0.0556L_1 + 0.0587L_2 + 0.1460L_3 + 0.3381L_4 + 0.3937L_5$
4	2	$\underline{E}(L) : 0.2071L_0 + 0.0540L_1 + 0.2341L_2 + 0.1167L_3 + 0.1778L_4 + 0.2103L_5$
		$\overline{E}(L) : 0.0270L_0 + 0.0452L_1 + 0.1230L_2 + 0.0889L_3 + 0.2786L_4 + 0.4373L_5$
3	4	$\underline{E}(L) : 0.2794L_0 + 0.1333L_1 + 0.2595L_2 + 0.1683L_3 + 0.1040L_4 + 0.0556L_5$
		$\overline{E}(L) : 0.0413L_0 + 0.1016L_1 + 0.1683L_2 + 0.1571L_3 + 0.2429L_4 + 0.2889L_5$
4	3	$\underline{E}(L) : 0.2698L_0 + 0.1429L_1 + 0.1680L_2 + 0.2595L_3 + 0.0857L_4 + 0.0738L_5$
		$\overline{E}(L) : 0.0286L_0 + 0.1016L_1 + 0.1317L_2 + 0.1810L_3 + 0.2206L_4 + 0.3365L_5$

Table 4.1: The NPI lower and upper expected multi-level losses for all groups combinations for Example 4.3.1

Example 4.3.1 The dataset used here is the same as in Example 2.3.1. The aim is to illustrate the selection of a ranked subset of the two best groups by using a

		Case 1		Case 2		Case 3	
ℓ_1	ℓ_2	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.2900	0.4190	0.5130	0.8234	0.6807	0.8941
2	1	0.2851	0.4063	0.5065	0.8147	0.6721	0.8783
1	3	0.2448	0.3811	0.3729	0.6954	0.5805	0.8321
3	1	0.2310	0.3505	0.3841	0.6548	0.5858	0.8025
1	4	0.2692	0.3989	0.4513	0.7533	0.6487	0.8676
4	1	0.2602	0.3824	0.4435	0.7348	0.6392	0.8478
2	3	0.2406	0.3808	0.3961	0.7070	0.5607	0.8211
3	2	0.2155	0.3499	0.3659	0.6551	0.5383	0.7839
2	4	0.2673	0.3932	0.4645	0.7590	0.6316	0.8587
4	2	0.2635	0.3859	0.4575	0.7438	0.6302	0.8475
3	4	0.1851	0.3325	0.2648	0.5984	0.4721	0.7560
4	3	0.1970	0.3473	0.2767	0.6258	0.4835	0.7771

Table 4.2: The NPI lower and upper expected multi-level losses for different loss values for Example 4.3.1. Case 1: $L_0 = 0.00$, $L_1 = 0.10$, $L_2 = 0.20$, $L_3 = 0.30$, $L_4 = 0.40$ and $L_5 = 0.50$; Case 2: $L_0 = 0.00$, $L_1 = 0.10$, $L_2 = 0.20$, $L_3 = 0.30$, $L_4 = 0.90$, and $L_5 = 1.00$; Case 3: $L_0 = 0.00$, $L_1 = 0.05$, $L_2 = 0.70$, $L_3 = 0.80$, $L_4 = 0.90$ and $L_5 = 1.00$

multi-level loss function, where the event of interest is $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_2}^f < X_{\ell_1}^f$. In this example, there are 1260 possible combinations of orderings of future observations from all 4 groups, and 12 different possible combinations of selecting a ranked subset of the two best groups. These 12 combinations are presented in Table 4.1. The purpose is to provide a graded penalty for different levels of mistakes in the selection of a ranked subset of the two best groups using a multi-level loss function.

To derive the NPI lower expected multi-level loss, it is necessary to go through each possible combination of orderings of future observations and determine the minimum possible loss. Starting with the first case, it is assumed that the groups in the subset are 1 as the best group and 2 as the second best group, and the groups 3 and 4 belong to NS. The minimum loss occurs if the best selected group 1 is as

		Case 4		Case 5		Case 6	
ℓ_1	ℓ_2	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.5385	0.7724	0.8460	0.9833	0.7873	0.9532
2	1	0.5211	0.7188	0.8682	0.9937	0.7825	0.9484
1	3	0.4902	0.7210	0.8111	0.9929	0.6841	0.9103
3	1	0.4580	0.6491	0.7698	0.9722	0.7151	0.9103
1	4	0.5070	0.7412	0.8516	0.9968	0.7675	0.9429
4	1	0.4865	0.7034	0.8087	0.9849	0.7627	0.9357
2	3	0.4670	0.7242	0.8079	0.9897	0.6556	0.8905
3	2	0.4230	0.6589	0.7119	0.9492	0.6516	0.8754
2	4	0.4969	0.7065	0.8500	0.9921	0.7397	0.9365
4	2	0.5116	0.7165	0.7929	0.9730	0.7389	0.9278
3	4	0.3972	0.6359	0.7206	0.9587	0.5873	0.8571
4	3	0.4114	0.6655	0.7302	0.9714	0.5873	0.8698

Table 4.3: The NPI lower and upper expected multi-level losses for different loss values for Example 4.3.1. Case 4: $L_0 = 0.00$, $L_1 = 0.50$, $L_2 = 0.51$, $L_3 = 0.52$, $L_4 = 0.53$ and $L_5 = 1.00$; Case 5: $L_0 = 0.00$, $L_1 = 1.00$, $L_2 = 1.00$, $L_3 = 1.00$, $L_4 = 1.00$ and $L_5 = 1.00$; Case 6: $L_0 = 0.00$, $L_1 = 0.00$, $L_2 = 1.00$, $L_3 = 1.00$, $L_4 = 1.00$ and $L_5 = 1.00$.

large as possible and the non-selected groups 3 and 4 are as small as possible. For example, for selecting a ranked subset of group 1 as the best group and group 2 as the second best group, starting with the first interval of each group, the probability mass of group 1 is assigned to $x_{1,1} = 1$, and the probability masses of groups 3 and 4 are assigned to $\max_{j_z \in NS}(x_{3,0}, x_{4,0}) = -\infty$. For group 2, its corresponding probability mass is assigned to any point (suppose the midpoint), excluding any ties that are identical to data points in other groups, of a chosen sub-interval that minimises the NPI lower expected multi-level loss, which is $(3, 4)$. Therefore, for the first combination of orderings, L_1 is the minimum possible loss because the best selected group 1 is as large as possible and the non-selected groups 3 and 4 are as small as possible. The probability mass $\dfrac{1}{1260}$ is assigned to L_1 for $\underline{E}(L_{ML})$. For L_1, after

		Case 7		Case 8		Case 9	
ℓ_1	ℓ_2	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.6142	0.9048	0.4460	0.8087	0.2063	0.5405
2	1	0.6111	0.8889	0.4429	0.8167	0.1460	0.4159
1	3	0.5476	0.8524	0.2563	0.6286	0.1484	0.4270
3	1	0.3929	0.7143	0.3063	0.6087	0.1254	0.2992
1	4	0.5722	0.8778	0.3643	0.7087	0.1365	0.4627
4	1	0.5238	0.8016	0.3667	0.7048	0.1397	0.3968
2	3	0.5286	0.8381	0.3111	0.6524	0.1024	0.4373
3	2	0.3762	0.7190	0.3008	0.6103	0.1143	0.3452
2	4	0.5722	0.8778	0.3944	0.7317	0.1167	0.3937
4	2	0.5048	0.8048	0.3881	0.7159	0.2103	0.4373
3	4	0.3278	0.6889	0.1595	0.5317	0.0556	0.2889
4	3	0.4190	0.7381	0.1595	0.5571	0.0738	0.3365

Table 4.4: The NPI lower and upper expected multi-level losses for different loss values for Example 4.3.1. Case 7: $L_0 = 0.00$, $L_1 = 0.00$, $L_2 = 0.00$, $L_3 = 1.00$, $L_4 = 1.00$ and $L_5 = 1.00$. Case 8: $L_0 = 0.00$, $L_1 = 0.00$, $L_2 = 0.00$, $L_3 = 0.00$, $L_4 = 1.00$ and $L_5 = 1.00$. Case 9: $L_0 = 0.00$, $L_1 = 0.00$, $L_2 = 0.00$, $L_3 = 0.00$, $L_4 = 0.00$ and $L_5 = 1.00$.

going through 1260 possible combinations of ordering of future observations, the total number of assigned probability masses is 74. Thus, C_1, which is the coefficient of L_1, is $\dfrac{74}{1260} = 0.0587$ as presented in Table 4.1.

To derive the NPI upper expected multi-level loss, the process is similar, but the maximum possible loss is considered. The maximum loss occurs if the best selected group 1 is as small as possible and the non-selected groups 3 and 4 are as large as possible. For the ranked subset where group 1 is the best and group 2 is the second best, starting with the first interval of each group, the probability mass of group 1 is assigned to $x_{1,0} = -\infty$, and the probability masses for groups 3 and 4 are assigned to $\max_{j_z \in NS}(x_{3,1}, x_{4,1}) = 3$. For group 2, its corresponding probability mass is assigned to any point (suppose the midpoint), excluding any ties that are

identical to data points in other groups, of a chosen sub-interval that maximises the NPI upper expected multi-level loss, which is $(-\infty, 3)$. Therefore, for the first combination of orderings, L_5 is the maximum possible loss. For L_5, after going through 1260 possible combinations of ordering of future observations, the total number of assigned probability masses is 681. Thus, C_5, which is the coefficient of L_5, is $\dfrac{681}{1260} = 0.5405$ as presented in Table 4.1.

Repeat this process for the rest of ordering combinations of future observations to get the NPI lower and upper expected multi-level losses for a ranked subset of selecting group 1 as the best group and group 2 as the second best. Apply the same process to all 11 remaining combinations of ranking two groups belonging to S.

Table 4.1 presents equations that are a linear combination of loss values and coefficients to calculate the NPI lower and upper expected multi-level loss values associated with 12 different combinations. The NPI lower and upper expected losses presented in this table are calculated using the method described in the two paragraphs provided earlier. Furthermore, each of these equations involves six coefficients C_0 through C_5, corresponding to loss values L_0 through L_5 respectively. C_c represents how many times L_c is indeed the actual possible loss.

Several key findings of Table 4.1 can be made, which are: coefficients corresponding to L_0 are consistently greater than those associated with L_1 for the NPI lower expected multi-level loss for all different combinations of selecting a ranked subset of the best two groups. This implies that the loss component represented by L_0 carries more weight in the calculation of the NPI lower expected multi-level loss compared to L_1 across all 12 combinations. Conversely, coefficients corresponding to L_1 are consistently greater than those associated with L_0 for the NPI upper expected multi-level loss. Moreover, for both NPI lower and upper expected multi-level losses, the sum of coefficients related to L_4 and L_5 is consistently greater than the sum of coefficients related to L_0 and L_1 across all 12 combinations. This suggests that the combined impact of L_4 and L_5 is more meaningful in determining the expected multi-level losses than that of L_0 and L_1.

For the NPI upper expected multi-level loss in Table 4.1, the coefficients related to L_1 are consistently small. This indicates that the contribution of L_1 to the NPI

upper expected multi-level loss is relatively minor compared to other loss components. In addition, the coefficients associated with L_4 are consistently greater than the coefficients associated with L_3 for all 12 combinations except selecting group 1 as the best and group 3 as the second best. This indicates that within the group of higher-loss components L_3 and L_4, L_4 has a more substantial influence on the NPI upper expected multi-level loss. In summary, the provided findings highlight the relative importance of different loss components and coefficients in the calculation of the NPI lower and upper expected multi-level losses, thereby offering valuable insights into the decision-making process.

In this example, some of these combinations have the same groups but in a different order, e.g., selecting group 1 as the best and group 2 as the second best group, while selecting group 2 as the best and group 1 as the second best group, have the same groups but in a different ranking, so out of the 12 combinations, 6 of them are unique combinations and the other 6 are just the same combinations with different group rankings within the subset. However, the loss coefficients of each similar two groups with respect to their ranking, say the coefficient of L_0 for selecting group 1 as the best and group 2 as the second best group and the coefficient of L_1 for selecting group 2 as the best and group 1 as the second best group are not interchangeable since the chosen sub-intervals for the middle group that minimise the NPI lower expected multi-level loss and maximise the NPI upper expected multi-level loss are not identical.

To understand the impact of different loss values, nine cases are considered, each with a unique set of values for the five loss levels, ranging from L_0 (no loss, best case) to L_5 (largest loss, worst case). These cases are reported in Tables 4.2, 4.3 and 4.4.

Table 4.2 presents three cases, each of which shows varying loss values. In Case 1, the chances of incurring losses are relatively low less than or equal 0.50. Case 2 demonstrates a higher loss potential in L_4 and L_5, while Case 3 shows relatively low loss values for L_0 to L_1. Table 4.3 presents an additional set of three cases. For Case 4, the loss values are defined as set at approximately 0.5. The last set of three cases is presented in Table 4.4. For Case 7, all loss values from L_0 to L_2 are set to

0 and the remaining loss values are set to 1. Case 8 sets all loss values from L_0 to L_3 to 0, while the rest are set to 1. Lastly, Case 9 explores a scenario in which loss values from L_0 to L_4 are set to zero and L_5 is set to one.

The chosen cases were assumed to explore various scenarios in which the loss values change, which range from low losses to higher losses. These values were arbitrarily chosen, and although many more cases could be considered, these are sufficient for our purposes. In addition, in some cases, the loss value equals zero, to show if the absence of loss in specific values affects the overall outcome, providing comparisons across different loss values.

Tables 4.2, 4.3 and 4.4 show that for each case the NPI lower and upper expected multi-level losses vary across different combinations of selected ranked groups due to different loss values. For the first row for Case 1 in Table 4.2, the NPI lower expected multi-level loss for selecting group 1 as the best and group 2 as the second best is 0.2900 and the NPI upper expected multi-level loss is 0.4190.

First, Cases 1, 2, 3, 4, 6, 7 and 8 are summarised. Despite the differences in the NPI lower and upper expected multi-level losses across those cases, the conclusions drawn from them remain consistent. From the comparisons for those cases, it is noticed that the combination of group 3 as the best and group 4 as the second best has the lowest NPI lower and upper expected multi-level losses and the combination of selecting group 1 as the best and 2 as the second best has the highest NPI lower and upper expected multi-level losses.

In Case 6, the NPI lower expected multi-level loss for the ranking preferences of selecting group 3 as the best and 4 as the second best and selecting group 4 as the best and 3 as the second best are identical. Similarly, the NPI upper expected multi-level loss for the ranking preferences selecting group 3 as the best and 1 as the second best and 1 as the best and selecting group 3 as the second best are identical as well, which are 0.5873 and 0.9103, respectively. Moreover, depending on Cases 1, 2, 3, 4, 6, 7, and 8, it can be said that a stable ranked subset selection is achieved, but there are some combination results that change.

Table 4.3 presents the NPI lower and upper expected multi-level losses for Case 5. These results are consistent with those shown in Table 2.1, as they follow a

direct relationship between the NPI lower and upper expected zero-one losses and the corresponding NPI lower and upper probabilities, as defined by Equations (3.13) and (3.14). It is evident that the NPI lower and upper expected multi-level losses when selecting group 3 as the best and 2 as the second best are the smallest. To clarify, for the NPI lower expected multi-level loss, the the coefficient of L_0 for the previous combination is the highest, also, since $L_0 = 0$ in this case, that term becomes irrelevant and has no impact on the overall expression. However, selecting group 2 as the best followed by selecting group 4 as the second best shows higher NPI lower and upper expected multi-level loss values when compared to selecting group 1 as the best followed by selecting group 2 as the second best. The NPI lower and upper expected multi-level losses when group 1 is ranked as the best and group 2 as the second best show a slight deviation from all other NPI lower and upper expected multi-level losses in the remaining cases. All of them show that the highest NPI lower and upper expected multi-level losses correspond to the combination of selecting group 1 as the best, followed by selecting group 2 as the second best.

In Case 7, the NPI lower and upper expected multi-level loss values for both selecting group 1 as the best and group 4 as the second best and selecting group 2 as the best and group 4 as the second best are found to be identical, $\underline{E}(L_{ML}) = 0.5722$ and $\overline{E}(L_{ML}) = 0.8778$. In Case 8, the combination of selecting group 4 as the best and group 3 as the second best has the same lower expected multi-level loss as the combination of selecting group 3 as the best and group 4 as the second best. This is because the total coefficients of L_4 and L_5 in Table 4.1 are equal in both cases.

In Case 9, the combination of selecting group 4 as the best and group 2 as the second best results in the largest NPI lower expected multi-level loss, equal to 0.2103. On the other hand, the combination of selecting group 1 as the best and group 2 as the second best exhibits the largest NPI upper expected multi-level loss, which is 0.5405. It is noticeable that the NPI upper expected multi-level loss for the following ranking preferences: selecting group 2 as the best and group 3 as the second best, and selecting group 4 as the best and group 2 as the second best are identical.

