Nonparametric Predictive Inference for Ranking and Selection using Loss Functions

Fatimah I. Alshihry[a], Frank P. A. Coolen[b], Tahani Maturi[b]

[a]King Abdulaziz University, Jeddah, Saudi Arabia

[b]Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, UK

Abstract

Selection and ranking of multiple groups are important decision problems where choices must be made about uncertain future outcomes. This paper develops a Nonparametric Predictive Inference (NPI) framework for ranking and selection using loss functions to quantify the consequences of suboptimal decisions. Uncertainty is quantified through NPI lower and upper expected losses, which capture the imprecision inherent in predictive inference under minimal modelling assumptions. For selection problems, zero–one, linear, and quadratic loss functions are applied in multiple comparison settings, including events involving the selection of the best group, a subset of the best groups, and a subset containing the best group. For ranking problems, a general multi-level loss function is introduced for ranked subsets of the best groups, assigning different penalties to different types of ranking errors and allowing a more flexible assessment of ranking quality. The framework is further extended to general selection-and-ranking problems in which groups are allocated to ordered buckets. Examples demonstrate that different loss specifications can lead to different assessments of competing decisions, while linear, quadratic, and multi-level losses provide additional information about the magnitude and severity of potential errors beyond binary evaluation. The results highlight the importance of carefully specifying loss functions and demonstrate the usefulness of NPI lower and upper expected losses for decision-making in ranking and selection problems under uncertainty.

Keywords: Nonparametric predictive inference (NPI); NPI-Bootstrap; Selection; Ranking; Imprecise probabilities; Loss functions.

1 Introduction

Ranking and selection problems arise in many areas of statistics, where the goal is to identify the best group or a subset of the best groups based on observed data. In many practical settings the interest is not only in selecting the best group but also in comparing several groups, ranking them, or selecting a subset that is likely to contain the best one. Decision making in such settings involves uncertainty about future observations, and therefore requires methods that can quantify the consequences of selecting suboptimal groups.

Nonparametric Predictive Inference (NPI) [2,9] provides a framework for predictive inference based on minimal modelling assumptions. NPI focuses on predicting future observations using observed data, typically resulting in lower and upper probabilities that reflect the imprecision arising from limited information. Previous work on NPI has addressed several ranking and selection problems, including pairwise comparisons [7] and subset selection [12], using NPI lower and upper probabilities to assess the likelihood of events of interest.

While probability-based comparisons provide useful information about the likelihood of different events, decision making often requires explicit consideration of the consequences of incorrect selections. Loss functions provide a natural way to quantify such consequences by assigning penalties to undesirable outcomes. A loss function measures the penalty incurred when a decision deviates from the desired outcome, and decisions can then be evaluated through their expected loss. Different loss functions emphasise different aspects of decision quality; for example, the zero–one loss distinguishes only between correct and incorrect decisions, whereas linear and quadratic losses quantify the magnitude of deviations [15,17,20].

This paper develops an NPI framework for ranking and selection based on loss functions. The aim is to evaluate decisions by means of NPI lower and upper expected losses, thereby extending probability-based NPI analysis to a decision-theoretic setting. Several loss functions are considered, including zero–one, linear, quadratic, and multi-level loss functions. These loss functions allow different types and severities of ranking errors to be incorporated into the decision process.

The methodology is first applied to multiple comparison problems involving the selection of the best group, the selection of a subset of the best groups, and the selection of a subset that includes the best group. The paper then considers the problem of selecting and ranking a subset of the best groups using a general multi-level loss function, where different penalties can be assigned to different types of ranking errors. In addition, a general selection-and-ranking framework is introduced in which groups are allocated to ordered buckets and the ranking of these buckets is evaluated using NPI lower and upper expected losses.

Through several examples based on real and simulated data, the proposed approach illustrates how loss functions can provide additional insight into ranking and selection decisions within the NPI framework. In particular, the analysis highlights the relationship between probability-based assessments and loss-based decision criteria, and shows how different loss functions influence the evaluation of competing ranking and selection strategies.

2 Loss Functions in Selection

In selection problems, a loss function quantifies the penalty incurred when suboptimal groups are selected. The aim is to choose a decision rule that minimises the expected loss associated with the selection. This section reviews loss functions commonly used in subset selection, providing background for the NPI-based developments introduced later.

Several methodologies in the literature have incorporated loss-based criteria in subset selection. Gupta and Panchapakesan [20] proposed a subset selection rule evaluated through loss functions, with emphasis on minimising incorrect selections. Berger and Gupta [5] investigated a minimax approach that minimises the maximum risk, equivalently maximising the minimum probability that the selected subset contains the best population. Hsu [24–26] developed procedures based on simultaneous confidence intervals for comparing each population with the best, establishing connections between multiple comparisons and subset selection. Finner and Giani [16] extended these ideas to more complex experimental designs. These approaches differ in formulation but emphasise the central role of loss-based evaluation in selection problems.

Various loss functions can be employed in the selection process, including zero–one, linear, and quadratic loss functions. The choice of loss function depends on the structure of the decision problem and the type of penalty one wishes to impose. We now introduce the general framework and summarise several commonly used loss functions.

Consider k populations characterised by one-dimensional parameters $\mu_1, \ldots, \mu_k$. Let $\mu^* = \{\mu_{[1]}, \ldots, \mu_{[k]}\}$ denote the ordered (increasing) rearrangement of $\mu_1, \ldots, \mu_k$, where $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$ and $\mu_{[k]}$ is the largest value; the brackets [ ] indicate rank order rather than original indices. Let $S \subseteq \{1, \ldots, k\}$ denote the selected subset. The zero–one loss introduced by Deely and Gupta [15] for selecting a subset that contains the best population is defined as

$$L_{\{0,1\}}(S, \mu^*) = I\{\mu_{[k]} \notin S\},$$

where $I\{\cdot\}$ denotes the indicator function. This loss assigns a unit penalty whenever the selected subset S fails to include the population with the largest parameter value $\mu_{[k]}$.

To quantify the magnitude of suboptimality, linear and quadratic loss functions are frequently used. The linear loss function [15] is given by

$$L_l(S, \mu^*) = \sum_{t \in S} c_t(\mu_{[k]} - \mu_{[t]}),$$

where $c_t > 0$ are penalty weights. This loss increases proportionally to the difference between the largest parameter and the parameters of the selected populations. The quadratic loss function [17] extends this idea by squaring the deviation,

$$L_q(S, \mu^*) = \sum_{t \in S} c_t(\mu_{[k]} - \mu_{[t]})^2,$$

thereby penalising larger deviations more heavily. Linear and quadratic losses often lead to similar conclusions, although the quadratic loss emphasises substantial discrepancies [4].

These loss functions emphasise different aspects of a selection decision. The zero–one loss considers only whether the desired selection event occurs, whereas the linear and quadratic losses additionally quantify the magnitude of deviation from the desired outcome, with the quadratic loss placing greater weight on larger deviations. The choice of loss function therefore depends on the decision context and on how different types of selection errors are to be penalised. A detailed discussion of the properties and broader relationships between these loss functions is beyond the scope of this paper; further details can be found in the cited literature.

Other formulations have also been proposed, such as the average loss function [15], defined as

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

measures the mean deviation within the selected subset. Goel and Rubin [18] introduced a combined linear loss incorporating both subset size and deviation from the best population,

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

capturing the trade-off between selecting larger subsets and ensuring inclusion of a population close to the best. Gupta and Miescke [19] proposed loss functions that reward either inclusion of all best populations or inclusion of at least one best population, allowing different treatments of ties. Lehmann [28,29] introduced an additive loss that separately penalises incorrect inclusion and incorrect exclusion of populations.

Loss functions evaluate the consequence of a decision for a fixed parameter configuration. The associated risk function, or expected loss [35], provides a measure of performance under uncertainty. Let $L(\mu^*, d)$ denote

the loss incurred by decision rule d when the true parameter vector is μ^*. The expected loss is

$$E\left[L(\mu^*, d(X))\right] = \int L(\mu^*, d(x))\, dP(x \mid \mu^*),$$

where $P(\cdot \mid \mu^*)$ denotes the probability distribution of the data X given μ^*. Minimising the expected loss provides a principled basis for evaluating competing decisions.

This formulation provides a theoretical basis for such evaluations. However, in practice, the true parameter vector μ^* and the corresponding distribution $P(\cdot \mid \mu^*)$ are unknown. NPI avoids this by focusing on predictive inference based on observable data and future observations.

3 Nonparametric Predictive Inference

Nonparametric Predictive Inference (NPI) [2,9] is a statistical framework based on Hill's assumption $A_{(n)}$ [21,22]. It provides predictive inferences for a future observation given n related observations without imposing parametric distributional assumptions, making it particularly suitable when limited prior knowledge is available or when additional modelling assumptions are undesirable.

To introduce assumption $A_{(n)}$, let $X_1, \ldots, X_n, X_{n+1}$ be continuous and exchangeable random quantities. Suppose the ordered observed values of $X_1, \ldots, X_n$ are $x_1 < \cdots < x_n$, and define $x_0 = -\infty$ and $x_{n+1} = \infty$. The observations partition the real line into $n+1$ intervals $I_j = (x_{j-1}, x_j)$ for $j = 1, \ldots, n+1$. Assumption $A_{(n)}$ assigns equal probability mass to each interval,

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

Throughout this paper, we denote a single future observation from a group by X^f. Thus, in the above formulation, X^f corresponds to X_{n+1}. Ties are assumed not to occur; if present, probability mass may be assigned to closed intervals or tied observations may be treated as differing by an arbitrarily small amount [23].

Assumption $A_{(n)}$ is the fundamental assumption underlying NPI [6]. It is a post-data assumption related to exchangeability and does not impose any further assumptions on the underlying distribution. In particular, $A_{(n)}$ assigns equal probabilities to the intervals between ordered observations for a future observable quantity, thereby providing a direct predictive probability based only on the available data [6]. Inferences based on $A_{(n)}$ are frequentist, predictive, and nonparametric, and the resulting predictive inference is *exactly calibrated* for every sample size n, which is a strong consistency property from a frequentist perspective [6,27]. Furthermore, NPI never yields conclusions that are in conflict with inference based on empirical probabilities [6].