In conclusion, the results emphasize that while a consistent ranked subset of the

two best groups can be selected, specific cases, such as Case 5 and Case 9, may show variability.

<div align="center">◇</div>

Example 4.3.2 This example presents a simulation with four groups, each sampled from Normal distributions with a standard deviation of 1 and a sample size of 25. The group means are 0.3, 0.2, 0.1, and 0.0, respectively. This study aims to provide a graded penalty for different levels of mistakes for selection of a ranked subset of the two best groups by using a multi-level loss function, with the primary objective of determining whether different conclusions can be drawn across nine different scenarios.

Table 4.5 presents equations that are a linear combination of loss values and coefficients to calculate the NPI lower and upper expected multi-level loss values associated with 12 different combinations. The NPI lower and upper expected losses presented in this table are calculated using the method described in Example 4.3.1.

Tables 4.6, 4.7 and 4.8 show the NPI lower and upper expected multi-level losses for nine cases to investigate the impact of different loss values on selecting a ranked subset of the two best groups. Similar loss values to those given in Example 4.3.1 are given here in this example, as listed in Tables 4.2, 4.3 and 4.4.

Tables 4.6, 4.7 and 4.8 show that the combinations of group selections yielding the smallest NPI lower and upper expected multi-level losses involve selecting group 2 as the best and group 1 as the second best. This trend is evident across cases, namely Case 1, 2 and 3. For the group preference when selecting group 1 as the best and 2 as the second best yields the smallest NPI lower and upper expected multi-level losses in Cases 6 and 8. For selecting group 2 as the best group and 4 as the second best group, it is evident that the lowest NPI lower and upper expected multi-level losses are observed in Case 7. Similarly, in the reverse ranking scenario where selecting group 4 as the best group and 2 as the second best group, the corresponding lowest NPI lower and upper expected multi-level losses occur in Case 5. In Case 4, two scenarios are considered: selecting 2 as the best and 1 as the second best, or selecting 2 as the best and 3 as the second best. The first scenario

ℓ_1	ℓ_2	$\underline{E}(L_{ML})$ and $\overline{E}(L_{ML})$
1	2	$\underline{E}(L) : 0.1208L_0 + 0.0978L_1 + 0.1732L_2 + 0.1671L_3 + 0.2119L_4 + 0.2292L_5$
		$\overline{E}(L) : 0.0743L_0 + 0.0928L_1 + 0.1472L_2 + 0.1620L_3 + 0.2340L_4 + 0.2896L_5$
2	1	$\underline{E}(L) : 0.1203L_0 + 0.0980L_1 + 0.1748L_2 + 0.1598L_3 + 0.2463L_4 + 0.2008L_5$
		$\overline{E}(L) : 0.0740L_0 + 0.0920L_1 + 0.1487L_2 + 0.1561L_3 + 0.2742L_4 + 0.2549L_5$
1	3	$\underline{E}(L) : 0.0906L_0 + 0.0825L_1 + 0.2032L_2 + 0.1682L_3 + 0.2595L_4 + 0.1961L_5$
		$\overline{E}(L) : 0.0502L_0 + 0.0797L_1 + 0.1708L_2 + 0.1636L_3 + 0.2775L_4 + 0.2582L_5$
3	1	$\underline{E}(L) : 0.1044L_0 + 0.0694L_1 + 0.1784L_2 + 0.1970L_3 + 0.2128L_4 + 0.2380L_5$
		$\overline{E}(L) : 0.0613L_0 + 0.0660L_1 + 0.1517L_2 + 0.1820L_3 + 0.2367L_4 + 0.3022L_5$
1	4	$\underline{E}(L) : 0.1170L_0 + 0.0773L_1 + 0.1772L_2 + 0.1660L_3 + 0.2561L_4 + 0.2064L_5$
		$\overline{E}(L) : 0.0734L_0 + 0.0735L_1 + 0.1489L_2 + 0.1613L_3 + 0.2807L_4 + 0.2623L_5$
4	1	$\underline{E}(L) : 0.0996L_0 + 0.0948L_1 + 0.1756L_2 + 0.1657L_3 + 0.2233L_4 + 0.2409L_5$
		$\overline{E}(L) : 0.0554L_0 + 0.0917L_1 + 0.1480L_2 + 0.1564L_3 + 0.2478L_4 + 0.3008L_5$
2	3	$\underline{E}(L) : 0.0971L_0 + 0.0928L_1 + 0.1967L_2 + 0.1649L_3 + 0.2757L_4 + 0.1727L_5$
		$\overline{E}(L) : 0.0545L_0 + 0.0859L_1 + 0.1692L_2 + 0.1498L_3 + 0.2986L_4 + 0.2422L_5$
3	2	$\underline{E}(L) : 0.1141L_0 + 0.0745L_1 + 0.1697L_2 + 0.1873L_3 + 0.1984L_4 + 0.2559L_5$
		$\overline{E}(L) : 0.0688L_0 + 0.0723L_1 + 0.1433L_2 + 0.1764L_3 + 0.2154L_4 + 0.3237L_5$
2	4	$\underline{E}(L) : 0.1121L_0 + 0.0979L_1 + 0.1839L_2 + 0.1421L_3 + 0.2573L_4 + 0.2068L_5$
		$\overline{E}(L) : 0.0686L_0 + 0.0923L_1 + 0.1549L_2 + 0.1379L_3 + 0.2809L_4 + 0.2654L_5$
4	2	$\underline{E}(L) : 0.1213L_0 + 0.0892L_1 + 0.1546L_2 + 0.1718L_3 + 0.2241L_4 + 0.2390L_5$
		$\overline{E}(L) : 0.0746L_0 + 0.0861L_1 + 0.1294L_2 + 0.1639L_3 + 0.2442L_4 + 0.3018L_5$
3	4	$\underline{E}(L) : 0.0977L_0 + 0.0683L_1 + 0.1839L_2 + 0.1785L_3 + 0.2228L_4 + 0.2487L_5$
		$\overline{E}(L) : 0.0589L_0 + 0.0648L_1 + 0.1538L_2 + 0.1676L_3 + 0.2488L_4 + 0.3061L_5$
4	3	$\underline{E}(L) : 0.0876L_0 + 0.0773L_1 + 0.1876L_2 + 0.1720L_3 + 0.2647L_4 + 0.2108L_5$
		$\overline{E}(L) : 0.0479L_0 + 0.0748L_1 + 0.1562L_2 + 0.1647L_3 + 0.2841L_4 + 0.2723L_5$

Table 4.5: The NPI lower and upper expected multi-level losses for all groups combinations for Example 4.3.2.

yields the smallest NPI upper expected multi-level loss and the second smallest NPI lower expected multi-level loss, while the second scenario results in the smallest NPI

		Case 1		Case 2		Case 3	
ℓ_1	ℓ_2	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.2939	0.3257	0.5144	0.5875	0.6797	0.7375
2	1	0.2916	0.3229	0.5151	0.5874	0.6775	0.7352
1	3	0.3012	0.3313	0.5290	0.5992	0.7105	0.7624
3	1	0.3058	0.3374	0.5313	0.6068	0.7155	0.7704
1	4	0.2986	0.3289	0.5299	0.6005	0.6976	0.7519
4	1	0.3041	0.3352	0.5362	0.6094	0.7022	0.7570
2	3	0.2948	0.3279	0.519	0.5983	0.6951	0.7534
3	2	0.3049	0.3368	0.5320	0.6064	0.7068	0.7626
2	4	0.2955	0.3266	0.5275	0.5998	0.6856	0.7415
4	2	0.3005	0.3322	0.5321	0.6052	0.6908	0.7476
3	4	0.3107	0.3401	0.5464	0.6175	0.7242	0.7750
4	3	0.3081	0.3379	0.5459	0.6161	0.7218	0.7728

Table 4.6: The NPI lower and upper expected multi-level losses for different loss values for Example 4.3.2

lower expected multi-level loss and the second smallest NPI upper expected multi-level loss. For selecting 2 as the best and 3 as the second best, this provides the smallest NPI lower expected multi-level loss in both Cases 4 and 9 and the smallest NPI upper expected multi-level loss in Case 9.

The combinations of group selections that result in the largest NPI lower and upper expected multi-level losses consistently involve the selection of group 3 as the best and 4 as the second best. This trend is consistently observed across multiple cases, namely Cases 1, 2, 3, 4 and 7. When the combination of 4 and 3, where 4 is selected as the best and 3 as the second best, this combination yields the following results: It leads to the highest NPI lower and upper expected multi-level losses in Case 5, 6 and 8. Lastly, selecting 3 as the best and 2 as the second best selected group, in Case 4 and 9, it yields the largest NPI upper expected multi-level loss.

In conclusion, the analysis of the nine cases in Example 4.3.2 shows more variation in the ranked subset of the two best groups, making the selection of the ranked

		Case 4		Case 5		Case 6	
ℓ_1	ℓ_2	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.5656	0.6194	0.8792	0.9257	0.7813	0.8328
2	1	0.5526	0.6032	0.8797	0.9259	0.7817	0.8339
1	3	0.5659	0.6173	0.9094	0.9498	0.8269	0.8701
3	1	0.5789	0.6327	0.8956	0.9387	0.8262	0.8727
1	4	0.5575	0.6076	0.8830	0.9266	0.8056	0.8532
4	1	0.5824	0.6347	0.9004	0.9446	0.8056	0.8529
2	3	0.5513	0.6075	0.9029	0.9455	0.8101	0.8597
3	2	0.5823	0.6389	0.8858	0.9312	0.8113	0.8588
2	4	0.5597	0.6111	0.8879	0.9314	0.7900	0.8390
4	2	0.5706	0.6255	0.8787	0.9254	0.7895	0.8393
3	4	0.5876	0.6359	0.9023	0.9411	0.8340	0.8763
4	3	0.5748	0.6256	0.9124	0.9521	0.8351	0.8773

Table 4.7: The NPI lower and upper expected multi-level losses for different loss
values for Example 4.3.2

subset of the two best groups inconsistent compared to the more consistent conclu-
sions observed in Example 4.3.1. This shows how different loss values and scenarios
can affect the ranking of the selected groups.

$\diamond$

4.4 Ranked subset of multiple best groups using general multi-level loss function

In this section, a ranked subset of multiple best groups using a general multi-level loss
function will be applied to provide a graded penalty for different levels of mistakes
for ranking w best groups within the subset. The number of possible ways to rank
two groups that belong to S together with the maximum future observation among
groups in NS is $(w+1)! - 1$. The definition of the general multi-level loss function

ℓ_1	ℓ_2	Case 7		Case 8		Case 9	
		$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$	$\underline{E}(L_{ML})$	$\overline{E}(L_{ML})$
1	2	0.6082	0.6856	0.4411	0.5236	0.2292	0.2896
2	1	0.6069	0.6851	0.4470	0.5291	0.2008	0.2549
1	3	0.6237	0.6994	0.4556	0.5358	0.1961	0.2582
3	1	0.6478	0.7210	0.4508	0.5390	0.2380	0.3022
1	4	0.6285	0.7043	0.4625	0.5431	0.2064	0.2623
4	1	0.6300	0.7049	0.4643	0.5485	0.2409	0.3008
2	3	0.6134	0.6905	0.4485	0.5408	0.1727	0.2422
3	2	0.6416	0.7155	0.4543	0.5391	0.2559	0.3237
2	4	0.6062	0.6842	0.4640	0.5463	0.2068	0.2654
4	2	0.6349	0.7099	0.4631	0.5459	0.2390	0.3018
3	4	0.6500	0.7225	0.4715	0.5549	0.2487	0.3061
4	3	0.6475	0.7211	0.4755	0.5564	0.2108	0.2723

Table 4.8: The NPI lower and upper expected multi-level losses for different loss values for Example 4.3.2

for the event of interest $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \cdots < X_{\ell_2}^f < X_{\ell_1}^f$ is as follows.

$$L_G\left(\max_{j_z \in NS} X_{j_z}^f, X_{\ell_w}^f, \ldots, X_{\ell_2}^f, X_{\ell_1}^f\right) = \begin{cases} L_0 & \text{if } \max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \cdots < X_{\ell_2}^f < X_{\ell_1}^f \\ L_1 & \text{if } \max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \cdots < X_{\ell_1}^f < X_{\ell_2}^f \\ \vdots & \quad \vdots \\ \vdots & \quad \vdots \\ L_{(w+1)!-1} & \text{if } X_{\ell_1}^f < X_{\ell_2}^f < \cdots < X_{\ell_w}^f < \max\limits_{j_z \in NS} X_{j_z}^f \end{cases}$$

$$(4.5)$$

The general multi-level loss function L_G assigns different loss values based on the ranking of future observations from the w groups in S and the maximum future observation that belongs to NS. Note that L_0 is zero because it is related to the correct ranking of the w groups that belong to S. For $\max\limits_{j_z \in NS} X_{j_z}^f < X_{\ell_1}^f < \cdots < X_{\ell_1}^f <$

$X_{\ell_2}^f$, the future observation for $X_{\ell_1}^f$ is smaller than $X_{\ell_2}^f$, which is an incorrect ranking. Therefore, the loss value is L_1, which is higher than L_0. There are $(w+1)! - 2$ other intermediate future observations that do not follow the desired ranking. Each incorrect ranking corresponds to a different loss value, gradually increasing as the ranking deviates more from the desired ranking. Finally, $X_{\ell_1}^f < X_{\ell_2}^f < \cdots < X_{\ell_w}^f < \max_{j_z \in NS} X_{j_z}^f$ is the worst scenario in which all future observations from S are smaller than the maximum future observations from NS. The loss is assumed to be lower when $\max_{j_z \in NS} X_{j_z}^f$ is the smallest compared to when $\max_{j_z \in NS} X_{j_z}^f$ is the second smallest, continuing in this manner until the worst case scenario where $\max_{j_z \in NS} X_{j_z}^f$ is the largest. The loss value $L_{(w+1)!-1}$ is the highest. The factorial number of possible rankings $((w+1)! - 1)$ can become very large, making the assessment process complex to manage for large w, especially with large and overlapping datasets.

To derive the NPI lower expected general multi-level loss, the probability mass corresponding to group ℓ_1 is assigned to the right endpoint per interval. The probability masses corresponding to groups $j_z \in NS$ are assigned to the left endpoint per interval to determine the minimum possible loss, L_c, where $c = 0, \ldots, (w+1)! - 1$.

To derive the NPI upper expected general multi-level loss, the probability mass corresponding to group ℓ_1, is assigned to the left endpoint per interval. Meanwhile, the probability masses corresponding to groups $j_z \in NS$ are assigned to the right endpoint per interval to determine the maximum possible loss, L_c, where $c = 0, \ldots, (w+1)! - 1$.

Now, with two or more groups in the middle, it becomes challenging to find the exact NPI lower expected general multi-level loss that determines the minimum possible loss. The number of groups in the middle for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \cdots < X_{\ell_2}^f < X_{\ell_1}^f$ is $w - 1$, making the probability mass allocation more complex. To address this, an independent optimisation process is carried out for each middle group. For each group X_{ℓ_i}, where $i = 2, \ldots, w$, its corresponding probability mass $\frac{1}{n_{\ell_i}+1}$ is assigned to any point (suppose the midpoint) belonging to a sub-interval within $(x_{\ell_i, u_{\ell_i}-1}, x_{\ell_i, u_{\ell_i}})$ that minimises the possible loss. Consequently, apply heuristic algorithm to determine the approximate NPI lower expected general multi-level loss by selecting the smallest NPI lower expected general multi-level loss as an ap-

proximate NPI lower expected general multi-level loss.

For the NPI upper expected general multi-level loss, the process is similar but aims to determine the maximum possible loss. The number of groups in the middle remains $w - 1$, and an optimisation process is applied independently to each group. For each group X_{ℓ_i}, where $i = 2, \ldots, w$, its corresponding probability mass $\frac{1}{n_{\ell_i} + 1}$ is assigned to a point (suppose the midpoint) belonging to a sub-interval within $(x_{\ell_i, u_{\ell_i} - 1}, x_{\ell_i, u_{\ell_i}})$ that maximises the possible loss. Consequently, apply heuristic algorithm to determine the approximate NPI upper expected general multi-level loss by selecting the largest NPI upper expected general multi-level loss as an approximate NPI upper expected general multi-level loss.

Then, for the NPI lower expected general multi-level loss, there is a total of $\prod_{t=1}^{k}(n_t + 1)$ combinations of orderings of future observations from all k groups (intervals). For each ordering, it is necessary to determine which of $L_0, \ldots, L_{(w+1)! - 1}$ represents the minimum possible loss. The probability $\dfrac{1}{\prod_{t=1}^{k}(n_t + 1)}$ is then assigned to the actual possible loss L_c, where $c = 0, \ldots, (w + 1)! - 1$, ensuring that the loss value that minimises the NPI lower expected general multi-level loss is selected. This process iterates over all combinations of orderings to determine the minimum possible loss. The procedure is repeated $w - 1$ times for the NPI lower expected general multi-level loss, and finally, the smallest lower expected loss is chosen as an approximate NPI lower expected general multi-level loss. Therefore, the approximate NPI lower expected general multi-level loss is denoted by

$$\underline{E}^{l}(L_G(\max_{j_z \in NS} X_{j_z}^{f}, \ldots, X_{\ell_1}^{f})) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t + 1)} \left[\min_{c} L_c \mid O_i \right] \qquad (4.6)$$

For the NPI upper expected general multi-level loss, among the total of the $\prod_{t=1}^{k}(n_t + 1)$ combinations of orderings of future observations from all k groups. For each ordering, it is necessary to determine which of $L_0, \ldots, L_{(w+1)! - 1}$ is the maximum possible loss. The probability is assigned to the actual possible loss L_c, ensuring that the loss value that maximises the NPI upper expected general multi-level loss is chosen. This process iterates over all combinations of orderings of future observations to determine the maximum possible loss. The procedure is repeated $w - 1$ times for the NPI upper expected general multi-level loss, and finally, the

largest NPI upper expected loss is selected as an approximate NPI upper expected general multi-level loss. The approximate NPI upper expected general multi-level loss is denoted by

$$\overline{E}^u(L_G(\max_{j_z \in NS} X_{j_z}^f, \ldots, X_{\ell_1}^f) = \frac{1}{\prod_{t=1}^{k}(n_t + 1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t+1)} \left[\max_c L_c \,|\, O_i \right] \qquad (4.7)$$

The following example illustrates the method presented using general multi-level loss to guide the ranking of future observations in the selection of a ranked subset of the three best groups and reflects the quality of their ranking.

Example 4.4.1 The dataset used in this example is similar to Example 2.4.1, taken from the lengthWeight dataset in the StatCharme Package in R contains variables which are: The chemical concentration, the fish were exposed to 6 levels of treatments and weight of the fish. The dataset is provided in Figure 2.2. Starting with the first case, it is assumed that the groups that belong to S are 1, 2 and 3, while the two groups that belong to NS are 4 and 5. The event of interest is $\max_{j_z \in NS}(X_4, X_5)^f < X_3^f < X_2^f < X_1^f$. The number of possible ways to rank the three groups that belong to S and two groups that belong to NS except the correct ranked subset is 24. The aim is to guide the ranking of future observations in the subset of the three best groups and to reflect on the quality of the ranking. There are a total of 60 possible combinations of selecting a ranked subset of the three best groups, since this number is rather large, the process is simplified by fixing the groups that belong to NS, which are 4 and 5, and focusing only on the different combinations of rankings of the three groups that belong to S. Therefore, there are 6 different possible combinations of selecting a ranked subset of the best three groups, when the groups in NS are 4 and 5, and the rest belong to S as presented in Tables 4.9, 4.10 and 4.11.

In Tables 4.9, 4.10 and 4.11, the approximate NPI lower and upper general expected multi-level losses for different selection of ranked subsets of the three best groups are presented. Each table represents a different selection of ranked subsets of three best groups. For example, as presented in Table 4.9, $\{4,5\} \prec 3 \prec 2 \prec 1$ represents this event $\max_{j_z \in NS}(X_4, X_5)^f < X_3^f < X_2^f < X_1^f$, where the best selected group

is 1, and groups 4 and 5 belong to non-selected groups. The notations $\underline{E_2^l}(L_G)$ and $\overline{E_2^u}(L_G)$ represent the NPI lower and upper general expected multi-level losses when optimizing over group 2, respectively. Similar notation is defined for all other groups. Each coefficient loss C_c multiplies by its corresponding loss values L_c, where $c = 0, \ldots, 23$. C_c represents how many times L_c is indeed the actual possible loss. The approximate NPI lower general expected multi-level loss shows considerable variation across all tables. For example, in Table 4.9, for the ranking preference when selecting group 1 as the best, group 2 as the second best, and group 3 as the third best, the approximate NPI lower general expected multi-level loss when the optimisation is performed over the loss coefficient of the group 2 at C_0 is 0.2482, where the number of assigned probability masses to find the loss coefficient of C_0 is 1929, while for the ranking preference when selecting group 2 as the best, group 1 as the second best and group 3 as the third best, it is 0.1515, where the number of assigned probability masses to find the loss coefficient of C_0 is 1178. Also, particularly for loss coefficients like C_6 and C_0, the coefficients can substantially affect the approximate NPI lower general expected multi-level loss across all tables since their values are generally large compared with the rest.

On the other hand, the approximate NPI upper general expected multi-level losses are generally more consistent across different group rankings within the subset, though some differences emerge. For example, in Table 4.11, for the ranking preference when selecting group 1 as the best, group 3 as the second best and group 2 as the third best, the approximate NPI upper general expected multi-level loss for the event $\max_{j_z \in NS}(X_4, X_5)^f < X_2^f < X_3^f < X_1^f$ when the optimisation is performed over group 2 with the loss coefficient at C_6 is 0.1538, but it rises slightly to 0.1744 for the ranking preference when selecting group 3 as the best, group 2 as the second best and group 1 as the third best when optimisation was performed over group 1. For C_6 and C_{18}, the coefficients affect the approximate NPI upper general expected multi-level loss across all tables due to their generally large values. These variations suggest that, while the approximate NPI upper general expected multi-level loss tends to be consistent, higher loss coefficient values can still show differences depending on the group rankings within the subset.

Tables 4.9 and 4.11 generally show zero losses under C_{10}, except for C_{12} and C_{14} for the approximate NPI lower general expected multi-level losses. For the approximate NPI upper general expected multi-level loss, C_3 and C_{13} show zero across all ranking preferences.

	$\{4,5\} \prec 3 \prec 2 \prec 1$				$\{4,5\} \prec 3 \prec 1 \prec 2$			
	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$
C_0	0.2482	0.0574	0.2693	0.0518	0.1515	0.1096	0.2063	0.0615
C_1	0.0419	0.0500	0.1319	0.0170	0.1386	0.0075	0.1394	0.0073
C_2	0.0478	0.0247	0.0505	0.0064	0.0343	0.0129	0.0849	0.0283
C_3	0.0231	0.0000	0.0000	0.0129	0.0324	0.0000	0.0093	0.0064
C_4	0.0826	0.0023	0.0027	0.0418	0.0698	0.0000	0.0147	0.0103
C_5	0.0185	0.0096	0.0185	0.0000	0.0463	0.0000	0.0185	0.0161
C_6	0.2279	0.1201	0.1914	0.1507	0.1471	0.1862	0.1803	0.1507
C_7	0.0553	0.0947	0.0995	0.0473	0.1438	0.0118	0.1106	0.0473
C_8	0.0972	0.0724	0.1223	0.0521	0.0829	0.0939	0.1157	0.0688
C_9	0.0424	0.0073	0.0066	0.0418	0.0459	0.0000	0.0131	0.0251
C_{10}	0.0000	0.0234	0.0000	0.0138	0.0000	0.0234	0.0000	0.0170
C_{11}	0.0000	0.0054	0.0000	0.0122	0.0000	0.0026	0.0000	0.0090
C_{12}	0.0000	0.0862	0.0671	0.0707	0.0671	0.0868	0.0671	0.0707
C_{13}	0.0000	0.0120	0.0000	0.0000	0.0000	0.0109	0.0000	0.0000
C_{14}	0.0401	0.0540	0.0401	0.0862	0.0401	0.0701	0.0401	0.0862
C_{15}	0.0000	0.0310	0.0000	0.0396	0.0000	0.0287	0.0000	0.0396
C_{16}	0.0000	0.0000	0.0000	0.0075	0.0000	0.0075	0.0000	0.0000
C_{17}	0.0000	0.0159	0.0000	0.0113	0.0000	0.0113	0.0000	0.0188
C_{18}	0.0000	0.1160	0.0000	0.1202	0.0000	0.1730	0.0000	0.1245
C_{19}	0.0000	0.0795	0.0000	0.0283	0.0000	0.0207	0.0000	0.0240
C_{20}	0.0000	0.0665	0.0000	0.0637	0.0000	0.0778	0.0000	0.0806
C_{21}	0.0000	0.0424	0.0000	0.0439	0.0000	0.0359	0.0000	0.0523
C_{22}	0.0000	0.0042	0.0000	0.0571	0.0000	0.0157	0.0000	0.0156
C_{23}	0.0000	0.0249	0.0000	0.0238	0.0000	0.0139	0.0000	0.0399

Table 4.9: The approximate NPI lower and upper expected general multi-level losses for Example 4.4.1, where the events of interest are $\max_{j_z \in NS}(X_4, X_5)^f < X_3^f < X_2^f < X_1^f$, and $\max_{j_z \in NS}(X_4, X_5)^f < X_3^f < X_1^f < X_2^f$

	$\{4,5\} \prec 1 \prec 3 \prec 2$				$\{4,5\} \prec 2 \prec 1 \prec 3$			
	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$
C_0	0.1757	0.1021	0.1399	0.0577	0.1878	0.1021	0.2325	0.0970
C_1	0.0959	0.0149	0.0761	0.0201	0.0838	0.0190	0.0345	0.0394
C_2	0.0316	0.0129	0.0509	0.0129	0.0316	0.0129	0.0575	0.0000
C_3	0.0208	0.0000	0.0185	0.0129	0.0401	0.0021	0.0185	0.0000
C_4	0.0818	0.0000	0.0671	0.0418	0.0864	0.0000	0.1161	0.0000
C_5	0.0579	0.0000	0.0417	0.0000	0.0247	0.0062	0.0139	0.0129
C_6	0.1692	0.1744	0.2281	0.1384	0.1803	0.1744	0.2467	0.1538
C_7	0.1124	0.0237	0.1415	0.0442	0.0921	0.0360	0.0442	0.0828
C_8	0.0764	0.0939	0.0571	0.0598	0.0764	0.0959	0.0829	0.0682
C_9	0.0478	0.0000	0.0394	0.0418	0.0432	0.0041	0.0459	0.0064
C_{10}	0.0069	0.0234	0.0000	0.0156	0.0123	0.0131	0.0000	0.0298
C_{11}	0.0023	0.0026	0.0000	0.0135	0.0062	0.0000	0.0000	0.0000
C_{12}	0.0640	0.0868	0.0671	0.0450	0.0610	0.0728	0.0671	0.0643
C_{13}	0.0093	0.0219	0.0000	0.0000	0.0185	0.0273	0.0000	0.0000
C_{14}	0.0386	0.0701	0.0725	0.1042	0.0370	0.0903	0.0401	0.0540
C_{15}	0.0000	0.0177	0.0000	0.0365	0.0000	0.0000	0.0000	0.0473
C_{16}	0.0046	0.0149	0.0000	0.0000	0.0093	0.0087	0.0000	0.0000
C_{17}	0.0000	0.0039	0.0000	0.0203	0.0000	0.0039	0.0000	0.0149
C_{18}	0.0031	0.1737	0.0000	0.1061	0.0062	0.1600	0.0000	0.1386
C_{19}	0.0000	0.0322	0.0000	0.0283	0.0000	0.0473	0.0000	0.0538
C_{20}	0.0000	0.0824	0.0000	0.0778	0.0031	0.1065	0.0000	0.0538
C_{21}	0.0000	0.0237	0.0000	0.0359	0.0000	0.0085	0.0000	0.0581
C_{22}	0.0000	0.0221	0.0000	0.0368	0.0000	0.0057	0.0000	0.0028
C_{23}	0.0000	0.0028	0.0000	0.0505	0.0000	0.0028	0.0000	0.0221

Table 4.10: The approximate NPI lower and upper expected general multi-level losses for Example 4.4.1, where the events of interest are $\max_{j_z \in NS}(X_4, X_5)^f < X_1^f < X_3^f < X_2^f$, and $\max_{j_z \in NS}(X_4, X_5)^f < X_2^f < X_1^f < X_3^f$

	$\{4,5\} \prec 2 \prec 3 \prec 1$				$\{4,5\} \prec 1 \prec 2 \prec 3$			
	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$
C_0	0.1651	0.0970	0.1535	0.0601	0.1515	0.1021	0.2325	0.0723
C_1	0.0586	0.0394	0.0625	0.0094	0.1386	0.0149	0.0345	0.0351
C_2	0.1003	0.0000	0.0509	0.0051	0.0343	0.0129	0.0575	0.0329
C_3	0.0278	0.0000	0.0185	0.0129	0.0602	0.0000	0.0185	0.0000
C_4	0.0748	0.0000	0.0671	0.0431	0.0698	0.0000	0.1161	0.0031
C_5	0.0139	0.0129	0.0417	0.0096	0.0185	0.0000	0.0139	0.0096
C_6	0.1921	0.1538	0.2423	0.1425	0.1471	0.1744	0.2467	0.1579
C_7	0.0664	0.0828	0.1273	0.0453	0.1438	0.0237	0.0442	0.0710
C_8	0.1111	0.0682	0.0571	0.0489	0.0829	0.0939	0.0829	0.0642
C_9	0.0502	0.0064	0.0394	0.0502	0.0459	0.0000	0.0459	0.0066
C_{10}	0.0000	0.0298	0.0000	0.0158	0.0000	0.0234	0.0000	0.0234
C_{11}	0.0000	0.0000	0.0000	0.0122	0.0000	0.0026	0.0000	0.0072
C_{12}	0.0594	0.0643	0.0671	0.0547	0.0671	0.0707	0.0671	0.0720
C_{13}	0.0000	0.0129	0.0000	0.0000	0.0000	0.0219	0.0000	0.0000
C_{14}	0.0802	0.0540	0.0725	0.0971	0.0401	0.0862	0.0401	0.0540
C_{15}	0.0000	0.0345	0.0000	0.0376	0.0000	0.0177	0.0000	0.0458
C_{16}	0.0000	0.0000	0.0000	0.0080	0.0000	0.0149	0.0000	0.0000
C_{17}	0.0000	0.0149	0.0000	0.0118	0.0000	0.0039	0.0000	0.0141
C_{18}	0.0000	0.1386	0.0000	0.1132	0.0000	0.1620	0.0000	0.1330
C_{19}	0.0000	0.0687	0.0000	0.024	0.0000	0.0322	0.0000	0.0467
C_{20}	0.0000	0.0538	0.0000	0.0665	0.0000	0.0941	0.0000	0.0665
C_{21}	0.0000	0.0432	0.0000	0.0435	0.0000	0.0237	0.0000	0.0554
C_{22}	0.0000	0.0028	0.0000	0.0608	0.0000	0.0221	0.0000	0.0042
C_{23}	0.0000	0.0221	0.0000	0.0279	0.0000	0.0028	0.0000	0.0249

Table 4.11: The approximate NPI lower and upper expected general multi-level losses for Example 4.4.1, where the events of interest are $\max_{j_z \in NS}(X_4, X_5)^f < X_2^f < X_3^f < X_1^f$, and $\max_{j_z \in NS}(X_4, X_5)^f < X_1^f < X_2^f < X_3^f$

ℓ_1	ℓ_2	ℓ_3					Approximate	
1	2	3	$\underline{E}_2^l(L)$	$\overline{E_2^u}(L)$	$\underline{E}_3^l(L)$	$\overline{E_3^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.4331	0.5415	0.4565	0.5910	0.4331	0.5910
2	1	3	$\underline{E}_1^l(L)$	$\overline{E_1^u}(L)$	$\underline{E}_3^l(L)$	$\overline{E_3^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.5270	0.5723	0.4928	0.5549	0.4928	0.5723
2	3	1	$\underline{E}_1^l(L)$	$\overline{E_1^u}(L)$	$\underline{E}_3^l(L)$	$\overline{E_3^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.5148	0.5821	0.5352	0.5632	0.5148	0.5821
3	1	2	$\underline{E}_1^l(L)$	$\overline{E_1^u}(L)$	$\underline{E}_2^l(L)$	$\overline{E_2^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.5214	0.5625	0.4862	0.5461	0.4862	0.5625
1	3	2	$\underline{E}_2^l(L)$	$\overline{E_2^u}(L)$	$\underline{E}_3^l(L)$	$\overline{E_3^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.4941	0.5453	0.5275	0.5794	0.4941	0.5794
3	2	1	$\underline{E}_1^l(L)$	$\overline{E_1^u}(L)$	$\underline{E}_2^l(L)$	$\overline{E_2^u}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
			0.5389	0.5747	0.4862	0.5632	0.4862	0.5747

Table 4.12: The approximate NPI lower and upper expected general multi-level losses for different loss values for Example 4.4.1. Case 1: $L_0 = 0.00, L_1 = 0.89, L_2 = 0.52, L_3 = 0.45, L_4 = 0.55, L_5 = 0.32, L_6 = 0.88, L_7 = 0.57, L_8 = 0.28, L_9 = 0.95, L_{10} = 0.69, L_{11} = 0.94, L_{12} = 0.45, L_{13} = 0.88, L_{14} = 0.24, L_{15} = 0.95, L_{16} = 0.89, L_{17} = 0.78, L_{18} = 0.99, L_{19} = 0.04, L_{20} = 0.64, L_{21} = 0.04, L_{22} = 0.67, L_{23} = 0.10$.