The assumption $A_{(n)}$ does not generally provide precise probabilities for many events of interest. Instead, optimal bounds for probabilities can be derived using de Finetti's fundamental theorem of probability [14], leading to lower and upper probabilities within the framework of imprecise probability [6]. For an event A, these are denoted by $\underline{P}(A)$ and $\overline{P}(A)$ and satisfy $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ and $\overline{P}(A) = 1 - \underline{P}(A^c)$. The lower probability can be interpreted as a maximal lower bound for the precise probability of A, while the upper probability is a minimal upper bound [9]. Hence, uncertainty is quantified by the interval $[\underline{P}(A), \overline{P}(A)]$, with the difference $\overline{P}(A) - \underline{P}(A)$ measuring the degree of imprecision [8].

NPI has been extensively developed for selection and ranking problems. Coolen [7] introduced NPI for pairwise comparison of real-valued observations based on one future observation per group. Coolen [12] extended this to multiple comparisons for $k \geq 2$ independent groups, focusing on three events: selection of the best group, selection of a subset of best groups, and selection of a subset that contains the best group. Subsequent work extended NPI to Bernoulli trials [11], multinomial data [3], right-censored data [30], tail-terminated data [31], and order statistics [10]. Ranking of multiple groups has also been investigated within NPI framework [13]. Recent work has further developed NPI methods for selection and ranking based on single future observations per group, including exact and approximate lower and upper probabilities for identifying and ordering the best groups, as well as extensions to bucket-based ranking structures and data-driven allocation procedures [1].

To incorporate decision-making, NPI is combined with loss functions through lower and upper expected losses. Let $L(X^f, d) \in \mathbb{R}$ denote the loss associated with decision d and future observation X^f. Under assumption $A_{(n)}$, each interval I_j carries probability mass $1/(n+1)$, while the exact location of X^f within an interval is unknown. The NPI lower and upper expected losses are therefore defined by taking the infimum and supremum of the loss within each interval:

$$\underline{E}(L(X^f, d)) = \sum_{j=1}^{n+1} \frac{1}{n+1} \inf_{x \in I_j} L(x, d), \tag{2}$$

$$\overline{E}(L(X^f, d)) = \sum_{j=1}^{n+1} \frac{1}{n+1} \sup_{x \in I_j} L(x, d). \tag{3}$$

These quantities provide bounds on the expected loss consistent with assumption $A_{(n)}$, thereby enabling decision-making under predictive imprecision. The NPI lower and upper expected losses are applied to pairwise and multiple comparison problems in the subsequent sections.

In this setting, the decision d specifies the selection rule, while $X^f = (X_1^f, \ldots, X_k^f)$ denotes the vector of future observations to which the decision is applied. Thus, $L(X^f, d)$ represents the loss associated with applying decision d to the future observations, and $L(x, d)$ denotes the corresponding loss for a particular

realisation x of X^f. The loss functions $L(S, \mu^*)$ introduced in Section 2 provide the general decision-theoretic formulation for selection, where the loss is expressed in terms of the unknown population parameters. In contrast, the NPI approach is parameter-free and predictive, with the specific loss functions in the following sections formulated directly in terms of future observations. The corresponding lower and upper expected losses are then obtained using the NPI framework described above.

4 NPI for selection using loss functions

This section applies the NPI decision-theoretic framework developed above to selection problems involving $k \geq 2$ independent groups. The framework is presented for three selection events: selecting the best group, selecting a subset of the best groups, and selecting a subset that includes the best group. Each event is examined using zero–one, linear, and quadratic loss functions to quantify the consequences of selecting suboptimal group(s) and to investigate how the choice of loss function influences the resulting selection decisions.

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 $i_j = 1, 2, \ldots, n_j + 1$. The ordered observed values for group j are $x_{j,1} < x_{j,2} < \ldots < x_{j,n_j}$, with $x_{j,0} = l_j$ and $x_{j,n_j+1} = r_j$ defined for ease of notation. Let X_j^f denote a single future observation from group j. Inference is based on Hill's assumption $A_{(n_j)}$ for each group. In the following subsections, the three selection events are examined using zero–one, linear, and quadratic loss functions.

4.1 Zero-one loss function

First, consider selecting the best group. For a fixed group ℓ, the event of interest is that X_ℓ^f is at least as large as all other future observations, that is, $X_\ell^f \geq \max_{j \neq \ell} X_j^f$. The zero–one loss function for this event is defined as

$$L_{\{0,1\}}(X_\ell^f, \max_{j \neq \ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j \neq \ell} X_j^f, \\ 1 & \text{if } X_\ell^f < \max_{j \neq \ell} X_j^f. \end{cases} \tag{4}$$

This loss assigns zero when the correct group is selected, that is, when X_ℓ^f is the largest future observation, and one otherwise.

Since the zero–one loss is the indicator of an incorrect selection, the NPI lower (upper) expected zero–one loss equals the NPI lower (upper) probability of the complementary event [12]. Hence, the NPI lower and

upper expected zero–one losses for $X_\ell^f \geq \max_{j\neq\ell} X_j^f$, respectively, are

$$\underline{E}(L_{\{0,1\}}(X_\ell^f, \max_{j\neq\ell} X_j^f)) = \underline{P}(X_\ell^f < \max_{j\neq\ell} X_j^f) = 1 - \overline{P}(X_\ell^f \geq \max_{j\neq\ell} X_j^f), \tag{5}$$

$$\overline{E}(L_{\{0,1\}}(X_\ell^f, \max_{j\neq\ell} X_j^f)) = \overline{P}(X_\ell^f < \max_{j\neq\ell} X_j^f) = 1 - \underline{P}(X_\ell^f \geq \max_{j\neq\ell} X_j^f). \tag{6}$$

Second, consider selection of a subset of the best groups. Let $S = \{\ell_1, \ldots, \ell_w\} \subset \{1, \ldots, k\}$ denote a subset of size w, where $1 \leq w \leq k-1$, and let $NS = \{1, \ldots, k\} \setminus S$ denote the non-selected groups. The event of interest is that all future observations from S exceed all future observations from NS, that is, $\min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$. The corresponding zero–one loss function is

$$L_{\{0,1\}}(\min_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f) = \begin{cases} 0 & \text{if } \min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f, \\ 1 & \text{if } \min_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f. \end{cases} \tag{7}$$

Again, the NPI lower and upper expected zero–one losses coincide with the NPI lower and upper probabilities of the complementary event [12]. Therefore,

$$\underline{E}(L_{\{0,1\}}(\min_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f)) = \underline{P}(\min_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f), \tag{8}$$

$$\overline{E}(L_{\{0,1\}}(\min_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f)) = \overline{P}(\min_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f). \tag{9}$$

Finally, consider selection of a subset that includes the best group. In this case, at least one future observation from S exceeds all future observations from NS, that is, $\max_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$. The zero–one loss function is

$$L_{\{0,1\}}(\max_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f) = \begin{cases} 0 & \text{if } \max_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f, \\ 1 & \text{if } \max_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f. \end{cases} \tag{10}$$

Consequently, the NPI lower and upper expected zero–one losses are

$$\underline{E}(L_{\{0,1\}}(\max_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f)) = \underline{P}(\max_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f), \tag{11}$$

$$\overline{E}(L_{\{0,1\}}(\max_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f)) = \overline{P}(\max_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f). \tag{12}$$

Similarly, these NPI lower and upper expected zero-one losses coincide with the NPI lower and upper probabilities of the complementary event [12].

4.2 Linear loss function

The same three selection events considered under the zero–one loss function are now evaluated using the linear loss function. In this case, the penalty reflects the magnitude of the deviation between the selected group(s) and the competing group(s) whenever an incorrect selection occurs.

The first scenario again involves selecting the best group. The linear loss for the event $X_\ell^f \geq \max_{j\neq\ell} X_j^f$ is defined as

$$L_l(X_\ell^f, \max_{j\neq\ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j\neq\ell} X_j^f, \\ X_\ell^f - \max_{j\neq\ell} X_j^f & \text{if } X_\ell^f < \max_{j\neq\ell} X_j^f. \end{cases} \tag{13}$$

This loss quantifies the deviation when X_ℓ^f fails to be the largest future observation, measuring the difference between X_ℓ^f and $\max_{j\neq\ell} X_j^f$.

To derive the NPI lower expected linear loss, the probability mass for group ℓ is assigned to the right endpoint per interval, while the probability mass for each group $j \neq \ell$ is assigned to the left endpoint per interval. For each configuration of interval indices, the lower expected loss is then computed via $L_l(x_{\ell,i_\ell}, \max_{j\neq\ell} x_{j,i_j-1})$. 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, yielding $L_l(x_{\ell,i_\ell-1}, \max_{j\neq\ell} x_{j,i_j})$. Therefore, the NPI lower and upper expected linear losses are

$$\underline{E}(L_l(X_\ell^f, \max_{j\neq\ell} X_j^f)) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j\in NS}}^{n_j+1} L_l(x_{\ell,i_\ell}, \max_{j\neq\ell} x_{j,i_j-1}) \tag{14}$$

$$\overline{E}(L_l(X_\ell^f, \max_{j\neq\ell} X_j^f)) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{i_\ell=1}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j\in NS}}^{n_j+1} L_l(x_{\ell,i_\ell-1}, \max_{j\neq\ell} x_{j,i_j}) \tag{15}$$

Here, $\sum_{\substack{i_j=1 \\ j\in NS}}^{n_j+1}$ represents multiple summations over all groups $j = 1, 2, \ldots, k$ excluding ℓ, and over all indices $i_j = 1, \ldots, n_j + 1$.

The second scenario involves selecting a subset of the best groups, with event $\min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$. The corresponding linear loss function is

$$L_l(\min_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f) = \begin{cases} 0 & \text{if } \min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f, \\ \min_{\ell\in S} X_\ell^f - \max_{j\in NS} X_j^f & \text{if } \min_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f. \end{cases} \tag{16}$$

To derive the NPI lower (upper) expected linear loss, probability mass for groups $\ell \in S$ is assigned to the right (left) endpoints per interval, while probability mass for groups $j \in NS$ is assigned to the left (right) endpoints per interval. The NPI lower and upper expected linear losses are then

$$\underline{E}(L_l(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\min_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}) \tag{17}$$

$$\overline{E}(L_l(\min_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\min_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}) \tag{18}$$

where $\sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1}$ represents multiple summations over the w selected groups and indices $i_\ell = 1, \ldots, n_\ell + 1$, with $\ell \in S$, and $\sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1}$ as defined above.

The third scenario considers selecting a subset that includes the best group, with event $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$. The corresponding linear loss is

$$L_l(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f) = \begin{cases} 0 & \text{if } \max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f, \\ \max_{\ell \in S} X_\ell^f - \max_{j \in NS} X_j^f & \text{if } \max_{\ell \in S} X_\ell^f < \max_{j \in NS} X_j^f. \end{cases} \tag{19}$$

With the same endpoint allocations, the NPI lower and upper expected linear losses are

$$\underline{E}(L_l(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\max_{\ell \in S} x_{\ell,i_\ell}, \max_{j \in NS} x_{j,i_j-1}) \tag{20}$$

$$\overline{E}(L_l(\max_{\ell \in S} X_\ell^f, \max_{j \in NS} X_j^f)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{\substack{i_\ell=1 \\ \ell \in S}}^{n_\ell+1} \sum_{\substack{i_j=1 \\ j \in NS}}^{n_j+1} L_l(\max_{\ell \in S} x_{\ell,i_\ell-1}, \max_{j \in NS} x_{j,i_j}) \tag{21}$$

The multiple summations are as defined above.

4.3 Quadratic loss function

The same events are now considered under quadratic loss, which penalises deviations by squaring the difference and therefore places greater weight on larger deviations than the linear loss function. For both linear and quadratic loss, future observations are assumed to be bounded by l_j and r_j, chosen below the minimum and above the maximum observed values, respectively.

For selecting the best group, the quadratic loss is

$$L_q(X_\ell^f, \max_{j\neq\ell} X_j^f) = \begin{cases} 0 & \text{if } X_\ell^f \geq \max_{j\neq\ell} X_j^f, \\ (X_\ell^f - \max_{j\neq\ell} X_j^f)^2 & \text{if } X_\ell^f < \max_{j\neq\ell} X_j^f. \end{cases} \tag{22}$$

For selecting a subset of the best groups with event $\min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$, the quadratic loss is

$$L_q(\min_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f) = \begin{cases} 0 & \text{if } \min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f, \\ (\min_{\ell\in S} X_\ell^f - \max_{j\in NS} X_j^f)^2 & \text{if } \min_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f. \end{cases} \tag{23}$$

For selecting a subset that includes the best group with event $\max_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$, the quadratic loss is

$$L_q(\max_{\ell\in S} X_\ell^f, \max_{j\in NS} X_j^f) = \begin{cases} 0 & \text{if } \max_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f, \\ (\max_{\ell\in S} X_\ell^f - \max_{j\in NS} X_j^f)^2 & \text{if } \max_{\ell\in S} X_\ell^f < \max_{j\in NS} X_j^f. \end{cases} \tag{24}$$

The NPI lower and upper expected quadratic losses for these events are derived in the same way as for the linear loss function, by assigning probability mass to the corresponding interval endpoints. Specifically, the expressions for the NPI lower and upper expected losses are obtained from those in the linear-loss case by replacing the loss function $L_l(\cdot)$ with $L_q(\cdot)$.

For linear and quadratic losses, bounded future observations are required to ensure that the corresponding expected losses are finite. In practice, the lower and upper bounds should be specified from genuine prior knowledge about the possible range of future observations and should not be determined from the observed data. It follows directly from the above expressions that, provided the lower and upper bounds are chosen below and above all observed values, respectively, the NPI lower expected linear and quadratic losses are invariant to the particular choice of bounds. This is because they depend only on the observed values $x_{j,i}$, $i = 1, \ldots, n_j$, for each group j. In contrast, the NPI upper expected losses explicitly involve the artificial boundary values $x_{j,0}$ and x_{j,n_j+1} and therefore depend on their specification. This reflects the imprecision inherent in NPI, where assumptions about possible extremes beyond the observed data affect the upper expectation only.

4.4 Strong and weak preference

This section introduces strong and weak preference criteria based on comparisons of NPI lower and upper expected losses across competing selections. A strong preference for event A over event B occurs when the NPI upper expected loss of A is smaller than the NPI lower expected loss of B, that is, $\overline{E}(L(A)) < \underline{E}(L(B))$.

If this condition does not hold, a weak preference may still be established when both the lower and upper expected losses of A are smaller than those of B, that is $\underline{E}(L(A)) < \underline{E}(L(B))$ and $\overline{E}(L(A)) < \overline{E}(L(B))$.

In this section, events A and B correspond to selecting the best group, selecting a subset of the best groups, or selecting a subset that includes the best group. By incorporating strong and weak preference criteria, we not only identify the preferred choice but also determine which groups or subsets are clearly inferior and which may still represent acceptable alternatives. Additionally, it can be useful to consider all events whose NPI lower and upper expected losses overlap with the lowest expected-loss interval among competing selections. This supports decision-making in which a non-optimal but still acceptable group or subset may be regarded as a reasonable alternative.

The following example quantifies the loss of incorrect NPI multiple comparisons using zero–one, linear, and quadratic loss functions, and compares whether these loss functions lead to consistent selection decisions.

Example 1. Simpson and Margolin [33] reported the Ames test data shown in Table 1, where six dose levels of the dye Acid Red 114 were tested using the Salmonella strain TA98. For each dose, the recorded values are replicate counts of mutation colonies; larger counts indicate stronger mutagenic activity.

The aim of this example is to illustrate NPI multiple comparisons based on loss functions by quantifying the penalty of selecting suboptimal dose levels using NPI lower and upper expected zero–one, linear, and quadratic losses. We consider three events of interest: selecting the best dose, $X_\ell^f \geq \max_{j\neq\ell} X_j^f$, selecting a subset of the best doses, $\min_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$, and selecting a subset that contains the best dose, $\max_{\ell\in S} X_\ell^f \geq \max_{j\in NS} X_j^f$, where S denotes the selected subset of doses and NS its complement. The dataset contains tied observations; to simplify the computation, we break ties by adding a small amount to tied values (here 0.1), which preserves all orderings within and across doses. To calculate linear and quadratic expected losses, we bound the future observations by $l = 7$ and $r = 99$, where l is smaller than the minimum observed value (8) and r is larger than the maximum observed value (98).

We first consider selecting the best dose. Table 2 reports the NPI lower and upper expected losses for each singleton subset $\{\ell\}$. Across all three loss functions, Dose 4 yields the smallest lower and upper expected losses (for zero–one loss, $\underline{E}(L_{\{0,1\}}) = 0.4304$ and $\overline{E}(L_{\{0,1\}}) = 0.7459$), indicating that it is the preferred choice under the "select-the-best" event. This is consistent with the data, as Dose 4 contains several of the largest observed counts (e.g. 82, 60.1, 59). At the other extreme, Doses 1 and 6 have the largest expected

Dose	Mutation colony counts								
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 1: Ames test mutation colony counts for six doses of Acid Red 114 (TA98). Ties have been broken by adding 0.1; these are therefore not the original reported values, Example 1.

	Zero–one loss		Linear loss		Quadratic loss	
$\{\ell\}$	$\underline{E}(L_{\{0,1\}})$	$\overline{E}(L_{\{0,1\}})$	$\underline{E}(L_l)$	$\overline{E}(L_l)$	$\underline{E}(L_q)$	$\overline{E}(L_q)$
$\{1\}$	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 2: NPI lower and upper expected losses for selecting the best dose, Example 1.

losses, reflecting their generally smaller observations (Dose 1 includes values such as 14 and 16, while Dose 6 includes 8 and 10). The linear and quadratic losses provide the same overall message but emphasise different aspects: linear loss penalises the magnitude of underperformance, while quadratic loss penalises large deviations more strongly, which leads to the heaviest penalties for clearly inferior choices such as Dose 6.

We next consider selecting a subset of the best doses under the stricter event $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, which requires every selected dose to outperform every non-selected dose in the next observation. The complete enumeration of all subsets and all loss functions produces long tables; therefore, we summarise the main findings here and refer the reader to [1] for the full results. Under zero–one loss, the subset $\{2, 3, 4, 5\}$ achieves the smallest NPI lower and upper expected losses among all subsets, with $\underline{E}(L_{\{0,1\}}) = 0.3948$ and $\overline{E}(L_{\{0,1\}}) = 0.7155$, and is therefore preferred. This choice is intuitively appealing because it excludes the clearly low-dose behaviour of Doses 1 and 6 while retaining the doses that contain the largest observations. A smaller subset such as $\{3, 4, 5\}$ provides a reasonable alternative with larger but still relatively low losses ($\underline{E}(L_{\{0,1\}}) = 0.5527$, $\overline{E}(L_{\{0,1\}}) = 0.8003$), illustrating how the interval $[\underline{E}, \overline{E}]$ can guide a trade-off between including more doses (to increase the chance that all selected doses dominate) and keeping the subset parsimonious.

For linear and quadratic loss, the same qualitative pattern emerges for this event: subsets containing Doses 4 and 3 tend to have the smallest lower and upper expected losses, whereas subsets that include Dose 6 and/or Dose 1 tend to have substantially larger losses. In particular, the subset $\{2, 3, 4, 5\}$ is again among

the best-performing subsets for linear loss, with $\underline{E}(L_l) = 2.2401$ and $\overline{E}(L_l) = 22.0169$ (reported in the full tables in [1]), and it remains highly competitive under quadratic loss. The agreement between linear and quadratic loss is expected here because both penalise the magnitude of underperformance, with quadratic loss placing additional weight on occasional large discrepancies between selected and non-selected doses.

The third event, selecting a subset that contains the best dose, $\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, is less restrictive than the previous event because it only requires that the overall maximum future observation arises from one of the selected doses. As a result, expected losses for this event are typically smaller than for $\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$, and the ranking of subsets is driven primarily by whether the subset includes the strongest doses. In this example, subsets containing Dose 4 (and typically also Dose 3) perform best under all three loss functions, while subsets dominated by Doses 1 and 6 perform poorly; full rankings and values are again provided in [1].

Finally, we illustrate strong and weak preference using the expected zero–one losses for selecting the best dose. Dose 4 is strongly preferred over Doses 1, 5, and 6 because $\overline{E}(L_{\{0,1\}})$ for Dose 4 equals 0.7459, which is smaller than the corresponding $\underline{E}(L_{\{0,1\}})$ values for Doses 1, 5, and 6 (0.8932, 0.8428, and 0.8880). In contrast, Dose 4 is only weakly preferred over Dose 3 because both bounds for Dose 4 are smaller than those for Dose 3, but the intervals overlap: $(0.4304, 0.7459)$ versus $(0.6216, 0.8537)$. Hence Dose 3 remains a reasonable alternative. Analogous preference comparisons can be made for subset selection; for example, the subset $\{2, 3, 4, 5\}$ is strongly preferred over subsets whose lower expected zero–one loss exceeds 0.7155, while several subsets remain only weakly dominated due to overlapping loss intervals.