Tables 4.12, 4.13, 4.14, 4.15 and 4.16 provide a comparison of the approximate NPI lower and upper expected general multi-level losses across five different cases, each using different values of L_c, where $c = 0, \ldots, 23$. These variations in L_c values result in differences in the approximate NPI lower and upper expected general multi-level losses for various group rankings within the subset within the subset, allowing for the evaluation of whether consistent conclusions can be drawn across the different cases.

In Case 1, the L_c values are randomly generated between 0 and 1, which results in varying the approximate NPI lower and upper expected general multi-level losses for the group rankings within the subset within the subset. Case 2 and the rest provide a scenario where the L_c values are regularly increasing. For Case 2, the L_c values

ℓ_1	ℓ_2	ℓ_3					Approximate	
1	2	3	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.0821	0.2339	0.0917	0.2449	0.0821	0.2449
2	1	3	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.1029	0.2333	0.0951	0.2440	0.0951	0.2440
2	3	1	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.1067	0.2317	0.1140	0.2454	0.1067	0.2454
3	1	2	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.1093	0.2259	0.0994	0.2285	0.0994	0.2285
1	3	2	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.1101	0.2274	0.1135	0.2453	0.1101	0.2453
3	2	1	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.1048	0.2328	0.0994	0.2306	0.0994	0.2328

Table 4.13: The approximate NPI lower and upper expected general multi-level losses for different loss values for Example 4.4.1. Case 2: $L_0 = 0.00, L_1 = 0.02, L_2 = 0.04, L_3 = 0.06, L_4 = 0.08, L_5 = 0.10, L_6 = 0.12, L_7 = 0.14, L_8 = 0.16, L_9 = 0.18, L_{10} = 0.20, L_{11} = 0.22, L_{12} = 0.24, L_{13} = 0.26, L_{14} = 0.28, L_{15} = 0.30, L_{16} = 0.32, L_{17} = 0.34, L_{18} = 0.36, L_{19} = 0.38, L_{20} = 0.40, L_{21} = 0.42, L_{22} = 0.44, L_{23} = 0.50.$

start small, ranging from 0.00 to 0.50. Case 3 represents an extreme scenario where all L_c values are 1 except $L_0 = 0$, making it a zero-one loss case which is a special case of the general multi-level loss. In Case 4, the L_c values are clustered around 0.50, increasing from 0.38 to 0.60. Case 5 further increases these values, ranging from 0.50 to 1.00, leading to consistently higher approximate NPI lower and upper expected general multi-level losses across all group rankings within the subset with the subset. Loss values in these cases chosen arbitrary. These cases help investigate the NPI method and determine whether they lead to consistent conclusions.

Across the five cases, the smallest approximate NPI lower expected general multi-level losses usually correspond to the ranking preference of selecting group 1 as the best, group 2 as the second best and group 3 as the third best. For example, in Case

ℓ_1	ℓ_2	ℓ_3					Approximate	
1	2	3	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.6768	0.9425	0.7306	0.9483	0.6768	0.9483
2	1	3	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L)$
			0.8483	0.8906	0.7937	0.9385	0.7937	0.9385
2	3	1	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L)$
			0.8226	0.8980	0.8600	0.9424	0.8226	0.9424
3	1	2	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L)$
			0.8122	0.8975	0.7674	0.9030	0.7674	0.9030
1	3	2	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.8348	0.9031	0.8464	0.9401	0.8348	0.9401
3	2	1	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.8576	0.8980	0.7674	0.9276	0.7674	0.9276

Table 4.14: The approximate NPI lower and upper expected general multi-level losses for different loss values for Example 4.4.1. Case 3: $L_0 = 0.00, L_1 = 1.00, L_2 = 1.00, L_3 = 1.00, L_4 = 1.00, L_5 = 1.00, L_6 = 1.00, L_7 = 1.00, L_8 = 1.00, L_9 = 1.00, L_{10} = 1.00, L_{11} = 1.00, L_{12} = 1.00, L_{13} = 1.00, L_{14} = 1.00, L_{15} = 1.00, L_{16} = 1.00, L_{17} = 1.00, L_{18} = 1.00, L_{19} = 1.00, L_{20} = 1.00, L_{21} = 1.00, L_{22} = 1.00, L_{23} = 1.00$.

1, the smallest approximate NPI lower expected general multi-level loss is 0.4331 and in Case 2, it is 0.0821. Similarly, in Case 3, it is 0.6768, in Case 4, it is 0.2915, and in Case 5, it is 0.4166. The smallest approximate NPI upper expected general multi-level loss corresponds to the ranking preference of selecting group 3 as the best, group 1 as the second best and group 2 as the third best for Cases 1, 2, 3, 4 and 5, with values of 0.5910, 0.2285, 0.9030, 0.4479 and 0.6792 respectively.

Similarly, the largest approximate NPI lower expected general multi-level loss typically occurs for the ranking preference of selecting group 1 as the best, group 3 as the second best and group 2 as the third best in all cases except Case 1. In Case 1, the largest approximate NPI lower expected general multi-level loss is 0.5148 corresponding to the ranking preference of selecting group 2 as the best, group 3

ℓ_1	ℓ_2	ℓ_3					Approximate	
1	2	3	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.2915	0.4652	0.3162	0.4729	0.2915	0.4729
2	1	3	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.3653	0.4459	0.3412	0.4685	0.3412	0.4685
2	3	1	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.3577	0.4481	0.3752	0.4704	0.3577	0.4704
3	1	2	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.3552	0.4450	0.3337	0.4479	0.3337	0.4479
1	3	2	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}_3^l(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.3639	0.4474	0.3699	0.4699	0.3639	0.4699
3	2	1	$\underline{E}_1^l(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E}_2^l(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.3697	0.4486	0.3337	0.4580	0.3337	0.4580

Table 4.15: The approximate NPI lower and upper expected general multi-level losses for different loss values for Example 4.4.1. Case 4: $L_0 = 0.00, L_1 = 0.38, L_2 = 0.39, L_3 = 0.40, L_4 = 0.41, L_5 = 0.42, L_6 = 0.43, L_7 = 0.44, L_8 = 0.45, L_9 = 0.46, L_{10} = 0.47, L_{11} = 0.48, L_{12} = 0.49, L_{13} = 0.50, L_{14} = 0.51, L_{15} = 0.52, L_{16} = 0.53, L_{17} = 0.54, L_{18} = 0.55, L_{19} = 0.56, L_{20} = 0.57, L_{21} = 0.58, L_{22} = 0.59, L_{23} = 0.60.$

as the second best and group 1 as the third best. The largest approximate NPI upper expected general multi-level loss follows a similar pattern, with the ranking preference of selecting group 1 as the best, group 2 as the second best, and group 3 as the third best having the largest value just for Case 3 is 0.9483, Case 4 is 0.4729, Case 5 is 0.7175, while for Case 2, the largest approximate NPI upper expected general multi-level loss occurs for selecting group 2 as the best, group 3 as the second best, and group 1 as the third best.

If the ranking with the lowest approximate NPI lower expected general multi-level loss is different from the ranking with the lowest approximate NPI upper expected general multi-level loss, then the choice depends on the decision-maker. For example, in Case 2, the lowest approximate NPI lower expected general multi-level

ℓ_1	ℓ_2	ℓ_3					Approximate	
1	2	3	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.4166	0.7036	0.4533	0.7175	0.4166	0.7175
2	1	3	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.5216	0.6782	0.4870	0.7122	0.4870	0.7122
2	3	1	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.5134	0.6801	0.5398	0.7149	0.5134	0.7149
3	1	2	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.5106	0.6740	0.4786	0.6792	0.4786	0.6792
1	3	2	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E_3^l}(L_G)$	$\overline{E_3^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.5223	0.6782	0.5327	0.714	0.5223	0.7140
3	2	1	$\underline{E_1^l}(L_G)$	$\overline{E_1^u}(L_G)$	$\underline{E_2^l}(L_G)$	$\overline{E_2^u}(L_G)$	$\underline{E}(L_G)$	$\overline{E}(L_G)$
			0.5276	0.6813	0.4786	0.6929	0.4786	0.6929

Table 4.16: The approximate NPI lower and upper expected general multi-level losses for different loss values for Example 4.4.1. Case 5: $L_0 = 0.00, L_1 = 0.50, L_2 = 0.52, L_3 = 0.54, L_4 = 0.56, L_5 = 0.60, L_6 = 0.62, L_7 = 0.64, L_8 = 0.66, L_9 = 0.68, L_{10} = 0.70, L_{11} = 0.72, L_{12} = 0.74, L_{13} = 0.76, L_{14} = 0.78, L_{15} = 0.80, L_{16} = 0.82, L_{17} = 0.84, L_{18} = 0.86, L_{19} = 0.88, L_{20} = 0.90, L_{21} = 0.92, L_{22} = 0.94, L_{23} = 1.00$.

loss corresponds to selecting group 1 as the best, group 2 as the second best, and group 3 as the third best. In contrast, the lowest approximate NPI upper expected general multi-level loss corresponds to selecting group 2 as the best, group 3 as the second best, and group 1 as the third best. A cautious decision-maker, who aims to minimise the worst-case loss, may focus on the approximate NPI upper expected general multi-level loss. This choice may be safer, as it leads to the smallest possible maximum loss. On the other hand, an optimistic decision-maker, who focuses on the best-case loss, may choose the ranking with the lowest approximate NPI lower expected general multi-level loss, as this minimises the loss if things go well.

Additionally, it is noticed that in Case 2, the approximate NPI lower and upper expected general multi-level losses are quite small, with values ranging from 0.0821

to 0.2454. This occurs because all L_c values are set to small values, less than 0.50. However, in Case 3, the approximate NPI lower and upper expected general multi-level losses are substantially larger, ranging from 0.6768 to 0.9483. This suggests that the loss function in Case 3 shows much higher approximate NPI lower and upper expected general multi-level losses compared to Case 2. This is because in Case 3, all the values of L_c are set to 1, which represents an extreme scenario.

It should also be noted that the results from Table 4.14, which are based on the zero-one loss function, do not exactly match the results from Table 2.3, which present approximate NPI lower and upper probabilities. This is because the direct relationship exists between the exact NPI lower and upper probabilities and the NPI lower and upper expected losses, not with the approximate bounds.

In conclusion, while there are some similarities in the optimal ranked subset of the three best groups corresponding to the smallest approximate NPI lower and upper expected multi-level losses. However, the largest approximate NPI lower and upper expected multi-level losses reveal key differences, particularly in Cases 1 and 2, show that the different structures of L_c values across cases can lead to different optimal ranked subset of the three best groups. These cases do not consistently produce the same ranked subset of the three best groups for the largest approximate NPI lower and upper expected general multi-level losses.

4.5   Concluding remarks

This chapter introduces NPI for selecting a ranked subset of w best groups, using both the zero-one loss function and a general multi-level loss function. Two specific cases are considered: $w = 2$, representing a ranked subset of the two best groups, and $w = 3$, representing a ranked subset of the three best groups.

The zero-one loss function introduced in Section 4.2 is a binary measure, indicating whether the ranking of future observations within a subset of w best groups is correct or not. It just classifies whether the ranking of subset of best groups is correct, without any further consideration. The zero-one loss function is a special case of the general multi-level loss presented in Sections 4.3 and 4.4.

A general multi-level loss function has been defined to provide a graded penalty for different levels of mistakes for selecting a ranked subset of w best groups, assigning different penalties to various types of group rankings within the subset within the subset. This flexibility is useful when certain rankings may have more serious consequences than others. As previously mentioned, two specific cases are considered: $w = 2$ in Section 4.3 and $w = 3$ in Section 4.4. The reason for presenting these two cases is as follows. For $w = 2$, there is only one group in the middle, making the optimisation process easier to perform. For $w = 3$ and beyond, there are two or more groups in the middle, making the optimisation process more complex and computationally demanding. Examples of real-valued data from the literature and simulated data are presented to investigate the performance of the NPI method for selecting a ranked subset of w best groups by using the NPI lower and upper expected general multi-level losses.

When the ranked subset contains the two best groups, the multi-level loss function is introduced as a method to guide the ranking of future observations in this subset and to reflect the quality of the ranking of groups belong to this subset. It assigns a range of penalties. An optimisation process is then performed to derive the NPI lower and upper expected multi-level losses, helping in decision-making regarding the group rankings within the subset within that subset.

In some cases, the loss values must be predetermined by the problem rather than derived from the data. The impact of different loss values on the decision-making process for selecting a ranked subset of the best groups was explored by analysing various cases with varying loss values. The results of different cases demonstrate that the multi-level loss function is highly sensitive to these specific loss values, leading to different conclusions.

The general case of the multi-level loss function for a ranked subset of w best groups is presented in Section 4.4. This generalisation is necessary because the NPI lower and upper expected multi-level losses cannot be directly derived when more than one group lies between the best group in S and the non-selected groups in NS. Therefore, an optimisation process is required to obtain the approximate NPI lower and upper expected general multi-level losses. Subsequently, various loss values

are assigned across different cases to explore how different penalty levels affect the evaluation of selecting a ranked subset of the best groups. Applying these loss values reveals varying conclusions in the group ranking evaluations.

An alternative method for defining the general multi-level loss function, which can be considered in future works, is through grouped losses; instead of assigning a distinct loss to every possible ranking, rankings can be grouped based on specific structural patterns, with a common loss assigned to each group. The first group loss indicates rankings where future observations from NS are smaller than those of S. The second grouped loss accounts for the ranking in which future observations from NS are smaller than those from S, except for one group, which is smaller than all future observations from NS and so on till the last grouped loss represents the ranking in which future observations from NS exceed those from S. This method may provide a more structured way to assign losses, reducing the number of distinct loss values and making the loss function easier to interpret and apply.

Finally, another approach is weighted grouped losses, instead of assigning a unique loss to every ranking, similar rankings are grouped together and a common loss value is assigned to each group. The weighted grouped loss is then defined as the average of the losses within each group. Those methods can reduce the number of different loss values while still differentiating between correct and incorrect rankings.

Chapter 5

5.1 Introduction

Ranking independent groups is a central problem addressed in this thesis, and this chapter extends that problem by focusing on ranking groups organised into buckets. Bucketing allows for clustering groups when a complete strict ranking is either unnecessary or impractical. At one extreme, having one bucket per group corresponds to a complete ranking as studied by Coolen-Maturi [20]. At the other extreme, combining all groups into a single bucket makes their relative rank irrelevant. Between these extremes, the minimal meaningful number of buckets is two. Special cases involving two buckets have been investigated by Coolen and van der Laan [18], including selecting the best group, selecting the best subset, and selecting a subset containing the best group.

Previous sections have examined related cases. Section 2.2 considers scenarios with $w + 1$ buckets, where w buckets each contain a single group and the remaining groups are placed in one bucket. Sections 2.3 and 2.4 explore cases with three and four buckets respectively, again with all but one bucket containing a single group. In contrast, this chapter assumes a fixed number of buckets, selected by the user to best fit the specific application.

The chapter begins by deriving NPI lower and upper probabilities for the ranking events of interest in Section 5.2. Section 5.3 addresses how to assign groups to buckets using statistics derived from the data, for example, ranking groups by median values to guide the allocation. Next, the probabilities of these events of interest are estimated using the NPI-bootstrap method described in Section 5.4. Their relation-

ship is illustrated through examples, showing how the bootstrap estimates typically lie between the NPI bounds.

Section 5.5 then considers how many buckets should be used by evaluating all possible group-to-bucket assignments, selecting the allocation that either maximises the lower and upper bounds for the NPI lower and upper probabilities or minimises NPI lower and upper expected losses. Finally, Section 5.6 applies two different loss functions to quantify the loss incurred from incorrect rankings of the buckets while ignoring the ranking of the groups within each bucket.

This chapter thus presents a flexible and practical framework for ranking independent groups organised into buckets, addressing both the allocation of groups and the evaluation of bucket rankings.

5.2 Generalising the event of interest with buckets

Suppose that there are k independent groups, $X_1, X_2, \ldots, X_k$, and the set of corresponding indices is $K = \{1, 2, \ldots, k\}$. The number of buckets s is assumed to be fixed, the number of ways to assign the groups to buckets, neglecting the rank of the groups within each bucket is $\frac{k!}{k_1! k_2! \ldots k_s!}$, where k_i is the number of groups assigned to bucket i, $i = 1, 2, \ldots, s$. Let $K_1, K_2, \ldots, K_s$ be the partition of the set of indices K, such as $K_i \cap K_t = \emptyset$ for $i \neq t$, $\cup_{i=1}^{s} K_i = K$ and $K_i \neq \emptyset$ for all $i = 1, 2, \ldots, s$. That is $k_i = ||K_i||$ and $\sum_{i=1}^{s} k_i = k$.