Overall, this example shows that while zero–one loss may produce slightly different fine-grained rankings than linear and quadratic loss, the main conclusions are consistent across the three events: Dose 4 is preferred when selecting a single best dose, and subsets concentrating on Doses $\{2, 3, 4, 5\}$ are preferred for both subset-selection events. This supports the use of NPI expected loss bounds as an assumption-lean summary of predictive uncertainty for multiple comparisons.

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 the behaviour of the method across different sample sizes and loss functions, and to assess whether the three loss functions lead to consistent conclusions.

Example 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 standard deviation 1. The group distributions are $X_1 \sim N(0, 1)$, $X_2 \sim N(0.2, 1)$, $X_3 \sim N(0.4, 1)$, and $X_4 \sim N(0.6, 1)$. To compute the NPI lower and upper

Event	n	Zero–one	Linear	Quadratic
Select the best group	10	0.547	0.567	0.555
	20	0.691	0.726	0.717
	50	0.804	0.827	0.818
	100	0.902	0.919	0.910
Subset of the best groups	10	0.549	0.531	0.472
	20	0.670	0.667	0.642
	50	0.799	0.811	0.779
	100	0.903	0.910	0.881
Subset that includes the best group	10	0.512	0.515	0.507
	20	0.660	0.648	0.628
	50	0.800	0.772	0.752
	100	0.880	0.854	0.823

Table 3: Hit rates, defined as the proportion of repetitions in which the correct group or subset is selected, Example 2.

expected linear and quadratic losses, future observations are bounded by $l = -5$ and $r = 5$.

To assess the stability of the conclusions under repeated sampling, the simulation is repeated h times for each sample size, generating data from the same Normal distributions in each repetition. The NPI lower and upper expected losses provided below are averaged over these h repetitions, with $h = 1000$ for $n = 10, 20, 50$ and $h = 900$ for $n = 100$, the latter reduced slightly due to the greater computational cost of the exact calculations at this sample size. In addition, for each repetition the group (or subset) with the smallest NPI lower expected loss is recorded as the selected best, and the resulting *hit rate*, the proportion of repetitions in which the selection matches the true best group or subset, is reported for each event and loss function.

Table 3 reports the hit rates for the three events of interest: selecting the best group, selecting a subset of the best groups, and selecting a subset that includes the best group. For all three events, the hit rate increases steadily with n, from around 0.5 when $n = 10$ to around 0.9 when $n = 100$, and the three loss functions produce very similar hit rates at every sample size, indicating that the choice of loss function has little effect on the reliability of the selection.

Tables 4, 5, and 6 present the average NPI lower and upper expected losses for the three events: selecting the best group ($X_\ell^f \geq \max_{j \neq \ell} X_j^f$), selecting a subset of the best groups ($\min_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$), and selecting a subset that includes the best group ($\max_{\ell \in S} X_\ell^f \geq \max_{j \in NS} X_j^f$), respectively.

For selecting the best group, Table 4 shows that the smallest average NPI lower and upper expected linear losses are consistently associated with selecting group 4 as the best, for all sample sizes. As before, the width of the NPI interval $[\underline{E}(L_l), \overline{E}(L_l)]$ decreases as n increases, reflecting reduced uncertainty with more data: for group 4 the interval width is about 1.40 when $n = 10$ and about 0.17 when $n = 100$. In contrast, selecting group 1 remains consistently the poorest decision, yielding the largest average lower and

	$n=10$		$n=20$		$n=50$		$n=100$	
ℓ	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
Zero–one loss								
1	0.7467	0.8929	0.7924	0.8707	0.8239	0.8565	0.8337	0.8503
2	0.6968	0.8574	0.7428	0.8292	0.7711	0.8073	0.7806	0.7990
3	0.6311	0.8094	0.6804	0.7759	0.7027	0.7433	0.7127	0.7333
4	0.5618	0.7574	0.5940	0.7014	0.6239	0.6690	0.6334	0.6564
Linear loss ($l=-5$, $r=5$)								
1	1.1848	2.7175	1.2652	2.0896	1.3281	1.6737	1.3439	1.5194
2	1.0092	2.5346	1.0770	1.8977	1.1345	1.4783	1.1499	1.3245
3	0.8220	2.3364	0.9022	1.7171	0.9334	1.2746	0.9471	1.1204
4	0.6583	2.1595	0.6988	1.5052	0.7405	1.0783	0.7519	0.9234
Quadratic loss ($l=-5$, $r=5$)								
1	2.6271	12.9491	2.8127	8.1550	2.9755	5.1596	3.0029	4.1008
2	2.1016	12.0604	2.2356	7.3675	2.3866	4.4793	2.4178	3.4691
3	1.5914	11.1505	1.7573	6.6857	1.8239	3.8182	1.8494	2.8504
4	1.1795	10.3662	1.2505	5.9308	1.3322	3.2277	1.3504	2.3006

Table 4: Average NPI lower and upper expected losses for selecting the best group, Example 2.

upper expected losses at every sample size.

For selecting a subset of the best groups, Table 5 indicates that choosing $\{2,3,4\}$ is consistently best across all sample sizes, yielding the smallest average lower and upper expected losses under every loss function. The subset $\{1,2\}$ is consistently the worst-performing choice at every sample size, again a more stable pattern than under a single simulation run.

For selecting a subset that includes the best group, Table 6 leads to the same qualitative conclusion as Table 5: subsets containing groups 2, 3, and 4 perform best, with $\{2,3,4\}$ again yielding the smallest average expected losses throughout. The expected losses for this event are, as expected, systematically smaller than for selecting a subset of the best groups, since the requirement is less stringent. In particular, for $n=50$ and $n=100$, nearly all average NPI lower and upper expected linear losses in Table 6 remain below 1.

Turning to the zero–one and quadratic losses, the same qualitative conclusions are reached throughout: group 4 (respectively $\{2,3,4\}$) consistently yields the smallest average expected loss across all sample sizes and both subset-selection events. As expected, quadratic loss produces substantially larger numerical values than linear loss, since it penalises larger deviations more heavily, but this does not change the overall decisions.

Strong and weak preference can also be assessed using the averaged expected losses. From Table 4, no strong preference among groups is observed for $n=10$ and $n=20$, since the NPI upper expected linear loss of group 4 exceeds the NPI lower expected linear loss of the other groups at these sample sizes. For $n=50$, group 4 is strongly preferred over groups 1 and 2, but not over group 3. For $n=100$, group 4 is strongly

	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
S	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
Zero–one loss								
$\{1,2\}$	0.8479	0.9373	0.8781	0.9237	0.8975	0.9157	0.9028	0.9120
$\{1,3\}$	0.8123	0.9170	0.8492	0.9024	0.8663	0.8882	0.8730	0.8840
$\{1,4\}$	0.7748	0.8941	0.8051	0.8683	0.8288	0.8546	0.8356	0.8486
$\{2,3\}$	0.7720	0.8922	0.8101	0.8723	0.8281	0.8539	0.8353	0.8484
$\{2,4\}$	0.7242	0.8608	0.7547	0.8280	0.7804	0.8105	0.7875	0.8028
$\{3,4\}$	0.6587	0.8156	0.6992	0.7826	0.7182	0.7532	0.7256	0.7434
$\{1,2,3\}$	0.7471	0.8934	0.7943	0.8723	0.8226	0.8553	0.8339	0.8505
$\{1,2,4\}$	0.6982	0.8584	0.7362	0.8234	0.7708	0.8070	0.7802	0.7986
$\{1,3,4\}$	0.6328	0.8107	0.6801	0.7759	0.7035	0.7441	0.7130	0.7336
$\{2,3,4\}$	0.5583	0.7546	0.5989	0.7057	0.6246	0.6697	0.6333	0.6563
Linear loss ($l = -5$, $r = 5$)								
$\{1,2\}$	1.4203	3.0428	1.4967	2.3624	1.5591	1.9193	1.5721	1.7547
$\{1,3\}$	1.2731	2.9120	1.3615	2.2352	1.4074	1.7714	1.4208	1.6054
$\{1,4\}$	1.1322	2.7831	1.1917	2.0730	1.2483	1.6153	1.2612	1.4473
$\{2,3\}$	1.1252	2.7766	1.2061	2.0869	1.2488	1.6158	1.2629	1.4489
$\{2,4\}$	0.9756	2.6351	1.0259	1.9120	1.0796	1.4488	1.0929	1.2801
$\{3,4\}$	0.8031	2.4675	0.8653	1.7541	0.8969	1.2673	0.9086	1.0964
$\{1,2,3\}$	1.1787	2.8076	1.2735	2.1517	1.3284	1.6965	1.3439	1.5311
$\{1,2,4\}$	1.0118	2.6334	1.0680	1.9417	1.1346	1.5010	1.1482	1.3345
$\{1,3,4\}$	0.8244	2.4354	0.8928	1.7615	0.9335	1.2973	0.9454	1.1304
$\{2,3,4\}$	0.6524	2.2495	0.7070	1.5674	0.7407	1.1012	0.7518	0.9351
Quadratic loss ($l = -5$, $r = 5$)								
$\{1,2\}$	3.2423	15.3244	3.4416	9.6388	3.6312	6.1456	3.6576	4.9192
$\{1,3\}$	2.7823	14.6925	3.0036	9.1092	3.1367	5.6062	3.1599	4.3984
$\{1,4\}$	2.3571	14.0792	2.4925	8.4668	2.6404	5.0567	2.6637	3.8747
$\{2,3\}$	2.3353	14.0552	2.5267	8.5184	2.6437	5.0620	2.6735	3.8858
$\{2,4\}$	1.9173	13.4059	2.0196	7.8552	2.1494	4.5052	2.1782	3.3581
$\{3,4\}$	1.4651	12.6576	1.5847	7.2688	1.6545	3.9367	1.6765	2.8184
$\{1,2,3\}$	2.6038	13.8812	2.8363	8.7253	2.9808	5.3899	3.0048	4.2194
$\{1,2,4\}$	2.1086	12.9916	2.2260	7.8516	2.3857	4.6909	2.4112	3.5724
$\{1,3,4\}$	1.5916	12.0183	1.7247	7.1210	1.8242	4.0190	1.8413	2.9457
$\{2,3,4\}$	1.1608	11.1502	1.2645	6.4064	1.3305	3.4151	1.3515	2.3995

Table 5: Average NPI lower and upper expected losses for selecting a subset of the best groups, Example 2.

preferred over all three remaining groups. Similar preference patterns are observed for the subset-selection events in Table 5, with $\{2,3,4\}$ becoming strongly preferred over an increasing number of competing subsets as n grows.