Let $X_1^f, X_2^f, \ldots, X_k^f$ be the next future observations from the k groups, and let the generalised event of interest be

$$\{X_t^f\}_{t \in K_1} < \{X_t^f\}_{t \in K_2} < \ldots < \{X_t^f\}_{t \in K_s} \tag{5.1}$$

which can also be written as

$$\left\{ \max_{t \in K_1} X_t^f < \min_{t \in K_2} X_t^f \right\} \wedge \left\{ \max_{t \in K_2} X_t^f < \min_{t \in K_3} X_t^f \right\} \wedge \ldots \wedge \left\{ \max_{t \in K_{s-1}} X_t^f < \min_{t \in K_s} X_t^f \right\} \tag{5.2}$$

or in a compact way, the event of interest is

$$A = \bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \right\} \tag{5.3}$$

Calculating the NPI lower and upper probabilities for the event A is computationally intensive and complex. As noted by Coolen-Maturi [20] in the case where each bucket contains only one group, obtaining exact NPI lower and upper probabilities can be challenging. Therefore, this chapter follows their approach by employing approximations that serve as a lower bound for the NPI lower probability and an upper bound for the NPI upper probability.

For each group t ($t = 1, 2, \ldots, k$), let the observed values be ordered as $x_{t,1} < x_{t,2} < \ldots < x_{t,n_t}$, and $x_{t,0} = -\infty$ and $x_{t,n_t+1} = \infty$ defined for ease of notation. The inference depends on Hill's assumption $A_{(n_t)}$ for each group t. Then the lower bound for the NPI lower probability and the upper bound for the NPI upper probability for the event of interest A in Equation (5.3) are

$$\underline{P}^L(A) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \prod_{i=1}^{s-1} I\left(\max_{t \in K_i}(x_{t,l_t}) < \min_{t \in K_{i+1}} (x_{t,l_t-1}) \right) \quad (5.4)$$

$$\overline{P}^U(A) = \frac{1}{\prod_{t=1}^k (n_t + 1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \prod_{i=1}^{s-1} \prod_{a=i}^{s} I\left(\max_{t \in K_i} (x_{t,l_t-1}) < \min_{t \in K_{a+1}} (x_{t,l_t}) \right)$$

$$(5.5)$$

For the lower bound for the NPI lower probability, total separation is required for the intervals across buckets. That is, for all possible configurations $\prod_{t=1}^k (n_t + 1)$, the number of configurations is counted, such that all intervals in bucket 1 are left of all intervals in bucket 2, and all intervals in bucket 2 are to the left of all intervals in bucket 3, and so on. For the upper bound of the NPI upper probability, all possible configurations of the intervals across buckets are counted for a given configuration. These bounds are presented in Section 2.2.2 for the event $\max_{j_z \in NS} X_{j_z}^f < X_{\ell_w}^f < \ldots < X_{\ell_2}^f < X_{\ell_1}^f$.

5.3 Assigning groups to buckets for defining events of interest

To address the question of how to allocate groups to buckets and select the event of interest, as discussed earlier in this chapter, one can use statistics derived from the data to rank the groups. For example, groups with the smallest medians could

Groups	Data										
X	1.2	1.3	0.7	0.5	0.9	0.2	0.2	1.3	1.0	0.8	0.3
	1.2	1.3	0.9	0.5	0.9	1.5	9.8	2.1	1.5	4.2	2.1
	0.9	0.1	1.1	0.9	7.7	4.7	1.7	1.1	0.3	20.0	0.0
	1.0	1.6	0.7	1.2	0.4	3.9	1.5	0.9	3.4	21.0	1.0
	4.0	1.1	2.2	12.0	0.5	4.7	3.0	0.1	4.2	0.7	0.9
	0.9	4.3	2.7	0.2	2.3	0.7	1.7	0.9			
Y	1.4	1.2	1.7	1.6	1.4	0.3	1.4	0.5	1.9	4.1	1.1
	0.0	2.6	1.5	0.5	1.1	1.1	0.4	3.3	1.8	2.6	2.2
	0.7	0.3	0.1	1.3	1.6	1.7	2.5	1.0	0.7		

Table 5.1: Fasting serum growth hormone levels for two groups: middle-aged X and elderly Y individuals for Example 5.3.1.

be placed in the first bucket, those with the next-smallest medians in the second bucket, and so on. There are many possible ways to allocate one or more groups to one or more buckets. However, to keep this section concise, it is assumed that the number of buckets and the allocation of groups to these buckets depend on the user's objective. The median of each group is represented as m_{ℓ_t}, where $t = 1, \ldots, k$. The ranked median will be denoted by

$$m_{\ell_1} < m_{\ell_2} < \ldots < m_{\ell_k} \tag{5.6}$$

It is assumed that each group has an associated median value, where the best group has the highest median m_{ℓ_k}, the second-best has the second-highest median $m_{\ell_{k-1}}$, and so on. This approach involves placing groups with the smallest medians in the first bucket, those with the second-smallest medians in the second bucket, and so on. However, this method does not necessarily yield the highest lower and upper bounds for the NPI lower and upper probabilities. The following example illustrates this point.

Example 5.3.1 The data from Hansen [44] examine the response of fasting serum growth hormone to exercise in individuals with maturity-onset diabetes. The primary objective was to investigate whether exercise could raise levels of this hormone.

Table 5.1 presents the observed values for two groups: group X represents middle-aged individuals, and group Y represents elderly individuals. Each value corresponds to the fasting serum growth hormone level (in ng/ml) measured after an exercise session.

The median of group X is 1.1 and the median of group Y is 1.4. It is clear that $m_X < m_Y$. Based on that, the event of interest is $X^f < Y^f$, which means that the future fasting serum growth hormone level of a middle-aged individual after an exercise session is less than that of an elderly individual, which could be expected to yield the largest NPI lower and upper probabilities. However, in this example, it was found that the complementary event, $X^f \geq Y^f$, actually yields slightly higher NPI lower and upper probabilities. Specifically, the results are close, but indicate that X^f is somewhat more likely to be greater than Y^f, with $\underline{P}^L(X^f \geq Y^f) = 0.4692$, $\overline{P}^U(X^f \geq Y^f) = 0.5151$, compared to $\underline{P}^L(X^f < Y^f) = 0.4599$, and $\overline{P}^U(X^f < Y^f) = 0.5058$. Although the difference is small, this example highlights that defining the event of interest just by ranking the medians does not always correspond to the highest NPI lower and upper probabilities.

$\diamond$

5.4   NPI-Bootstrap for estimating the probability of the event

Deriving the exact NPI lower and upper probabilities for the general event of interest is computationally complex, as discussed in Section 5.2, where we introduced approximations in the form of bounds. In this section, we focus on a specialised application of the NPI-Bootstrap method, as proposed by Coolen and BinHimd [14], to approximate the probability of the general event of interest itself, rather than its NPI lower and upper probabilities. We begin with an overview of the NPI-Bootstrap method, followed by an explanation of how a particular case of this method can be used to approximate our event of interest. As will be demonstrated, this approach offers an attractive and computationally efficient alternative for estimating the prob-

ability of the event. We start by outlining the NPI-B method introduced by Coolen and BinHimd [14]. This method involves generating $n + 1$ intervals between the n observations from the original data and then randomly selecting one of these intervals.

The real line $(-\infty, \infty)$ is divided into three parts: the first part is $(-\infty, x_1)$ or $(0, x_1)$, the second part contains all the intervals between x_1 and x_n, and the last interval is (x_n, ∞). Regarding the second part, which contains $n - 1$ intervals, one interval is sampled randomly. Then, sample a future observation uniformly from that interval, and then this observation is added to the original data, resulting in $n + 1$ observations. This results in the formation of a partition comprising $n + 2$ intervals, from which the second observation is selected.

This method handles the first interval $(-\infty, x_1)$ and the last interval (x_n, ∞) differently using the Normal distribution tails for real-valued data. One future observation is sampled from the tail of the Normal distribution with mean $\mu = \frac{x_1 + x_n}{2}$ and standard deviation $\sigma = \frac{x_n - \mu}{\Phi^{-1}(\frac{n}{n+1})}$, where Φ denotes the standard Normal cumulative distribution function.

The procedure continues until a predetermined number, denoted as m, of future observations is sampled. This set of m observations forms a single NPI-B sample (excluding the n original data observations). Repeat all these steps B times, where B is a chosen integer value, to get a total of NPI-B bootstrap samples.

However, this process is performed only once, as in this thesis the number of future observations considered is one per group, $m = 1$. Therefore, one interval is sampled randomly and then the future observation is sampled either uniformly from the middle intervals or from a Normal distribution over the first $(-\infty, x_1)$ or the last (x_n, ∞) interval. In NPI-B, all possible orderings of the new observations among the past observations are equally likely to occur, $\frac{1}{\binom{n+m}{n}}$; in this thesis, that reduces to $\frac{1}{n + 1}$. Thus, this NPI-bootstrap sample will be used to estimate the probability of the event of interest, $A = \bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \right\}$, as follows

$$P_B(A) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \prod_{i=1}^{s-1} I\left(\max_{t \in K_i} (x^*_{t,l_t}) < \min_{t \in K_{i+1}} (x^*_{t,l_t}) \right) \quad (5.7)$$

where $x^*_{t,l_t} \in (x_{t,l_t-1}, x_{t,l_t})$ refers to the bootstrap-sampled observation. The sampling is done with equal probability $1/(n_t+1)$ across all (n_t+1) intervals defined by the ordered data. For the middle intervals, those between the first and last observation values are sampled uniformly. For the first and last intervals, sampling is performed from assumed Normal tails.

In addition, 1000 bootstrap replications are used, a number commonly adopted in the literature, especially for hypothesis testing and constructing confidence intervals.

The lower bound for the NPI lower probability and the upper bound for the NPI upper probability of the event $A = \bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \right\}$ are briefly illustrated in the following example. In this example, these bounds obtained using Equations (5.4) and (5.5) are compared with the probability of the event estimated using the NPI-bootstrap method based on Equation (5.7).

Example 5.4.1 The dataset provided in this example is sourced from the literature, Simpson and Margolin [68]. The data record revertant-colony counts from three replicate Ames assays on Salmonella TA98 at six dose levels of Acid Red 114. The number of visible revertant colonies at each dose levels was nine. The classification of six doses into different buckets based on their medians should be performed first. The total number of ways to arrange the doses in s buckets can be considered, ensuring that the first bucket contains the lowest medians and the last bucket contains the highest medians. In this example, we will consider the following scenario in which $k = 6$ doses are divided into $s = 3$ buckets, $k_1 = 2, k_2 = 2$ and $k_3 = 2$, based on their medians $m_1 = 22, m_2 = 25, m_3 = 35, m_4 = 43, m_5 = 28$ and $m_6 = 16$, the ranked median is $m_6 < m_1 < m_2 < m_5 < m_3 < m_4$. Therefore, the event of interest is $\left\{ \max(X_6^f, X_1^f) < \min(X_2^f, X_5^f) \right\} \wedge \left\{ \max(X_2^f, X_5^f) < \min(X_3^f, X_4^f) \right\}$. Note that X_t^f is the future observation from group t, where $t = 1, \ldots, 6$. The goal is to find lower and upper bounds for the NPI lower and upper probabilities for the event and then to find an estimated probability using the NPI-bootstrap method.

For the lower bound for the NPI lower probability, total separation for the intervals is required across buckets 1, 2 and 3. The lower bound for the NPI lower probability is

$$
\underline{P}^L\left(\left\{\max\left(X_6^f, X_1^f\right) < \min\left(X_2^f, X_5^f\right)\right\} \wedge \left\{\max\left(X_2^f, X_5^f\right) < \min\left(X_3^f, X_4^f\right)\right\}\right)
$$
$$
= \frac{1}{10^6}\sum_{l_1=1}^{n_1+1}\ldots\sum_{l_6=1}^{n_6+1} I\left(\max\left(x_{6,l_6}, x_{1,l_1}\right) < \min\left(x_{2,l_2-1}, x_{5,l_5-1}\right)\right)
$$
$$
I\left(\max\left(x_{2,l_2}, x_{5,l_5}\right) < \min\left(x_{3,l_3-1}, x_{4,l_4-1}\right)\right) = 0.1213
$$

For the upper bound for the NPI upper probability, all possible configurations of the intervals across buckets 1, 2 and 3 are counted. The upper bound for the NPI upper probability is

$$
\overline{P}^U\left(\left\{\max\left(X_6^f, X_1^f\right) < \min\left(X_2^f, X_5^f\right)\right\} \wedge \left\{\max\left(X_2^f, X_5^f\right) < \min\left(X_3^f, X_4^f\right)\right\}\right)
$$
$$
= \frac{1}{10^6}\sum_{l_1=1}^{n_1+1}\ldots\sum_{l_6=1}^{n_6+1} I\left(\max\left(x_{6,l_6-1}, x_{1,l_1-1}\right) < \min\left(x_{2,l_2}, x_{5,l_5}\right)\right)
$$
$$
I\left(\max\left(x_{6,l_6-1}, x_{1,l_1-1}\right) < \min\left(x_{3,l_3}, x_{4,l_4}\right)\right)
$$
$$
I\left(\max\left(x_{2,l_2-1}, x_{5,l_5-1}\right) < \min\left(x_{3,l_3}, x_{4,l_4}\right)\right) = 0.2168
$$

To estimate the probability of this event using the NPI-bootstrap method described above, note that the original dataset has endpoints at 8 and 98, as reported by Simpson and Margolin [68]. Start by randomly sampling one interval from each dose; then, one future observation is sampled either uniformly from any intervals between $(8, 98)$ or normally from interval $(-\infty, 8)$ or $(98, \infty)$. This process is repeated 1000 times to approximate the probability. Using Equation (5.7), the estimated probability is found to be 0.1500. The estimated probability result clearly lies in between the lower bound for the NPI lower probability which is 0.1213 and the upper bound for the NPI upper probability which is 0.2168.

$\diamond$

Groups	Data									
1	49	50	53	57	59	70	71	78	82	85
2	67	75	90	94	97	114	127	130	140	151
3	60	63	66	81	98	110	128	133	144	157
4	89	91	100	104	118	120	123	137	147	149

Table 5.2: Observed values for four groups with ten observations each for Example 5.5.1.

5.5 Selecting the number of buckets

In Section 5.2, it is assumed that the number of buckets is fixed. However, when it comes to assigning groups into buckets, there are several approaches to consider. For example, one could consider all possible allocation methods and select one that meets a specific criterion, such as maximising the lower and upper bounds for the NPI lower and upper probabilities or minimising the lower and upper bounds for the NPI lower and upper losses of an event. Alternatively, one could use statistics drawn from the data to rank the groups, such as putting those with the smallest medians in the first bucket, those with the second smallest medians in the second bucket, and so on. Ultimately, the allocation method may depend on the user and the specific application at hand. When there are many groups and buckets to consider, exploring all the options to assign groups into buckets can be challenging. However, with a small number of groups, all possible events of interest can be thoroughly examined.

Next, an example is provided with just four groups which are manageable to consider all bucket options: 2, 3 and 4. The goal is to investigate how the number of buckets and the number of groups within the buckets impact the selection and ranking.

Example 5.5.1 Neuhauser et al. [62] provide an example of four groups with ten observations each. Table 5.2 shows this dataset. The number of ways to partition k groups into s ordered, non-empty buckets, where the order of groups within each bucket does not matter, is given by multiplying $s!$ by the Stirling number of the

Event	$\underline{P}^L$	$\overline{P}^U$	P_B
$X_1^f < \min(X_2^f, X_3^f, X_4^f)$	0.5314	0.8144	0.6870
$X_2^f < \min(X_1^f, X_3^f, X_4^f)$	0.0382	0.1700	0.0840
$X_3^f < \min(X_1^f, X_2^f, X_4^f)$	0.1134	0.2689	0.2060
$X_4^f < \min(X_1^f, X_2^f, X_3^f)$	0.0000	0.1102	0.0170
$\max(X_1^f, X_2^f) < \min(X_3^f, X_4^f)$	0.2288	0.3815	0.3330
$\max(X_1^f, X_3^f) < \min(X_2^f, X_4^f)$	0.3026	0.4696	0.4050
$\max(X_1^f, X_4^f) < \min(X_2^f, X_3^f)$	0.1325	0.2630	0.2070
$\max(X_2^f, X_3^f) < \min(X_1^f, X_4^f)$	0.0191	0.0914	0.0380
$\max(X_2^f, X_4^f) < \min(X_1^f, X_3^f)$	0.0000	0.0460	0.0030
$\max(X_3^f, X_4^f) < \min(X_1^f, X_2^f)$	0.0000	0.0544	0.0070
$\max(X_1^f, X_2^f, X_3^f) < X_4^f$	0.2759	0.4929	0.4040
$\max(X_1^f, X_2^f, X_4^f) < X_3^f$	0.2097	0.3908	0.3100
$\max(X_1^f, X_3^f, X_4^f) < X_2^f$	0.1974	0.3847	0.3160
$\max(X_2^f, X_3^f, X_4^f) < X_1^f$	0.0000	0.0952	0.0020

Table 5.3: Four groups allocating into two buckets, $s = 2$ for Example 5.5.1

second kind, $S(k, s)$. The Stirling number $S(k, s)$ counts the number of ways to partition k labelled groups into s unordered, non-empty buckets, while multiplying by $s!$ accounts for the ordering of these buckets, converting unordered partitions into ordered partitions. The number of ordered partitions is given by the formula $s! \cdot S(k, s)$, with initial values $S(0, 0) = 1$, and $S(k, 0) = S(0, s) = 0$ for $k, s > 0$ [36]. Stirling numbers of the second kind can be efficiently calculated in R using the *stirling*2 function from the *Brobdingnag* package.