Overall, the results confirm that group 4 remains the preferred selection, and $\{2,3,4\}$ the preferred subset, across all sample sizes and all three loss functions, when the simulation is repeated many times rather than run once. The hit rates in Table 3 provide direct, quantitative confirmation that this conclusion becomes increasingly reliable as the sample size grows, while the narrowing NPI loss intervals confirm that uncertainty decreases correspondingly.

	$n=10$		$n=20$		$n=50$		$n=100$	
S	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$	$\overline{E}(L)$
Zero–one loss								
$\{1,2\}$	0.4950	0.7190	0.5597	0.6811	0.6044	0.6554	0.6189	0.6449
$\{1,3\}$	0.4377	0.6642	0.5014	0.6241	0.5379	0.5896	0.5520	0.5782
$\{1,4\}$	0.3774	0.6047	0.4209	0.5441	0.4613	0.5132	0.4739	0.5003
$\{2,3\}$	0.3953	0.6226	0.4559	0.5791	0.4868	0.5387	0.4997	0.5261
$\{2,4\}$	0.3358	0.5623	0.3759	0.4986	0.4104	0.4621	0.4218	0.4480
$\{3,4\}$	0.2810	0.5050	0.3189	0.4403	0.3446	0.3956	0.3551	0.3811
$\{1,2,3\}$	0.2426	0.4382	0.2986	0.4060	0.3310	0.3761	0.3436	0.3666
$\{1,2,4\}$	0.1906	0.3689	0.2241	0.3196	0.2567	0.2973	0.2667	0.2873
$\{1,3,4\}$	0.1426	0.3032	0.1708	0.2572	0.1927	0.2289	0.2010	0.2194
$\{2,3,4\}$	0.1071	0.2533	0.1293	0.2076	0.1435	0.1761	0.1497	0.1663
Linear loss ($l=-5$, $r=5$)								
$\{1,2\}$	0.5382	1.4514	0.6067	1.0853	0.6608	0.8583	0.6775	0.7772
$\{1,3\}$	0.4392	1.3240	0.5069	0.9713	0.5472	0.7383	0.5602	0.6566
$\{1,4\}$	0.3464	1.2026	0.3851	0.8310	0.4263	0.6101	0.4382	0.5309
$\{2,3\}$	0.3740	1.2352	0.4361	0.8877	0.4676	0.6531	0.4796	0.5731
$\{2,4\}$	0.2918	1.1245	0.3258	0.7590	0.3572	0.5355	0.3681	0.4580
$\{3,4\}$	0.2239	1.0282	0.2541	0.6730	0.2760	0.4477	0.2844	0.3710
$\{1,2,3\}$	0.2035	0.6776	0.2505	0.5001	0.2790	0.3818	0.2884	0.3403
$\{1,2,4\}$	0.1461	0.5886	0.1708	0.3999	0.1955	0.2901	0.2038	0.2516
$\{1,3,4\}$	0.0994	0.5132	0.1173	0.3318	0.1344	0.2226	0.1396	0.1841
$\{2,3,4\}$	0.0687	0.4621	0.0832	0.2871	0.0918	0.1752	0.0960	0.1381
Quadratic loss ($l=-5$, $r=5$)								
$\{1,2\}$	0.8911	5.0351	0.9956	3.0841	1.0927	1.9375	1.1207	1.5435
$\{1,3\}$	0.6863	4.6131	0.7930	2.7766	0.8609	1.6581	0.8783	1.2768
$\{1,4\}$	0.5076	4.2299	0.5596	2.4155	0.6234	1.3715	0.6398	1.0137
$\{2,3\}$	0.5610	4.3213	0.6543	2.5510	0.7064	1.4670	0.7235	1.1038
$\{2,4\}$	0.4110	3.9826	0.4539	2.2320	0.4997	1.2146	0.5160	0.8733
$\{3,4\}$	0.2934	3.6912	0.3312	2.0259	0.3622	1.0402	0.3720	0.7108
$\{1,2,3\}$	0.2789	1.9987	0.3405	1.2213	0.3821	0.7401	0.3937	0.5733
$\{1,2,4\}$	0.1878	1.8026	0.2162	1.0323	0.2471	0.5797	0.2586	0.4256
$\{1,3,4\}$	0.1177	1.6417	0.1367	0.9083	0.1589	0.4725	0.1636	0.3210
$\{2,3,4\}$	0.0757	1.5376	0.0913	0.8331	0.1010	0.4013	0.1059	0.2567

Table 6: Average NPI lower and upper expected losses for selecting a subset that includes the best group, Example 2.

5 NPI for ranking subsets using loss functions

In decision-making problems involving comparisons of multiple populations, classical statistical tools such as hypothesis testing have well-known limitations. Although they are useful for assessing evidence of differences or equivalence between populations, they are less suitable when the main objective is to *rank* populations (e.g. to identify the largest, second-largest, and so on). In such settings, it is often more relevant to select a ranked subset of the best groups in order to support subsequent decisions. Loss functions can then be used to quantify the cost incurred when the ranking of the selected groups deviates from the optimal ranking.

The aim here 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.

In this section, we consider a general multi-level loss function for scenarios in which future observations from a ranked subset of three or more best groups are to be ordered. The loss function provides a refined assessment of ranking performance by assigning different penalties to different types of ranking errors, reflecting that some deviations from the target ordering may be more consequential than others. The zero–one loss can be viewed as a special case of this framework, where only a binary distinction is made between correct and incorrect rankings and all ranking errors receive the same penalty. The proposed multi-level loss function extends this idea by assigning graded penalties to different types of ranking mistakes when ordering the w selected groups. It therefore allows the quality of the ranking procedure to be evaluated while distinguishing between mild and severe departures from the desired ordering.

Let k independent groups be observed, and let $S = \{\ell_1, \ell_2, \ldots, \ell_w\} \subset \{1, 2, \ldots, k\}$ denote a selected subset of size w, with complement $NS = \{1, 2, \ldots, k\} \setminus S$. We are interested in the event that the w selected groups not only outperform the non-selected groups, but are also correctly ordered:

$$\max_{j_z \in NS} X^f_{j_z} \;<\; X^f_{\ell_w} \;<\; \cdots \;<\; X^f_{\ell_2} \;<\; X^f_{\ell_1}.$$

Define the augmented set of $(w+1)$ quantities $\Big\{ \max\limits_{j_z \in NS} X^f_{j_z},\; X^f_{\ell_w},\; \ldots,\; X^f_{\ell_2},\; X^f_{\ell_1} \Big\}$, whose relative order can occur in $(w+1)!$ different ways. Exactly one ordering corresponds to the target event above; all remaining $(w+1)! - 1$ orderings represent incorrect rankings. We therefore define a general multi-level loss function that assigns a distinct loss value to each ordering.

$$L_G\Big(\max_{j_z \in NS} X^f_{j_z},\; X^f_{\ell_w}, \ldots, X^f_{\ell_2},\; X^f_{\ell_1} \Big) = \begin{cases} L_0, & \text{if } \max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \cdots < X^f_{\ell_2} < X^f_{\ell_1}, \\ L_1, & \text{if } \max\limits_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \cdots < X^f_{\ell_1} < X^f_{\ell_2}, \\ \vdots & \vdots \\ L_{(w+1)!-1}, & \text{if } X^f_{\ell_1} < X^f_{\ell_2} < \cdots < X^f_{\ell_w} < \max\limits_{j_z \in NS} X^f_{j_z}. \end{cases} \tag{25}$$

Here $L_0 = 0$ corresponds to a correct ranking of the w selected groups, with all of them exceeding the best non-selected group. Any other ordering incurs a positive loss, with larger values reserved for more undesirable outcomes. For example, the ordering

$$\max_{j_z \in NS} X^f_{j_z} < X^f_{\ell_w} < \cdots < X^f_{\ell_1} < X^f_{\ell_2}$$

represents an inversion between $X^f_{\ell_1}$ and $X^f_{\ell_2}$ and is penalised by $L_1 > L_0$. At the opposite extreme, the ordering

$$X^f_{\ell_1} < X^f_{\ell_2} < \cdots < X^f_{\ell_w} < \max_{j_z \in NS} X^f_{j_z}$$

is the worst-case scenario, in which *all* selected groups fall below the best non-selected group; this is assigned the largest loss $L_{(w+1)!-1}$. Since the number of orderings grows factorially with w, direct assessment becomes increasingly demanding as w increases, particularly when the data are large or contain substantial overlap.

As in the earlier sections, NPI is applied under Hill's assumption $A_{(n)}$. For the *lower* expected general multi-level loss, probability masses are placed to favour the target ordering: the mass for the "best" selected group $X^f_{\ell_1}$ is assigned to the *right* endpoint of its interval, while masses for non-selected groups are assigned to the *left* endpoints of their intervals, so as to minimise the attainable loss. For the *upper* expected loss, the assignments are reversed to maximise the attainable loss: the mass for $X^f_{\ell_1}$ is assigned to the *left* endpoint, while masses for non-selected groups are assigned to the *right* endpoints.

When $w \geq 3$, there are $w-1$ "middle" selected groups, $X^f_{\ell_2}, \ldots, X^f_{\ell_w}$, and their probability-mass placement affects which orderings are feasible within a given combination of NPI intervals. Determining the exact minimum (or maximum) attainable loss over all placements can therefore be computationally challenging. To obtain practical bounds, we use a heuristic optimisation in which each middle group is treated separately: for each $i = 2, \ldots, w$, the mass $1/(n_{\ell_i}+1)$ is assigned to a point (e.g. the midpoint, avoiding ties when needed) within a sub-interval of $(x_{\ell_i, u_{\ell_i}-1}, x_{\ell_i, u_{\ell_i}})$ chosen to minimise (for the lower bound) or maximise (for the upper bound) the attainable loss. Repeating this for all middle groups yields an *approximate* lower and upper expected general multi-level loss.

There are $\prod_{t=1}^{k}(n_t + 1)$ combinations of NPI interval orderings across the k groups. For each ordering O_i, we determine which loss values among $\{L_0, \ldots, L_{(w+1)!-1}\}$ are feasible under the endpoint/sub-interval assignments and then select the smallest feasible loss (for the lower bound) or the largest feasible loss (for the upper bound). The resulting approximations are

$$\underline{E}^{\,l}\Big(L_G(\max_{j_z \in NS} X^f_{j_z}, \ldots, X^f_{\ell_1})\Big) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t+1)} \Big[\min_c L_c \,\Big|\, O_i\Big], \tag{26}$$

$$\overline{E}^{\,u}\Big(L_G(\max_{j_z \in NS} X^f_{j_z}, \ldots, X^f_{\ell_1})\Big) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{i=1}^{\prod_{t=1}^{k}(n_t+1)} \Big[\max_c L_c \,\Big|\, O_i\Big], \tag{27}$$

where $c \in \{0, 1, \ldots, (w+1)! - 1\}$ indexes the possible orderings and O_i denotes the ith combination of NPI intervals.

It is worth noting that the zero–one loss function can be viewed as a special case of the proposed multi-

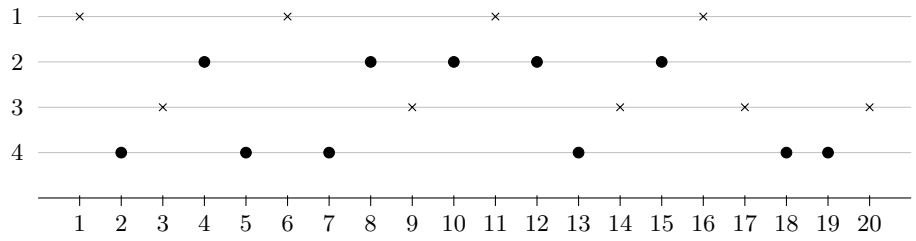


Figure 1: Four-group dataset used in Example 3.

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 7: NPI lower and upper expected multi-level losses for all groups combinations, Example 3

level loss framework. This is obtained by assigning $L_0 = 0$ to the target ordering and setting $L_c = 1$ for all remaining $(w+1)! - 1$ incorrect orderings. In this special case, all ranking errors are treated equally, and the loss function only distinguishes between a correct and an incorrect ranking. Consequently, the corresponding NPI lower and upper expected losses reduce to the NPI lower and upper probabilities associated with the target ranking event and its complement.

Example 3. We consider the dataset from [13], consisting of $k = 4$ groups with sample sizes $n_1 = 4$, $n_2 = 5$, $n_3 = 5$, and $n_4 = 6$ (see Figure 1). Let $S = \{\ell_1, \ell_2\}$ denote the selected subset and $NS = \{j_1, j_2\}$ the non-selected groups. The aim is to illustrate the selection of a ranked subset of the two best groups using

		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 8: NPI lower and upper expected multi-level losses for different loss values for Example 3. 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 $L_5 = 1.00$

		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 9: NPI lower and upper expected multi-level losses for different loss values for Example 3. 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 $L_5 = 1.00$.

the multi-level loss function. The event of interest is $\max_{j_z \in NS} X^f_{j_z} < X^f_{\ell_2} < X^f_{\ell_1}$, where the selected subset is $S = \{\ell_1, \ell_2\}$ with ℓ_1 ranked best and ℓ_2 ranked second, and NS is the complement of S.

In this example there are 1260 possible combinations of orderings of future observations from the $k = 4$ groups, and hence 12 ordered choices for selecting and ranking two groups (all (ℓ_1, ℓ_2) pairs). To compute $\underline{E}(L_{ML})$ for a given ordered pair (ℓ_1, ℓ_2), we consider each of the 1260 orderings and determine the smallest loss level among $L_0, \ldots, L_5$ that can occur under that ordering when the "best" selected group ℓ_1 is taken as large as possible and the non-selected groups are taken as small as possible (with ℓ_2 placed within a sub-interval chosen to minimise the loss, avoiding ties with observed values as described earlier). The

		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 10: NPI lower and upper expected multi-level losses for different loss values for Example 3. 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 $L_5 = 1.00$.

ordering then contributes probability mass $1/1260$ to the corresponding loss level. Summing over all orderings yields the coefficients in Table 7. For example, for $(\ell_1, \ell_2) = (1, 2)$, the coefficient of L_1 in $\underline{E}(L_{ML})$ equals $74/1260 = 0.0587$, meaning that L_1 is the minimum attainable loss in 74 of the 1260 orderings.

The computation of $\overline{E}(L_{ML})$ is analogous, but now we determine the largest attainable loss level for each ordering under the endpoint allocations that make ℓ_1 as small as possible and the non-selected groups as large as possible (again choosing the sub-interval for ℓ_2 to maximise the loss). For $(\ell_1, \ell_2) = (1, 2)$, the coefficient of L_5 in $\overline{E}(L_{ML})$ equals $681/1260 = 0.5405$, meaning that the worst-case attainable loss is L_5 in 681 orderings. Repeating this procedure for all ordered pairs gives the full set of linear combinations in Table 7.

Table 7 highlights how the coefficients shift between lower and upper expectations. For $\underline{E}(L_{ML})$, coefficients attached to the smallest loss levels (in particular L_0) are typically larger than those attached to L_1, reflecting the optimistic endpoint allocation used for the lower expectation. In contrast, for $\overline{E}(L_{ML})$ the coefficient attached to L_5 is often substantial, reflecting the pessimistic endpoint allocation that drives the upper expectation. Moreover, across all 12 ordered pairs, the combined weight on the higher-loss levels (L_4 and L_5) is typically larger than the combined weight on the lower-loss levels (L_0 and L_1), indicating that severe ranking errors are feasible under many orderings even when the selected subset contains strong groups.

Note that there are six unordered pairs of groups, but twelve ordered pairs. Ordered pairs that involve the same two groups in opposite order, such as $(1, 2)$ and $(2, 1)$, need not yield "interchangeable" coefficient patterns: the ranking direction changes the role of ℓ_1 and ℓ_2, and the sub-interval choice for the middle group

(used to minimise or maximise expected loss) can differ between the two orderings. Consequently, reversing the order may change both the coefficients and the resulting expected loss.

To study sensitivity to the numerical choice of loss levels, nine sets of loss values are considered, ranging from gradually increasing penalties (Case 1) to extreme threshold-type penalties (Cases 7–9). These cases are reported in Tables 8–10. For each case, Table 7 is converted into numerical lower and upper expected multi-level losses by substituting the chosen values of $L_0, \ldots, L_5$. For example, under Case 1, the numerical NPI lower and upper expected multi-level losses for $(\ell_1, \ell_2) = (1, 2)$ are $\underline{E}(L_{ML}) = 0.2900$ and $\overline{E}(L_{ML}) = 0.4190$ (Table 8).

Across Cases 1, 2, 3, 4, 6, 7, and 8, the ranking conclusions are stable: selecting group 3 as best and group 4 as second best yields the smallest lower and upper expected multi-level losses, while selecting group 1 as best and group 2 as second best yields the largest losses. Some cases also produce ties in the induced rankings, driven by the structure of the loss values. For instance, in Case 6, the lower expected multi-level loss for $(3, 4)$ equals that for $(4, 3)$, and in Case 7 the pairs $(1, 4)$ and $(2, 4)$ have identical values, reflecting that certain loss levels are set equal (or to zero) so that only the coefficients of the remaining levels affect the outcome.

Case 5 corresponds to taking $L_1 = \cdots = L_5 = 1$ (and $L_0 = 0$), so the multi-level loss reduces to the zero-one loss for the event of interest. Hence the results in Table 9 are consistent with the relationship between expected zero-one loss and NPI probabilities discussed earlier. In this case, selecting group 3 as best and group 2 as second best gives the smallest expected loss values, and the pattern differs slightly from the "stable" conclusion above because all incorrect outcomes are penalised equally, regardless of severity.

Case 9 (Table 10) isolates the most severe error by taking $L_5 = 1$ and $L_0 = \cdots = L_4 = 0$, so the numerical expected loss depends only on the coefficient of L_5. This leads to different behaviour: the ordered pair $(4, 2)$ has the largest lower expected multi-level loss (equal to its lower coefficient on L_5), while $(1, 2)$ has the largest upper expected multi-level loss (equal to its upper coefficient on L_5). Such behaviour is expected because Case 9 ignores all errors except the single worst category.

Overall, this example shows how the multi-level loss framework can support selecting and ranking the two best groups, and how the resulting decision can be stable across a broad range of loss specifications, while extreme or degenerate specifications (such as Cases 5 and 9) can shift the preferred ordering by collapsing the grading structure of the loss.

A more extensive numerical study, including related settings with larger sample sizes and simulated populations, is reported in [1]. The additional results confirm the sensitivity of the selected ranked subset to the specification of the loss levels and further illustrate the flexibility of the proposed multi-level loss

framework.

The previous example focused on the special case of ranking two selected groups. The following example extends the framework to a ranked subset of the three best groups, illustrating how the general multi-level loss can be used to guide the ranking of future observations and how the resulting expected-loss bounds reflect the quality of the induced ranking.

Example 4. This example illustrates the use of the general multi-level loss function for ranking a subset of the three best groups. The data are taken from the `lengthWeightData` dataset in the `StatCharrms` R package [34], which contains treatment concentration and fish weight information. In this paper, five treatment groups are used, labelled $1, \ldots, 5$. The aim is to rank the three best groups based on their future observations. In general, selecting and ranking three groups from five leads to $\binom{5}{3} \times 3! = 60$ possible ranked selections. These correspond to first choosing a subset of three groups and then ranking the groups within that subset.

To keep the presentation manageable, attention is restricted to one configuration where the selected subset is fixed as $S = \{1, 2, 3\}$ and the non-selected subset is $NS = \{4, 5\}$. Once the subset is fixed, only the ranking of the three selected groups remains to be determined. Hence there are $3! = 6$ possible rankings: $1 \succ 2 \succ 3$, $1 \succ 3 \succ 2$, $2 \succ 1 \succ 3$, $2 \succ 3 \succ 1$, $3 \succ 1 \succ 2$, and $3 \succ 2 \succ 1$.