When $k = 4$ groups are partitioned into s buckets, the number of ways to assign these groups varies with s. When there is only one bucket, $s = 1$, there is exactly one way to assign all groups together. For two buckets, $s = 2$, the total number of ordered partitions is 14. For three buckets, $s = 3$, resulting in 36 ways to assign the groups. Finally, when each group forms its own bucket, $s = 4$, there are 24 possible ordered partitions.

From Table 5.2, it is clear that all the observations that belong to group 1 are less

Event	$\underline{P}^L$	$\overline{P}^U$	P_B
$(X_1^f < X_2^f) \wedge (X_2^f < \min(X_3^f, X_4^f))$	0.1933	0.2930	0.2510
$(X_1^f < X_3^f) \wedge (X_3^f < \min(X_2^f, X_4^f))$	0.2056	0.3080	0.2620
$(X_1^f < X_4^f) \wedge (X_4^f < \min(X_2^f, X_3^f))$	0.1325	0.2126	0.1540
$(X_2^f < X_1^f) \wedge (X_1^f < \min(X_3^f, X_4^f))$	0.0355	0.1150	0.0860
$(X_2^f < X_3^f) \wedge (X_3^f < \min(X_1^f, X_4^f))$	0.0027	0.0343	0.0150
$(X_2^f < X_4^f) \wedge (X_4^f < \min(X_1^f, X_3^f))$	0.0000	0.0200	0.0020
$(X_3^f < X_1^f) \wedge (X_1^f < \min(X_2^f, X_4^f))$	0.0970	0.1961	0.1400
$(X_3^f < X_2^f) \wedge (X_2^f < \min(X_1^f, X_4^f))$	0.0164	0.0496	0.0220
$(X_3^f < X_4^f) \wedge (X_4^f < \min(X_1^f, X_2^f))$	0.0000	0.0225	0.0080
$(X_4^f < X_1^f) \wedge (X_1^f < \min(X_2^f, X_3^f))$	0.0000	0.0658	0.0090
$(X_4^f < X_2^f) \wedge (X_2^f < \min(X_1^f, X_3^f))$	0.0000	0.0193	0.0010
$(X_4^f < X_3^f) \wedge (X_3^f < \min(X_1^f, X_2^f))$	0.0000	0.0245	0.0030
$(X_1^f < \min(X_2^f, X_3^f)) \wedge (\max(X_2^f, X_3^f) < X_4^f)$	0.1927	0.2802	0.2470
$(X_1^f < \min(X_2^f, X_4^f)) \wedge (\max(X_2^f, X_4^f) < X_3^f)$	0.1889	0.2796	0.2620
$(X_1^f < \min(X_3^f, X_4^f)) \wedge (\max(X_3^f, X_4^f) < X_2^f)$	0.1498	0.2330	0.2150
$(X_2^f < \min(X_1^f, X_3^f)) \wedge (\max(X_1^f, X_3^f) < X_4^f)$	0.0175	0.0705	0.0540
$(X_2^f < \min(X_1^f, X_4^f)) \wedge (\max(X_1^f, X_4^f) < X_3^f)$	0.0208	0.0590	0.0480
$(X_2^f < \min(X_3^f, X_4^f)) \wedge (\max(X_3^f, X_4^f) < X_1^f)$	0.0000	0.0300	0.0020
$(X_3^f < \min(X_1^f, X_2^f)) \wedge (\max(X_1^f, X_2^f) < X_4^f)$	0.0658	0.1242	0.0860
$(X_3^f < \min(X_1^f, X_4^f)) \wedge (\max(X_1^f, X_4^f) < X_2^f)$	0.0476	0.0941	0.0730
$(X_3^f < \min(X_2^f, X_4^f)) \wedge (\max(X_2^f, X_4^f) < X_1^f)$	0.0000	0.0387	0.0000
$(X_4^f < \min(X_1^f, X_2^f)) \wedge (\max(X_1^f, X_2^f) < X_3^f)$	0.0000	0.0380	0.0150
$(X_4^f < \min(X_1^f, X_3^f)) \wedge (\max(X_1^f, X_3^f) < X_2^f)$	0.0000	0.0421	0.0070
$(X_4^f < \min(X_2^f, X_3^f)) \wedge (\max(X_2^f, X_3^f) < X_1^f)$	0.0000	0.0220	0.0000

Table 5.4: Four groups allocating into three buckets, $s = 3$ for Example 5.5.1, part 1

than all the observations that belong to group 4. In addition, Tables 5.3, 5.4, 5.5 and 5.6 present the lower and upper bounds for the NPI lower and upper probabilities and the estimated probability P_B for all possible events considered. It is clear that

Event	$\underline{P}^L$	$\overline{P}^U$	P_B
$(\max(X_1^f, X_2^f) < X_3^f) \wedge (X_3^f < X_4^f)$	0.1018	0.1878	0.1540
$(\max(X_1^f, X_2^f) < X_4^f) \wedge (X_4^f < X_3^f)$	0.1270	0.2201	0.2140
$(\max(X_1^f, X_3^f) < X_2^f) \wedge (X_2^f < X_4^f)$	0.1550	0.2569	0.2380
$(\max(X_1^f, X_3^f) < X_4^f) \wedge (X_4^f < X_2^f)$	0.1475	0.2472	0.1810
$(\max(X_1^f, X_4^f) < X_3^f) \wedge (X_3^f < X_2^f)$	0.0499	0.1190	0.1060
$(\max(X_1^f, X_4^f) < X_2^f) \wedge (X_2^f < X_3^f)$	0.0826	0.1594	0.1190
$(\max(X_2^f, X_3^f) < X_1^f) \wedge (X_1^f < X_4^f)$	0.0191	0.0473	0.0280
$(\max(X_2^f, X_3^f) < X_4^f) \wedge (X_4^f < X_1^f)$	0.0000	0.0365	0.0030
$(\max(X_2^f, X_4^f) < X_1^f) \wedge (X_1^f < X_3^f)$	0.0000	0.0105	0.0040
$(\max(X_2^f, X_4^f) < X_3^f) \wedge (X_3^f < X_1^f)$	0.0000	0.0288	0.0000
$(\max(X_3^f, X_4^f) < X_1^f) \wedge (X_1^f < X_2^f)$	0.0000	0.0178	0.0050
$(\max(X_3^f, X_4^f) < X_2^f) \wedge (X_2^f < X_1^f)$	0.0000	0.0291	0.0000

Table 5.5: Four groups allocating into three buckets, $s = 3$ for Example 5.5.1, part 2

the estimated probability is either between the lower and upper bounds for the NPI lower and upper probabilities or equal to one of them. For example, from Table 5.4 and for the event $(X_4^f < \min(X_2^f, X_3^f)) \wedge (\max(X_2^f, X_3^f) < X_1^f)$, the lower bound for the NPI lower probability is equal to the estimated probability, which is 0. Also, in Table 5.6, for the event $X_3^f < X_4^f < X_1^f < X_2^f$, the lower bound for the NPI lower probability, the upper bound for the NPI upper probability and estimated probabilities have an identical value of 0. Some events result in the lower bound for the NPI lower probability, the upper bound for the NPI upper probability or the estimated probability of 0. The reason is that these events consider either group 4 to provide the smallest future observation or group 1 to provide the largest future observation, such as $X_4^f < \min(X_1^f, X_2^f, X_3^f)$ and $\max(X_2^f, X_3^f, X_4^f) < X_1^f$.

It is also clear that the lower bound for the NPI lower probability and the upper bound for the NPI upper probability for all bucket options $s = 2$, $s = 3$, and $s = 4$ are mostly small values, except for some events in Table 5.3, where they are around 0.5. For example, the lower bound for the NPI lower probability for the events

Event	$\underline{P}^L$	$\overline{P}^U$	P_B
$X_1^f < X_2^f < X_3^f < X_4^f$	0.0870	0.1282	0.1080
$X_1^f < X_2^f < X_4^f < X_3^f$	0.1063	0.1578	0.1320
$X_1^f < X_3^f < X_2^f < X_4^f$	0.1057	0.1520	0.1315
$X_1^f < X_3^f < X_4^f < X_2^f$	0.0999	0.1496	0.1303
$X_1^f < X_4^f < X_2^f < X_3^f$	0.0826	0.1217	0.1120
$X_1^f < X_4^f < X_3^f < X_2^f$	0.0499	0.0834	0.0800
$X_2^f < X_1^f < X_3^f < X_4^f$	0.0148	0.0547	0.0310
$X_2^f < X_1^f < X_4^f < X_3^f$	0.0208	0.0590	0.0380
$X_2^f < X_3^f < X_1^f < X_4^f$	0.0027	0.0158	0.0110
$X_2^f < X_3^f < X_4^f < X_1^f$	0.0000	0.0147	0.0010
$X_2^f < X_4^f < X_1^f < X_3^f$	0.0000	0.0000	0.0000
$X_2^f < X_4^f < X_3^f < X_1^f$	0.0000	0.0153	0.0000
$X_3^f < X_1^f < X_2^f < X_4^f$	0.0494	0.1002	0.0760
$X_3^f < X_1^f < X_4^f < X_2^f$	0.0476	0.0941	0.0850
$X_3^f < X_2^f < X_1^f < X_4^f$	0.0164	0.0240	0.0170
$X_3^f < X_2^f < X_4^f < X_1^f$	0.0000	0.0212	0.0010
$X_3^f < X_4^f < X_1^f < X_2^f$	0.0000	0.0000	0.0000
$X_3^f < X_4^f < X_2^f < X_1^f$	0.0000	0.0176	0.0030
$X_4^f < X_1^f < X_2^f < X_3^f$	0.0000	0.0339	0.0060
$X_4^f < X_1^f < X_3^f < X_2^f$	0.0000	0.0312	0.0030
$X_4^f < X_2^f < X_1^f < X_3^f$	0.0000	0.0041	0.0020
$X_4^f < X_2^f < X_3^f < X_1^f$	0.0000	0.0117	0.0010
$X_4^f < X_3^f < X_1^f < X_2^f$	0.0000	0.0109	0.0000
$X_4^f < X_3^f < X_2^f < X_1^f$	0.0000	0.0103	0.0000

Table 5.6: Four groups allocating into four buckets, $s = 4$ for Example 5.5.1

$X_1^f < \min(X_2^f, X_3^f, X_4^f)$ and the upper bound for the NPI upper probability for the events $\max(X_1^f, X_3^f) < \min(X_2^f, X_4^f)$ are around 0.5 which might be considered a bit large compared with the rest.

In conclusion, Table 5.3 shows that the number of buckets leading to a good

decision, defined as those with high values for the lower bound of the NPI lower probability and the upper bound of the NPI upper probability, is two $s = 2$. Specifically, one group is in the first bucket and three groups are in the second bucket for the event $X_1^f < \min(X_2^f, X_3^f, X_4^f)$. The lower bound for the NPI lower probability for this event is 0.5314, the upper bound for the NPI upper probability for this event is 0.8144, and the estimated probability equal to 0.6870. Moreover, for every event considered, the estimated probability P_B always lies between or equal to either the lower bound for the NPI lower probability $\underline{P}^L$ or the upper bound for the NPI upper probability $\overline{P}^U$. Also, many events under $s = 3$ or $s = 4$ yield very small NPI lower and upper probabilities, often zero.

$$\diamond$$

5.6   General event of selection and ranking based on loss functions

In this section, the aim is to measure the penalty associated with the ranking of the suboptimal buckets and to reflect on the quality of the ranking, without considering the ranking of the groups within each bucket, using a loss function for the event $A = \bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \right\}$. The general multi-level loss function considered in Chapter 4 could be considered in this section. It should be noted that the general multi-level loss function provides graded penalty for different level of ranking errors, which leads to large number of loss levels. For example, in Example 5.5.1, the number of loss levels should be 75, each representing a different ordering of the groups. Therefore, linear and quadratic loss functions are considered as an alternative of a general multi-level loss function. The linear loss function is considered in Section 5.6.1, while the quadratic loss function is considered in Section 5.6.2. They focus on penalising the suboptimal ordering of buckets without accounting for the ranking of groups within each bucket.

5.6.1   Linear loss function

The linear loss function measures the distance between future observations in the bucket i and future observations in the bucket $i+1$. The penalty increases linearly with the difference. This helps to measure the penalty associated with ranking the suboptimal s buckets in terms of one or more future observations within each bucket. The linear loss function for the general event $A = \bigcap\limits_{i=1}^{s-1} \left\{ \max\limits_{t \in K_i} X_t^f < \min\limits_{t \in K_{i+1}} X_t^f \right\}$ can be defined as

$$L_l\left(\bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f, \min_{t \in K_{i+1}} X_t^f \right\} \right) = \begin{cases} 0 & \text{if A is true} \\ F & \text{otherwise} \end{cases} \tag{5.8}$$

where $F = \sum\limits_{i=1}^{s-1} \max \left\{ \max\limits_{t \in K_i} X_t^f - \min\limits_{t \in K_{i+1}} X_t^f, 0 \right\}$. If the condition, which is A is not satisfied, the loss that is F is defined as the sum of the individual loss between consecutive buckets.

This loss function measures the penalty associated with the suboptimal ranking of observations in each bucket K_i relative to those in the next bucket K_{i+1}, considering all buckets from 1 to s. A lower value of $L_l\left(\sum\limits_{i=1}^{s-1} \left\{ \max\limits_{t \in K_i} X_t^f, \min\limits_{t \in K_{i+1}} X_t^f \right\} \right)$ indicates a smaller deviation from the desired ranking between the future observations in bucket i and the future observations in bucket $i+1$, while a larger value suggests a greater deviation between the future observations in bucket i and the future observations in bucket $i+1$.

The discussion now turns to the NPI lower and upper expected linear losses. Calculating the NPI lower and upper expected linear losses for the event A is computationally intensive and complex. Therefore, this chapter employs approximations that serve as a lower bound for the NPI lower expected linear loss and an upper bound for the NPI upper expected linear loss.

For each group t ($t = 1, 2, \ldots, k$), let the observed values be ordered as $x_{t,1} < x_{t,2} < \ldots < x_{t,n_t}$, and $x_{t,0} = l$ and $x_{t,n_t+1} = r$ defined for ease of notation. future observations are bounded by l and r, set below the minimum and above the maximum observed data points, respectively. The appearance of $-\infty$ or ∞ affects the calculation of the lower and upper bounds for the NPI lower and upper expected linear losses. The inference depends on Hill's assumption $A_{(n_t)}$ for each group t.

For the lower bound for the NPI lower expected linear loss, the sum of the losses for pairwise comparison across buckets is calculated while the probability mass for groups belonging to bucket K_i is assigned to the left endpoint per interval, and the probability mass for groups belonging to bucket K_{i+1} is assigned to the right endpoint per interval. For the upper bound of the NPI upper probability, again the sum of the losses for pairwise comparison across buckets is calculated while the probability mass for groups belonging to bucket K_i is assigned to the right endpoint per interval, and the probability mass for groups belonging to bucket K_{i+1} is assigned to the left endpoint per interval. Then, the lower bound for the NPI lower expected linear loss and the upper bound for the NPI upper expected linear loss for the event of interest A in Equation (5.3) are

$$\underline{E}^L(L_l(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_l\Big(\max_{t\in K_i}(x_{t,l_t-1}) < \min_{t\in K_{i+1}}(x_{t,l_t})\Big)$$
(5.9)

$$\overline{E}^U(L_l(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_l\Big(\max_{t\in K_i}(x_{t,l_t}) < \min_{t\in K_{a+1}}(x_{t,l_t-1})\Big)$$
(5.10)

Next, for further investigation, the quadratic loss function will be presented to measure the penalty associated with the ranking of the suboptimal s buckets in terms of one or more future observations within each bucket.

5.6.2   Quadratic loss function

The quadratic loss function for the general event $\bigcap_{i=1}^{s-1}\left\{\max_{t\in K_i} X_t^f < \min_{t\in K_{i+1}} X_t^f\right\}$ can be defined as

$$L_q\left(\bigcap_{i=1}^{s-1}\left\{\max_{t\in K_i} X_t^f, \min_{t\in K_{i+1}} X_t^f\right\}\right) = \begin{cases} 0 & \text{if A is true} \\ F & \text{otherwise} \end{cases}$$
(5.11)

where $F = \sum_{i=1}^{s-1}\left\{\max\left\{\max_{t\in K_i} X_t^f - \min_{t\in K_{i+1}} X_t^f, 0\right\}\right\}^2$.

The quadratic loss function measures the penalty associated with ranking suboptimal among consecutive buckets. In other words, the quadratic loss function

measures the difference between future observations in the bucket i and those in bucket $i+1$ across s buckets, squaring these differences to penalise larger gaps more heavily.

Then the lower bound for the NPI lower expected quadratic loss and the upper bound for the NPI upper expected quadratic loss for the event of interest A

$$\underline{E}^L(L_q(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_q\Big(\max_{t\in K_i}(x_{t,l_t-1}) < \min_{t\in K_{i+1}}(x_{t,l_t}) \Big)$$

(5.12)

$$\overline{E}^U(L_q(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_q\Big(\max_{t\in K_i}(x_{t,l_t}) < \min_{t\in K_{a+1}}(x_{t,l_t-1}) \Big)$$

(5.13)

Next, an example will be provided to illustrate and discuss the general event of selection and ranking based on the linear and quadratic loss function.

Example 5.6.1 In Example 5.5.1, the goal was to investigate and determine the optimal number of buckets and the optimal number of groups within buckets for four groups. Now, in this example, the aim is to measure the penalty associated with the ranking of suboptimal buckets for all possible number of buckets and the number of groups within buckets for four groups, and to examine whether the results align with those obtained in Example 5.5.1. The decision here depends on the lower and upper bounds of the NPI lower and upper expected linear and quadratic losses given in Sections 5.6.1 and 5.6.2 respectively.

From Equations (5.8) and (5.11) the loss is only calculated when the buckets do not follow the desired ranking. In addition, the loss is calculated based on given penalty values, which are linear and quadratic deviations between two consecutive buckets. The quadratic penalty penalises larger deviations more heavily, as seen with some events of interest in Tables 5.7, 5.8, 5.9 and 5.10.

To calculate the lower and upper bounds for the NPI lower and upper linear and quadratic losses, an assumption is made that the data observations are bounded by the values 48 and 158, as discussed in Section 3.4.2.

First, it is notable that in Tables 5.8, 5.9, and 5.10, some events of interest have either the lower bound for the NPI lower linear or quadratic expected losses equal

Event	$\underline{E^L}(L_l)$	$\overline{E^U}(L_l)$	$\underline{E^L}(L_q)$	$\overline{E^U}(L_q)$
$X_1^f < \min(X_2^f, X_3^f, X_4^f)$	2.221	13.255	35.316	804.748
$X_2^f < \min(X_1^f, X_3^f, X_4^f)$	39.513	53.898	2474.744	3862.426
$X_3^f < \min(X_1^f, X_2^f, X_4^f)$	35.573	49.706	2395.757	3768.141
$X_4^f < \min(X_1^f, X_2^f, X_3^f)$	47.889	62.385	3081.985	4546.632
$\max(X_1^f, X_2^f) < \min(X_3^f, X_4^f)$	18.567	31.247	955.981	1993.767
$\max(X_1^f, X_3^f) < \min(X_2^f, X_4^f)$	15.173	27.365	749.644	1737.983
$\max(X_1^f, X_4^f) < \min(X_2^f, X_3^f)$	25.778	39.014	1387.926	2541.986
$\max(X_2^f, X_3^f) < \min(X_1^f, X_4^f)$	44.554	59.088	2995.033	4571.409
$\max(X_2^f, X_4^f) < \min(X_1^f, X_3^f)$	50.478	63.391	3412.675	4916.874
$\max(X_3^f, X_4^f) < \min(X_1^f, X_2^f)$	49.007	63.089	3305.446	4931.718
$\max(X_1^f, X_2^f, X_3^f) < X_4^f$	14.076	28.190	555.556	1626.071
$\max(X_1^f, X_2^f, X_4^f) < X_3^f$	26.292	40.953	1551.024	2892.147
$\max(X_1^f, X_3^f, X_4^f) < X_2^f$	22.152	36.928	1126.253	2369.710
$\max(X_2^f, X_3^f, X_4^f) < X_1^f$	60.706	76.518	4537.619	6359.053

Table 5.7: Linear and quadratic loss functions for four groups allocating into two buckets, $s = 2$ for Example 5.6.1

to zero or both the lower and upper bounds for the NPI lower and upper linear or quadratic expected losses equal to zero, while this situation does not occur in Table 5.7. This occurs mainly when future observations related to the group 1 are less than those related to group 4. Also, These tables clearly present scenarios with more buckets. As a result, the events of interest imply additional conditions on the ranking, which leads to smaller values for the lower and upper bounds for the NPI lower and upper linear and quadratic expected losses. The smallest lower and upper bounds for the NPI lower and upper linear and quadratic expected losses in Table 5.7 are related to the event $X_1^f < \min(X_2^f, X_3^f, X_4^f)$, where $\underline{E^L}(L_l) = 2.221$, $\overline{E^U}(L_l) = 13.255$, $\underline{E^L}(L_q) = 35.316$, and $\overline{E^U}(L_q) = 804.748$, this result aligns with the result shown in Table 5.3, where this event has the highest NPI lower and upper probabilities to occur. The smallest NPI lower and upper linear and quadratic expected losses across all cases are associated with the events $X_3^f < X_1^f < X_4^f < X_2^f$

Event	$\underline{E}^L(L_l)$	$\overline{E^U}(L_l)$	$\underline{E}^L(L_q)$	$\overline{E^U}(L_q)$
$(X_1^f < X_2^f) \wedge (X_2^f < \min(X_3^f, X_4^f))$	0.286	5.593	5.186	466.394
$(X_1^f < X_3^f) \wedge (X_3^f < \min(X_2^f, X_4^f))$	0.038	4.694	0.524	360.501
$(X_1^f < X_4^f) \wedge (X_4^f < \min(X_2^f, X_3^f))$	0.000	5.850	0.000	526.135
$(X_2^f < X_1^f) \wedge (X_1^f < \min(X_3^f, X_4^f))$	5.445	16.031	335.222	1200.933
$(X_2^f < X_3^f) \wedge (X_3^f < \min(X_1^f, X_4^f))$	17.684	28.320	1261.878	2268.915
$(X_2^f < X_4^f) \wedge (X_4^f < \min(X_1^f, X_3^f))$	21.533	31.798	1582.906	2613.513
$(X_3^f < X_1^f) \wedge (X_1^f < \min(X_2^f, X_4^f))$	2.140	11.619	137.821	947.729
$(X_3^f < X_2^f) \wedge (X_2^f < \min(X_1^f, X_4^f))$	21.209	31.940	1615.259	2676.125
$(X_3^f < X_4^f) \wedge (X_4^f < \min(X_1^f, X_2^f))$	21.120	31.396	1632.086	2671.193
$(X_4^f < X_1^f) \wedge (X_1^f < \min(X_2^f, X_3^f))$	8.718	20.828	528.525	1538.427
$(X_4^f < X_2^f) \wedge (X_2^f < \min(X_1^f, X_3^f))$	26.714	37.846	1906.299	2999.489
$(X_4^f < X_3^f) \wedge (X_3^f < \min(X_1^f, X_2^f))$	22.760	34.040	1638.281	2717.004
$(X_1^f < \min(X_2^f, X_3^f)) \wedge (\max(X_2^f, X_3^f) < X_4^f)$	4.847	10.937	242.651	986.402
$(X_1^f < \min(X_2^f, X_4^f)) \wedge (\max(X_2^f, X_4^f) < X_3^f)$	6.068	9.768	410.346	1027.228
$(X_1^f < \min(X_3^f, X_4^f)) \wedge (\max(X_3^f, X_4^f) < X_2^f)$	6.705	12.166	403.525	1168.968
$(X_2^f < \min(X_1^f, X_3^f)) \wedge (\max(X_1^f, X_3^f) < X_4^f)$	20.402	40.808	1846.823	4302.378
$(X_2^f < \min(X_1^f, X_4^f)) \wedge (\max(X_1^f, X_4^f) < X_3^f)$	39.041	59.850	4017.900	6817.316
$(X_2^f < \min(X_3^f, X_4^f)) \wedge (\max(X_3^f, X_4^f) < X_1^f)$	56.803	68.761	6260.650	7937.445
$(X_3^f < \min(X_1^f, X_2^f)) \wedge (\max(X_1^f, X_2^f) < X_4^f)$	19.175	38.060	1735.271	4043.018
$(X_3^f < \min(X_1^f, X_4^f)) \wedge (\max(X_1^f, X_4^f) < X_2^f)$	31.139	50.804	3081.594	5666.770
$(X_3^f < \min(X_2^f, X_4^f)) \wedge (\max(X_2^f, X_4^f) < X_1^f)$	48.878	58.343	5389.616	6687.953
$(X_4^f < \min(X_1^f, X_2^f)) \wedge (\max(X_1^f, X_2^f) < X_3^f)$	38.410	61.591	4005.643	7071.520
$(X_4^f < \min(X_1^f, X_3^f)) \wedge (\max(X_1^f, X_3^f) < X_2^f)$	31.887	55.492	3189.693	6206.252
$(X_4^f < \min(X_2^f, X_3^f)) \wedge (\max(X_2^f, X_3^f) < X_1^f)$	66.109	81.858	7427.897	9625.469

Table 5.8: Linear and quadratic loss functions for four groups allocating into three buckets, $s = 3$ for Example 5.6.1, part 1

and $X_2^f < X_1^f < X_4^f < X_3^f$, which are equal to zero. However, the highest lower and upper bounds for NPI linear and quadratic expected losses for this table occur for the event $\max(X_2^f, X_3^f, X_4^f) < X_1^f$, where $\underline{E}^L(L_l) = 60.706$, $\overline{E^U}(L_l) = 76.518$, $\underline{E}^L(L_q) =$

Event	$\underline{E}^L(L_l)$	$\overline{E^U}(L_l)$	$\underline{E}^L(L_q)$	$\overline{E^U}(L_q)$
$(\max(X_1^f, X_2^f) < X_3^f) \wedge (X_3^f < X_4^f)$	2.056	8.773	89.101	612.815
$(\max(X_1^f, X_2^f) < X_4^f) \wedge (X_4^f < X_3^f)$	9.019	18.662	583.764	1446.125
$(\max(X_1^f, X_3^f) < X_2^f) \wedge (X_2^f < X_4^f)$	3.502	10.427	163.646	725.799
$(\max(X_1^f, X_3^f) < X_4^f) \wedge (X_4^f < X_2^f)$	7.172	15.886	426.192	1163.642
$(\max(X_1^f, X_4^f) < X_3^f) \wedge (X_3^f < X_2^f)$	5.181	13.508	283.499	971.810
$(\max(X_1^f, X_4^f) < X_2^f) \wedge (X_2^f < X_3^f)$	10.575	20.253	678.641	1551.636
$(\max(X_2^f, X_3^f) < X_1^f) \wedge (X_1^f < X_4^f)$	0.000	6.733	0.000	609.901
$(\max(X_2^f, X_3^f) < X_4^f) \wedge (X_4^f < X_1^f)$	32.951	47.775	2551.931	4038.842
$(\max(X_2^f, X_4^f) < X_1^f) \wedge (X_1^f < X_3^f)$	8.355	16.523	566.728	1340.080
$(\max(X_2^f, X_4^f) < X_3^f) \wedge (X_3^f < X_1^f)$	29.390	42.683	2253.864	3614.108
$(\max(X_3^f, X_4^f) < X_1^f) \wedge (X_1^f < X_2^f)$	3.653	11.284	237.752	943.668
$(\max(X_3^f, X_4^f) < X_2^f) \wedge (X_2^f < X_1^f)$	34.936	48.984	2696.366	4128.334

Table 5.9: Linear and quadratic loss functions for four groups allocating into three buckets, $s = 3$ for Example 5.6.1, part 2

4537.619, and $\overline{E^U}(L_q) = 6359.053$. The reason is that this event considers group 1 more likely to provide the largest future observation, even though all observations for group 1 are less than those for group 4.

Generally, from Examples 5.5.1 and 5.6.1, events involving either group 4 being more likely to produce the smallest future observation, group 1 being more likely to produce the largest future observation, or the event that the future observation for group 4 exceeds that of group 1 tend to result in high lower and upper bounds for the NPI lower and upper expected loss values, both linear and quadratic, and low NPI lower and upper probabilities.

In addition, tables that present scenarios with more buckets tend to lead to smaller values for the lower and upper bounds of the NPI lower and upper linear and quadratic expected losses. This is because events of interest imply additional ranking conditions, which supports better decision-making. Also, the imprecision cannot be calculated directly, as the results are expressed in terms of lower and upper bounds for the losses rather than exact NPI lower and upper linear and quadratic

Event	$\underline{E}^L(L_l)$	$\overline{E^U}(L_l)$	$\underline{E}^L(L_q)$	$\overline{E^U}(L_q)$
$X_1^f < X_2^f < X_3^f < X_4^f$	0.000	0.737	0.000	62.846
$X_1^f < X_2^f < X_4^f < X_3^f$	0.000	1.498	0.000	135.712
$X_1^f < X_3^f < X_2^f < X_4^f$	0.000	0.815	0.000	65.764
$X_1^f < X_3^f < X_4^f < X_2^f$	0.000	1.174	0.000	97.403
$X_1^f < X_4^f < X_2^f < X_3^f$	0.000	1.914	0.000	179.771
$X_1^f < X_4^f < X_3^f < X_2^f$	0.000	1.189	0.000	104.985
$X_2^f < X_1^f < X_3^f < X_4^f$	0.000	0.788	0.000	63.899
$X_2^f < X_1^f < X_4^f < X_3^f$	0.000	0.000	0.000	0.000
$X_2^f < X_3^f < X_1^f < X_4^f$	0.000	2.518	0.000	224.359
$X_2^f < X_3^f < X_4^f < X_1^f$	3.681	6.825	286.626	587.012
$X_2^f < X_4^f < X_1^f < X_3^f$	3.120	7.265	221.037	604.278
$X_2^f < X_4^f < X_3^f < X_1^f$	7.454	12.165	589.633	1046.658
$X_3^f < X_1^f < X_2^f < X_4^f$	0.000	0.333	0.000	29.348
$X_3^f < X_1^f < X_4^f < X_2^f$	0.000	0.000	0.000	0.000
$X_3^f < X_2^f < X_1^f < X_4^f$	0.000	3.019	0.000	284.030
$X_3^f < X_2^f < X_4^f < X_1^f$	6.239	9.908	504.383	856.285
$X_3^f < X_4^f < X_1^f < X_2^f$	1.380	4.908	98.008	433.614
$X_3^f < X_4^f < X_2^f < X_1^f$	8.200	13.103	671.869	1147.608
$X_4^f < X_1^f < X_2^f < X_3^f$	0.898	1.496	56.344	106.720
$X_4^f < X_1^f < X_3^f < X_2^f$	0.128	1.090	7.197	84.861
$X_4^f < X_2^f < X_1^f < X_3^f$	3.129	7.459	216.352	604.266
$X_4^f < X_2^f < X_3^f < X_1^f$	7.604	11.890	587.259	997.270
$X_4^f < X_3^f < X_1^f < X_2^f$	0.906	4.235	60.393	355.252
$X_4^f < X_3^f < X_2^f < X_1^f$	6.272	10.121	485.211	853.674

Table 5.10: Linear and quadratic loss functions for four groups allocating into four buckets, $s = 4$ for Example 5.6.1

expected losses.

$\diamond$

5.7 Concluding remarks

This chapter introduces NPI for the general event of selection and ranking, $A = \bigcap_{i=1}^{s-1} \left\{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \right\}$, which addresses the problem of ranking buckets that contain one or more independent groups. Each group must be assigned to one and only one bucket. The focus is on the ranking of the buckets, ignoring the rank of the groups within each bucket. A predetermined number of buckets s and the number of groups in each bucket k are assumed. In addition, an estimated probability and NPI lower and upper probabilities are obtained.

Two main questions are considered and answered: how to assign the groups into these buckets and how many buckets should be used? Two counterexamples are provided to illustrate the method and reach conclusions. To answer the first question, decision-making processes based on median values are performed. The median is considered because it is often preferred in nonparametric statistics compared to the mean or mode. The approach involves placing groups with the smallest medians in the first bucket, those with the second-smallest medians in the second bucket, and so on. It was found that this approach does not always result in the highest NPI lower and upper probabilities.

To answer the second question, all possible allocation methods are considered, and selects the allocation that either maximises the NPI lower and upper probabilities of an event, as detailed in Section 5.5, or minimises the NPI lower and upper expected losses. When dealing with a large number of groups and a potential number of buckets, evaluating all possibilities can be challenging. However, with a smaller number of groups, it is easy to thoroughly examine all potential events of interest. The findings in the example indicate that the optimal number of buckets consists of two buckets, $s = 2$, with one group in the first bucket and three groups in the second bucket, resulting in the highest NPI lower and upper probabilities.

The bootstrap method is used, which provides an estimated probability for the general event. Additionally, NPI lower and upper probabilities are derived for this event. It can be expected that the estimated probability lies between the NPI lower and upper probabilities, which was indeed confirmed in all the examples provided

in this chapter.