For each ranking, the loss function depends on the relative ordering of four quantities: the maximum future observation among the non-selected groups, $\max(X_4^f, X_5^f)$, together with the three future observations from the selected groups. For example, if the ranking is $1 \succ 2 \succ 3$, the event of interest is $\max(X_4^f, X_5^f) < X_3^f < X_2^f < X_1^f$. Since four quantities are compared, there are $4! = 24$ possible orderings. One ordering corresponds to the correct ranking and is assigned loss $L_0 = 0$, while the remaining 23 correspond to incorrect rankings with loss values $L_1, \ldots, L_{23}$.

The approximate NPI lower and upper expected losses for each ranking are linear combinations of the loss values $L_0, \ldots, L_{23}$ with coefficients determined by the NPI ordering probabilities. The full coefficient tables for the six rankings are lengthy and are therefore omitted here; they are reported in [1].

To examine the influence of the loss specification, several sets of loss values were considered. In this paper, we report representative results for three cases. Case 2 uses gradually increasing small losses with $L_c = 0.02c$ for $c = 0, \ldots, 22$ and $L_{23} = 0.50$. Case 3 corresponds to the zero–one loss specification with $L_0 = 0$ and $L_c = 1$ for $c = 1, \ldots, 23$. Case 5 uses larger increasing losses with $L_0 = 0$, $L_1 = 0.50$, $L_c = 0.50 + 0.02(c-1)$ for $c = 2, \ldots, 22$, and $L_{23} = 1.00$. The complete results for all considered loss specifications are provided in [1].

Table 11 shows that the ranking $1 \succ 2 \succ 3$ consistently yields the smallest approximate lower expected

Ranking (ℓ_1, ℓ_2, ℓ_3)	Case 2		Case 3		Case 5	
	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$	$\underline{E}(L)$	$\overline{E}(L)$
$1 \succ 2 \succ 3$	0.0821	0.2449	0.6768	0.9483	0.4166	0.7175
$1 \succ 3 \succ 2$	0.1101	0.2453	0.8348	0.9401	0.5223	0.7140
$2 \succ 1 \succ 3$	0.0951	0.2440	0.7937	0.9385	0.4870	0.7122
$2 \succ 3 \succ 1$	0.1067	0.2454	0.8226	0.9424	0.5134	0.7149
$3 \succ 1 \succ 2$	0.0994	0.2285	0.7674	0.9030	0.4786	0.6792
$3 \succ 2 \succ 1$	0.0994	0.2328	0.7674	0.9276	0.4786	0.6929

Table 11: Approximate NPI lower and upper expected general multi-level losses for three representative loss specifications, Example 4.

Ranking	Exact Lower	Exact Upper	Heuristic Lower	Heuristic Upper
$1 \succ 2 \succ 3$	0.0868	0.2533	0.0821	0.2449
$2 \succ 1 \succ 3$	0.1002	0.2647	0.0951	0.2440
$1 \succ 3 \succ 2$	0.1017	0.2666	0.1067	0.2454
$3 \succ 2 \succ 1$	0.1382	0.3063	0.0994	0.2285
$2 \succ 3 \succ 1$	0.1326	0.2958	0.1101	0.2453
$3 \succ 1 \succ 2$	0.1251	0.2877	0.0994	0.2328

Table 12: Comparison of exact and heuristic NPI lower and upper expected general multi-level losses for the six possible rankings in Case 2, Example 4.

loss across the considered cases. This suggests that groups 1, 2, and 3 are plausibly ordered in this way when decisions are based on the lower expected loss. In contrast, the ranking $3 \succ 1 \succ 2$ tends to produce the smallest approximate upper expected loss, indicating that the ranking preferred under a more cautious perspective may differ from the ranking preferred under an optimistic perspective. The scale of the expected losses is also influenced by the magnitude of the loss values themselves, as illustrated by the larger values obtained under the zero–one specification.

To assess the accuracy of the proposed approximation, we compare the heuristic results with the exact NPI lower and upper expected general multi-level losses for Case 2. For this example, the exact calculation involves 7,776 possible combinations. Table 12 reports the exact and approximate values for the six possible rankings. The results show that the heuristic approximation provides values close to the exact bounds, preserves the ordering of the candidate rankings, and slightly underestimates the exact lower and upper expected losses in most cases.

Overall, this example demonstrates how the general multi-level loss function can be used to compare candidate rankings of a selected subset of groups. While some rankings remain favourable across different loss specifications, the preferred ranking may depend on whether decisions are based on lower or upper expected loss values. The complete numerical results underlying this example are provided in [1].

6 General event of selection and ranking based on loss functions

This section considers a general event of selection and ranking in which groups are allocated to ordered buckets, and the interest lies in the ranking of these buckets rather than in the ranking of groups within each bucket. The aim is to measure the penalty associated with suboptimal bucket rankings and to evaluate the quality of the resulting ordering using loss functions.

Let $K_1, \ldots, K_s$ denote s ordered buckets containing one or more independent groups. The construction of these buckets can be guided by summary statistics derived from the data, such as group medians. For example, groups may be ordered according to their medians and then assigned to buckets accordingly. However, such data-driven allocations do not necessarily lead to optimal decisions in terms of NPI probabilities or expected losses. In addition to determining how groups are allocated, one must also choose the number of buckets. Different choices may be considered depending on the objective, such as maximising NPI lower and upper probabilities or minimising expected losses. When the number of groups is small, it is feasible to examine all possible allocations and numbers of buckets, allowing a systematic comparison of their impact on the resulting rankings.

Given the number of buckets and the group allocation, the event of interest is

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

which states that the maximum future observation from bucket K_i is smaller than the minimum future observation from bucket K_{i+1} for all $i = 1, \ldots, s-1$. Hence the buckets are correctly ordered according to their future observations.

In principle, the general multi-level loss function introduced earlier could also be applied in this setting. However, such a loss assigns a distinct penalty to every possible ordering of the groups. This leads to a very large number of loss levels. Specifying and interpreting such a large number of loss levels is often impractical. For this reason, we instead consider linear and quadratic loss functions as simpler alternatives. These losses penalise deviations from the desired bucket ordering while ignoring the ordering of groups within each bucket. Both loss functions measure the extent to which the ordering condition in event A is violated.

Linear loss function

The linear loss function measures the deviation between consecutive buckets. Specifically, it evaluates the difference between the largest future observation in bucket K_i and the smallest future observation in bucket K_{i+1}. The penalty increases linearly with the size of this deviation. For the event A, the linear loss function

is defined as

$$L_l\left(\bigcap_{i=1}^{s-1}\left\{\max_{t\in K_i} X_t^f, \min_{t\in K_{i+1}} X_t^f\right\}\right) = \begin{cases} 0 & \text{if } A \text{ is satisfied,} \\ D_l & \text{otherwise,} \end{cases} \tag{28}$$

where $D_l = \sum_{i=1}^{s-1} \max\left\{\max_{t\in K_i} X_t^f - \min_{t\in K_{i+1}} X_t^f, 0\right\}$. Thus, when the ordering condition is violated, the loss is defined as the sum of the deviations between consecutive buckets. Smaller values of L_l indicate that the bucket ordering is close to the desired ranking, whereas larger values correspond to larger violations.

We now consider the NPI lower and upper expected linear losses. Exact calculation of these expectations for event A is computationally demanding, so we derive bounds that serve as a lower bound for the NPI lower expected loss and an upper bound for the NPI upper expected loss.

For each group t ($t = 1, \ldots, k$), let the observed values be ordered as $x_{t,1} < x_{t,2} < \ldots < x_{t,n_t}$, and define $x_{t,0} = l$ and $x_{t,n_t+1} = r$ for notational convenience. The values l and r bound future observations below the minimum and above the maximum observed values, respectively. This avoids infinite bounds when computing expected losses. The inference is based on Hill's assumption $A_{(n_t)}$ for each group t.

To obtain the lower bound for the NPI lower expected linear loss, the probability mass for groups in bucket K_i is assigned to the left endpoint of each interval, while the probability mass for groups in bucket K_{i+1} is assigned to the right endpoint. For the upper bound of the NPI upper expected linear loss, the assignments are reversed. This leads to the following bounds for event A:

$$\underline{E}^L(L_l(A)) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{l_1=1}^{n_1+1}\sum_{l_2=1}^{n_2+1}\cdots\sum_{l_k=1}^{n_k+1}\sum_{i=1}^{s-1} L_l\Big(\max_{t\in K_i}(x_{t,l_t-1}) < \min_{t\in K_{i+1}}(x_{t,l_t})\Big), \tag{29}$$

$$\overline{E}^U(L_l(A)) = \frac{1}{\prod_{t=1}^k (n_t+1)} \sum_{l_1=1}^{n_1+1}\sum_{l_2=1}^{n_2+1}\cdots\sum_{l_k=1}^{n_k+1}\sum_{i=1}^{s-1} L_l\Big(\max_{t\in K_i}(x_{t,l_t}) < \min_{t\in K_{i+1}}(x_{t,l_t-1})\Big). \tag{30}$$

Quadratic loss function

The quadratic loss function is defined similarly to the linear loss, but squares the deviations between consecutive buckets. This leads to larger penalties when the violations of the desired ordering are substantial. For event A, the quadratic loss is defined as

$$L_q\left(\bigcap_{i=1}^{s-1}\left\{\max_{t\in K_i} X_t^f, \min_{t\in K_{i+1}} X_t^f\right\}\right) = \begin{cases} 0 & \text{if } A \text{ is satisfied,} \\ D_q & \text{otherwise,} \end{cases} \tag{31}$$

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 13: Observed values for four groups with ten observations each, Example 5.

where $D_q = \sum_{i=1}^{s-1} \left(\max \left\{ \max_{t \in K_i} X_t^f - \min_{t \in K_{i+1}} X_t^f, 0 \right\} \right)^2$. Hence, the corresponding lower and upper bounds for the NPI lower and upper expected quadratic losses are given by

$$\underline{E}^L(L_q(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_q\Big(\max_{t \in K_i}(x_{t,l_t-1}) < \min_{t \in K_{i+1}}(x_{t,l_t}) \Big), \tag{32}$$

$$\overline{E}^U(L_q(A)) = \frac{1}{\prod_{t=1}^{k}(n_t+1)} \sum_{l_1=1}^{n_1+1} \sum_{l_2=1}^{n_2+1} \cdots \sum_{l_k=1}^{n_k+1} \sum_{i=1}^{s-1} L_q\Big(\max_{t \in K_i}(x_{t,l_t}) < \min_{t \in K_{i+1}}(x_{t,l_t-1}) \Big). \tag{33}$$

The following example illustrates the application of these loss functions for evaluating bucket rankings.