Finally, in this chapter, two different loss functions are applied to measure the penalty associated with the ranking of suboptimal s buckets in terms of one or more future observations within each bucket and to see if these two loss functions lead to the same conclusion. Additionally, it is examined whether the results obtained based on the lower and upper bounds for the NPI lower and upper linear and quadratic expected loss functions align with the results obtained based on the NPI lower and upper probabilities. The findings indicate that the results based on expected loss and probabilities are well aligned, leading both to the same conclusion. Furthermore, it was observed that the results from the lower and upper bounds for the NPI lower and upper expected linear and quadratic losses are consistent with each other.

Chapter 6

Conclusions

This chapter offers a concise summary of the main findings of this thesis and high-lights challenges for future research. The thesis applied Nonparametric Predictive Inference (NPI) to various selection and ranking problems based on a single future observation from each group. In Chapter 2, several important findings were presented on using NPI for selecting and ranking real-valued future observations for multiple events of interest. The study derived exact NPI lower and upper probabilities for a scenario, which is the first scenario, involving selecting a ranked subset of the two best groups, an optimisation process was performed, and the exact NPI lower and upper probabilities were derived. Determining the exact NPI lower and upper probabilities is straightforward in this case, as there is only one group in the middle. Even with a large and overlapping number of observations, obtaining exact NPI lower and upper probabilities remains possible. A simulation study was also conducted where the means were assumed and arranged from smallest to largest, with the worst groups having the smallest means and the best group having the largest mean. The NPI method performed well in this setup, and the simulation results led to the conclusion that selecting the group with the largest mean as the best is a well-considered decision.

In the second scenario, when the number of ranked groups within a subset is three or more, it is not possible to simultaneously optimise all groups. Therefore, an optimisation process and a heuristic algorithm were performed independently for each group in the middle, and then the approximate NPI lower and upper proba-

bilities were obtained. The lower and upper bounds for the NPI probabilities are derived because calculating the exact (or approximate) NPI lower and upper probabilities is computationally demanding. The results indicate that the approximate NPI lower and upper probabilities are nearly equal to their respective bounds.

As a result, ways to improve the NPI method for selecting a ranked subset of two groups, and to reduce the computational effort to calculate the exact lower (upper) NPI probabilities were identified. The first (last) sub-interval in the first interval often yields the minimum (maximum) NPI lower (upper) probability, while the last (first) sub-interval in the last interval does the same. For intervals in between, the minimum (maximum) number of assigned probability masses comes from the first (last) sub-interval of each interval.

Additionally, in Chapter 2, the selection of a partially ranked subset is introduced, offering a simpler and more practical approach for comparison, especially when a fully ranked subset of the best groups is either challenging or unnecessary.

Finally, a comparison was made between the proposed method and Bechhofer's and Gupta's methods from the literature. It was found that the indifference zone approach introduced by Bechhofer and the NPI method for selecting a ranked subset of the two best populations yield both the same selected subset and the same ranking within the subset. However, Gupta's method does not always yields the same selected subset as Bechhofer's method or the NPI method.

Chapter 3 explored how to quantify the loss of making incorrect selections by investigating pairwise and multiple comparisons using three types of loss functions: zero-one, linear, and quadratic loss functions. Each loss function offered a clear understanding of the evaluation of the selection process. The zero-one loss function, simple to calculate, indicated whether a selection was correct or not, directly correlating with imprecise probabilities. The linear loss function provided more detailed information by measuring the deviation from the best group. The quadratic loss function further detailed the assessment by penalising larger errors more heavily.

Through various examples, a clear relationship was observed between the conclusions drawn from the zero-one, linear, and quadratic loss functions in pairwise comparisons. In the case of multiple comparisons, the examples used both real-

valued and simulated data, offering a thorough examination of how the method performs in different scenarios. The findings across these data types reveal that, while the zero-one, linear, and quadratic loss functions generally yield consistent results, slight variations in selection can occur depending on the specific loss function applied.

In Chapter 4, NPI for ranking a subset of the best w groups using zero-one and general multi-level loss functions was introduced, focusing on the $w = 2$ and $w = 3$ cases. The zero-one loss function provides a basic binary evaluation of whether the ranked subset of the best w groups is correct, while the general multi-level loss function offers a more nuanced approach by assigning different penalties for various incorrect ranked groups within the subset. This approach is useful when certain ranked groups within the subset have more consequences. Through real and simulated data examples, the NPI lower and upper expected zero-one and general multi-level losses were calculated to quantify the loss across different scenarios involving a ranked groups within the subset.

For $w = 2$, the multi-level loss function is introduced as a more advanced method to evaluate the loss of making a non-optimal ranking of the groups within the subset, penalties range from L_0 to L_5. An optimisation process was performed to derive the NPI lower and upper expected multi-level losses, aiding in decision-making regarding subset group rankings.

A ranked subset of the best w groups using general multi-level loss function is presented, and a special case for the ranked subset of the three best groups is derived. This method is necessary because NPI lower and upper expected losses cannot be directly calculated when more than one group exists between the non-selected groups and the largest selected group. In this case, an optimisation process is required to find the approximate NPI lower and upper expected multi-level losses.

In some cases, the loss values are set by the problem instead of the data. This helps decision-makers evaluate the consequences and make informed decisions. The effect of different loss values on various scenarios involving a ranked subset of the best groups is investigated. The analysis shows differences in conclusions across various scenarios of subset group ranking.

In Chapter 5, the focus is on the ranking of the buckets, without considering the order of the groups within the buckets. Each group is placed in exactly one bucket, each bucket containing at least one group. A predetermined number of buckets s and a specific number of groups in each bucket are assumed. In addition, using NPI-Bootstrap method along with the lower and upper bounds for the NPI lower and upper probabilities.

Two key questions were addressed: how should the groups be assigned to the buckets? And how many buckets should be used? Two real-world examples were provided to illustrate the method and draw conclusions. To answer them, all potential allocation methods were considered and selected the one that aligned with specific criteria, such as maximising or minimising the probability of a particular event. Although evaluating all possibilities is challenging with a large number of groups and buckets, it becomes manageable when dealing with fewer groups. Another approach was to use statistics from the data to rank the groups, such as placing the groups with the smallest medians in the first bucket, the second smallest in the second bucket, and so on. However, it was found that this method did not always yield the highest lower and upper bounds for the NPI lower and upper probabilities.

The NPI-bootstrap method was used to provide an estimated probability for the general event of selection and ranking, and it was found that the estimated probability consistently fall between the lower and upper bounds for the NPI lower and upper probabilities, as confirmed in all the examples presented in Chapter 5.

Finally, two different loss functions were applied to quantify the loss from making non-optimal selection and ranking. Also, a comparison between the results based on the NPI lower and upper linear and quadratic expected losses and those based on the lower and upper bounds for the NPI lower and upper probabilities. The findings indicated that the results of both the expected loss functions and the bounds of the probabilities were in agreement, leading to the same conclusion. Furthermore, the NPI lower and upper expected linear and quadratic losses produced consistent results.

To provide further insight into the proposed NPI method for selection and ranking, simulations of other distributions should be considered. However, in this thesis,

only data simulated from the Normal distribution were used.

For future research, it will be of interest to apply different loss functions to the ranking of future observations from subsets of various sizes. Additionally, ranking future observations from different data types, including ordinal, bivariate data, right-censored data, or lifetime data, remains interesting for exploration. In addition, an important topic for future research is to extend the use of NPI in ranking and selection from one future observation to m future observations and to measure the loss of making incorrect selections or ranking of groups.

Bibliography

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

[2] Baker, R. *Multinomial Nonparametric predictive inference: Selection, Classification and Subcategory Data.* PhD thesis, Durham University. UK, 2010. `http://etheses.dur.ac.uk/257/`.

[3] Bechhofer, R.E. A single-sample multiple decision procedure for ranking means of Normal populations with known variances. *The Annals of Mathematical Statistics*, 25: 16–39, 1954.

[4] Bechhofer, R.E. and Kiefer, J. and Sobel, M. *Sequential Identification and Ranking Procedures.* University of Chicago Press, Chicago, 1968.

[5] Berger, J.O. *Statistical Decision Theory and Bayesian Analysis.* Springer, New York, 1985.

[6] Berger, R.L. and Gupta, S.S. Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. *Scandinavian Journal of Statistics*, 7: 21–26, 1980.

[7] Boole, G. *An Investigation of the Laws of Thought: On which are Founded the Mathematical Theories of Logic and Probabilities.* Walton and Maberley, London, 1854.

[8] Carroll, R.J. and Gupta, S.S. On the probabilities of rankings of k populations with applications. *Journal of Statistical Computation and Simulation*, 5: 145–157, 1977.

[9] Coolen, F.P.A. Comparing two populations based on low stochastic structure assumptions. *Statistics and Probability Letters*, 29: 297–305, 1996.

[10] Coolen, F.P.A. Low structure imprecise predictive inference for Bayes' problem. *Statistics and Probability Letters*, 36: 349–357, 1998.

[11] Coolen, F.P.A. On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International*, 20: 193–202, 2004.

[12] Coolen, F.P.A. On Nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, 15: 21–47, 2006.

[13] Coolen, F.P.A. and Augustin, T. Learning from multinomial data: a nonparametric predictive alternative to the Imprecise Dirichlet Model. In *International Symposium on Imprecise Probabilities and Their Applications*, volume 5, pages 125–134, 2005.

[14] Coolen, F.P.A. and BinHimd, S. Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov-Smirnov test. *Journal of Statistical Theory and Practice*, 14, 2020.

[15] Coolen, F.P.A. and Coolen-Maturi, T. and Alqifari, H.N. Nonparametric predictive inference for future order statistics. *Communications in Statistics: Theory and methods*, 47(10): 2527–2548, 2018. ISSN 0361-0926.

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

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

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

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

[20] Coolen-Maturi, T. The ordering of future observations from multiple groups. *Communications in Statistics: Simulation and Computation*, 51: 7526–7543, 2022.

[21] Cramér, H. *On the Mathematical Theory of Risk*. Centraltryckeriet, Stockholm, 1930.

[22] De Finetti, B. *Theory of Probability*. Wiley, London, 1974.

[23] Deely, J.J. and Gupta, S.S. On the properties of subset selection procedures. *Sankhyā: The Indian Journal of Statistics, Series A*, 30: 37–50, 1968.

[24] Desu, M.M. and Sobel, M. A fixed subset-size approach to the selection problem. *Biometrika*, 55: 401–410, 1968.

[25] Desu, M.M. and Sobel, M. Nonparametric procedures for selecting fixed-size subsets. In S.S. Gupta and J. Yackel, editors, *Statistical Decision Theory and Related Topics*, pages 359–376. Academic Press, New York, 1971.

[26] Dobson, A.J. *Introduction to Statistical Modelling*. Springer, New York, NY, 2013.

[27] Dudewicz, E.J. and Dalal, S.R. Allocation of observations in ranking and selection with unequal variances. *Sankhyā: The Indian Journal of Statistics, Series B*, 37: 28–78, 1975.

[28] Ferguson, T.S. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.

[29] Finner, H. and Giani, G. Duality between multiple testing and selecting. *Journal of Statistical Planning and Inference*, 54: 201–227, 1996.

[30] Freeman, H. and Kuzmack, A. and Maurice, R. Multivariate t and the ranking problem. *Biometrika*, 54: 305–308, 1967.

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

[32] Girshick, M. A. and Savage, L. J. Quadratic loss functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 53–73, Berkeley, CA, 1950. University of California Press.

[33] Goel, P.K. and Rubin, H. On selecting a subset containing the best population-a Bayesian approach. *The Annals of Statistics*, 5: 969–983, 1977.

[34] Good, I.J. *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press, Minneapolis, MN, 1983.

[35] Govindarajulu, Z. and Harvey, C. Bayesian procedures for ranking and selection problems. *Annals of the Institute of Statistical Mathematics*, 26: 35–53, 1974.

[36] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, Reading, Massachusetts, 1994.

[37] Gupta, S.S. *On a Decision Rule for a Problem in Ranking Means*. PhD thesis, University of North Carolina, USA, 1956.

[38] Gupta, S.S. On some multiple decision (selection and ranking) rules. *Technometrics*, 7: 225–245, 1965.

[39] Gupta, S.S. and Huyett, M.J. and Sobel, M. Selection and ranking with Binomial populations. *Transactions of American Society for Quality Control*, 9: 635–644, 1957.

[40] Gupta, S.S. and Miescke, K.J. On the performance of subset selection rules under normality. *Journal of Statistical Planning and Inference*, 103: 101–115, 2002.

[41] Gupta, S.S. and Panchapakesan, S. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Society for Industrial and Applied Mathematics, Philadelphia, 2002.

[42] Gupta, S.S. and Sobel, M. On a statistic which arises in selection and ranking problems. *The Annals of Mathematical Statistics*, 28: 957–967, 1957.

[43] Gupta, S.S. and Sobel, M. Selecting a subset containing the best of several binomial populations. In I. Olkin, H. Hotelling, and J. Wolfowitz, editors, *Contributions to Probability and Statistics*, pages 389–413. Stanford University Press, Stanford, CA, 1960.

[44] Hansen, A.P. Abnormal serum growth hormone response to exercise in maturity-onset diabetics. *Diabetes*, 22: 619–628, 1973.

[45] Hashemi, S.J. and Ahmed, S. and Khan, F. Loss functions and their applications in process safety assessment. *Process Safety Progress*, 33: 285–291, 2014.

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

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

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

[49] Hong, Y. and Lee, T.H. Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics*, 85: 1048–1062, 2003.

[50] Hsu, J.C. Simultaneous confidence intervals for all distances from the best. *The Annals of Statistics*, 9: 1026–1034, 1981.

[51] Hsu, J.C. Simultaneous inference with respect to the best treatment in block designs. *Journal of the American Statistical Association*, 77: 461–467, 1982.

[52] Hsu, J.C. *Multiple Comparisons: Theory and Methods.* Chapman and Hall, London, 1996.

[53] Kumar, N. A general class of subset selection procedures. *Journal of Combinatorics, Information and System Sciences*, 36: 135–158, 2011.

[54] Kumar, N. Nonparametric subset selection procedure for location parameters. *Model Assisted Statistics and Applications*, 6: 337–347, 2011.

[55] Lehmann, E.L. A theory of some multiple decision problems. *The Annals of Mathematical Statistics*, 28: 547–572, 1957.

[56] Lehmann, E.L. Some model problems of selection. *The Annals of Mathematical Statistics*, 32: 990–1012, 1961.

[57] Lehmann, E.L. and Romano, J.P. and Casella, G. *Testing Statistical Hypotheses.* Springer, New York, 1986.

[58] Liese, F. and Miescke, K.J. *Statistical Decision Theory. Estimation, Testing and Selection.* Springer, New York, 2008.

[59] Mahamunulu, D.M. Some fixed-sample ranking and selection problems. *The Annals of Mathematical Statistics*, 38: 1079–1091, 1967.

[60] Maturi, T. *Nonparametric predictive inference for Multiple Comparisons*. PhD thesis, Durham University, UK, 2010. `http://etheses.dur.ac.uk/230/`.

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

[62] Neuhauser, M. and Liu, P.Y. and Hothorn, L. A. Nonparametric tests for trend: Jonckheere's test, a modification and a maximum test. *Biometrical Journal*, 40: 899–909, 1998.

[63] Paulson, E. A multiple decision procedure for certain problems in the analysis of variance. *The Annals of Mathematical Statistics*, 20: 95–98, 1949.

[64] Rizvi, M.H. and Sobel, M. Nonparametric procedures for selecting a subset containing the population with the largest q-quantile. *The Annals of Mathematical Statistics*, 38: 1788–1803, 1967.

[65] Savage, L.J. *The Subjective Basis of Statistical Practice*. University of Michigan, Ann Arbor, MI, 1961.

[66] Schafer, R.E. A single-sample complete ordering procedure for certain populations. In F. Proschan and R. J. Serfling, editors, *Reliability and Biometry: Statistical Analysis of Lifelength*, pages 413–427. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1974.

[67] Shafer, G. Constructive decision theory. *International Journal of Approximate Reasoning*, 79: 45–62, 2016.

[68] Simpson, D.G. and Margolin, B.H. Recursive nonparametric testing for dose-response relaflonships subject to downturns at high doses. *Biometrika*, 73: 589–596, 1986.

[69] Sobel, M. Nonparametric procedures for selecting the t population with the largest q-quantiles. *The Annals of Mathematical Statistics*, 38: 1804–1816, 1967.

[70] Sobel, M. Selecting a subset containing at least one of the t best populations. In P. R. Krishnaiah, editor, *Multivariate Analysis II*, pages 411–428. Academic Press, New York, 1969.

[71] Tong, Y.L. On partitioning a set of Normal populations by their locations with respect to a control using a single-stage, a two-stage and a sequential procedure. *The Annals of Mathematical Statistics*, 40: 1300–1324, 1969.

[72] Wald, A. Statistical decision functions. *The Annals of Mathematical Statistics*, 20: 165–205, 1949.

[73] Wald, A. *Statistical Decision Theory*. Wiley, New York, 1950.

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

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

[76] Weichselberger, K. *Elementare Grundbegriffe einer Allgemeineren Wahrschein-lichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept.* Physika, Heidelberg, 2001.