Example 5. This example uses data from Neuhauser et al. [32], consisting of four groups with ten observations each, as shown in Table 13. The aim is to measure the penalty associated with suboptimal bucket rankings using the NPI lower and upper expected linear and quadratic loss functions.

The four groups are allocated to s ordered, non-empty buckets, and the event of interest is $A = \bigcap_{i=1}^{s-1} \{ \max_{t \in K_i} X_t^f < \min_{t \in K_{i+1}} X_t^f \}$, which represents the desired ordering of the buckets in terms of future observations.

For $k = 4$ groups, there are multiple ways to allocate the groups into ordered buckets. The number of such allocations into s buckets is given by $s!\, S(k, s)$, where $S(k, s)$ is the Stirling number of the second kind. This results in 1, 14, 36, and 24 possible allocations for $s = 1, 2, 3, 4$, respectively. The case $s = 1$ is not of interest, as all groups in a single bucket and no ordering between buckets is involved.

To compute the NPI lower and upper expected losses, future observations are assumed to be bounded by 48 and 158, where the lower bound is smaller than the minimum observed value and the upper bound is larger than the maximum observed value. Such bounds are required for the linear and quadratic loss functions. In the following, representative cases are considered to examine how the linear and quadratic loss functions evaluate these bucket orderings.

Table 14 presents the results for the case of two buckets ($s = 2$). All possible allocations of the four groups into two ordered buckets are considered, and the lower and upper bounds for the NPI lower and upper expected linear and quadratic losses are calculated. From this table, the smallest lower and upper bounds for both the expected linear and quadratic losses correspond to the event $X_1^f < \min(X_2^f, X_3^f, X_4^f)$. For this

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 14: Lower and upper bounds for the NPI lower and upper expected linear and quadratic losses for four groups allocated into two buckets ($s = 2$), Example 5.

event, $\underline{E}^L(L_l) = 2.221$, $\overline{E}^U(L_l) = 13.255$, $\underline{E}^L(L_q) = 35.316$, and $\overline{E}^U(L_q) = 804.748$, indicating the smallest expected penalty among all two-bucket allocations. In contrast, the largest lower and upper bounds occur for the event $\max(X_2^f, X_3^f, X_4^f) < X_1^f$, with $\underline{E}^L(L_l) = 60.706$, $\overline{E}^U(L_l) = 76.518$, $\underline{E}^L(L_q) = 4537.619$, and $\overline{E}^U(L_q) = 6359.053$. This event imposes a ranking that contradicts the observed data and therefore leads to the greatest expected penalty.

The analysis can be extended to cases with more buckets. For three buckets ($s = 3$) and four buckets ($s = 4$), all possible ordered allocations were evaluated in the same way. These results show that several events have very small lower bounds, and in some cases both the lower and upper bounds of the expected losses are equal to zero. In particular, events such as $X_3^f < X_1^f < X_4^f < X_2^f$ and $X_2^f < X_1^f < X_4^f < X_3^f$ lead to zero expected loss under both loss functions. For comparison, the sample medians of the four groups are $m_1 = 64.5$, $m_2 = 105.5$, $m_3 = 104$, and $m_4 = 119$, giving the ordering $m_1 < m_3 < m_2 < m_4$. This ordering does not correspond to the configurations with the smallest expected losses, indicating that allocation based solely on sample medians does not necessarily lead to optimal bucket rankings. The full results for the cases $s = 3$ and $s = 4$ are omitted here for brevity and are reported in [1].

Overall, the linear and quadratic loss functions lead to consistent conclusions regarding the relative quality of bucket rankings. The quadratic loss produces larger numerical penalties for poor rankings because it penalises larger deviations more heavily, but the overall ordering of preferred events remains similar for both loss functions.

7 Conclusion

This paper developed a Nonparametric Predictive Inference (NPI) framework for ranking and selection problems based on loss functions. The aim was to extend probability-based NPI analysis to decision-making settings by evaluating competing decisions through NPI lower and upper expected losses.

The paper first introduced the use of loss functions within NPI for selection problems. Three commonly used loss functions were considered: the zero–one, linear, and quadratic loss functions. These provide different ways of quantifying the penalty associated with selecting suboptimal groups. The zero–one loss distinguishes only between correct and incorrect decisions, whereas the linear and quadratic losses quantify the magnitude of deviations from the desired outcome, with the quadratic loss assigning greater penalties to larger deviations.

The methodology was then developed for multiple comparison problems, where decisions involve selecting and ranking groups based on future observations. Three types of events were considered: selecting the best group, selecting a subset of the best groups, and selecting a subset that includes the best group. The NPI lower and upper expected losses provide a framework for evaluating these decisions and quantifying the potential consequences of incorrect selections or rankings. Competing decisions can be compared using the concepts of strong and weak preference, where strong preference occurs when the NPI upper expected loss of one decision is smaller than the NPI lower expected loss of another, and weak preference occurs when both the lower and upper expected losses of one decision are smaller than those of another.

The paper subsequently extended the framework to the selection and ranking of a subset of the w best groups through a general multi-level loss function. This loss function assigns different penalties to different types of ranking errors, allowing ranking quality to be assessed by distinguishing between mild and severe departures from the desired ordering. The NPI lower and upper expected multi-level losses were obtained through optimisation procedures that determine the minimum and maximum attainable losses. For ranked subsets containing multiple best groups, the presence of intermediate groups increases the complexity of the optimisation problem, and approximate procedures were developed to obtain the corresponding NPI lower and upper expected losses. The examples demonstrated that ranking decisions may depend on the specified loss values, highlighting the importance of selecting a loss structure that reflects the relative severity of different ranking errors.

Finally, the paper introduced a general framework for selection and ranking based on ordered buckets of groups. In this setting, the focus is on the ranking of the buckets rather than the ranking of groups within each bucket. The NPI lower and upper expected linear and quadratic losses were used to quantify the

penalty associated with suboptimal bucket rankings. The results showed that allocation rules based solely on sample summaries, such as medians, do not necessarily lead to the highest NPI lower and upper probabilities or the smallest expected losses. Instead, optimal allocation can be identified by evaluating possible bucket configurations using NPI probabilities or expected losses.

Overall, the proposed framework extends Nonparametric Predictive Inference from probability-based assessment to decision-oriented ranking and selection problems by incorporating loss functions and NPI lower and upper expected losses. It provides a principled approach for evaluating competing decisions under predictive uncertainty.

Future research may extend this work in several directions. Alternative specifications of loss functions could be explored to further enhance the flexibility of the framework. For example, grouped or weighted grouped loss functions may provide a structured way to reduce the number of distinct loss values while retaining the ability to distinguish between different types of ranking errors. Further work could also investigate computational strategies for efficiently evaluating NPI lower and upper expected losses in problems involving larger numbers of groups or more complex ranking structures. An interesting direction is the investigation of multi-level loss functions for ordinal-scale data. Unlike linear and quadratic losses, which typically require a metric scale, multi-level losses may provide a natural framework for decision problems involving ordered categories. Further research is needed to explore the properties and potential applications of such extensions.

Acknowledgement(s)

This research was conducted as part of the first author's PhD at Durham University, supported by a studentship from the University of Jeddah, Saudi Arabia, and the Saudi Cultural Bureau in London.

Disclosure Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

No funding was received.

References

[1] Alshihry, F. I. (2025). *Nonparametric Predictive Inference for Selection and Ranking*. PhD thesis, Durham University. `https://etheses.dur.ac.uk/16269/`.

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

[3] Baker, R. (2010). *Multinomial Nonparametric predictive inference: Selection, Classification and Subcategory Data*. PhD thesis, Durham University. UK. `http://etheses.dur.ac.uk/257/`.

[4] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

[5] Berger, R.L. and Gupta, S.S. (1980). Minimax subset selection rules with applications to unequal variance (unequal sample size) problems. *Scandinavian Journal of Statistics*, 7: 21–26.

[6] Coolen, F. P. A. and Coolen-Maturi, T. (2025). Nonparametric predictive inference. In Lovric, M., editor, *International Encyclopedia of Statistical Science*. Springer.

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

[8] Coolen, F.P.A. (2004). On the use of imprecise probabilities in reliability. *Quality and Reliability Engineering International*, 20: 193–202.

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

[10] Coolen, F.P.A. and Coolen-Maturi, T. and Alqifari, H.N. (2018). Nonparametric predictive inference for future order statistics. *Communications in Statistics: Theory and methods*, 47(10): 2527–2548.

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

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

[13] Coolen-Maturi, T. (2022). The ordering of future observations from multiple groups. *Communications in Statistics: Simulation and Computation*, 51: 7526–7543.

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

[15] Deely, J.J. and Gupta, S.S. (1968). On the properties of subset selection procedures. *Sankhyā: The Indian Journal of Statistics, Series A*, 30: 37–50.

[16] Finner, H. and Giani, G. (1996). Duality between multiple testing and selecting. *Journal of Statistical Planning and Inference*, 54: 201–227.

[17] Girshick, M. A. and Savage, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 53–73, Berkeley, CA. University of California Press.

[18] Goel, P.K. and Rubin, H. (1977). On selecting a subset containing the best population-a Bayesian approach. *The Annals of Statistics*, 5: 969–983.

[19] Gupta, S.S. and Miescke, K.J. (2002). On the performance of subset selection rules under normality. *Journal of Statistical Planning and Inference*, 103: 101–115.

[20] Gupta, S.S. and Panchapakesan, S. (2002). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Society for Industrial and Applied Mathematics, Philadelphia.

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

[22] Hill, B.M. (1988). De Finetti's theorem, induction, and A_n or Bayesian nonparametric predictive inference (with discussion). In Bernardo, J., DeGroot, M., Lindley, D., and Smith, A., editors, *Bayesian Statistics*, volume 3, pages 211–241. Oxford University Press.

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

[24] Hsu, J.C. (1981). Simultaneous confidence intervals for all distances from the best. *The Annals of Statistics*, 9: 1026–1034.

[25] Hsu, J.C. (1982). Simultaneous inference with respect to the best treatment in block designs. *Journal of the American Statistical Association*, 77: 461–467.

[26] Hsu, J.C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman and Hall, London.

[27] Lawless, J. F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika*, 92(3):529–542.

[28] Lehmann, E.L. (1957). A theory of some multiple decision problems. *The Annals of Mathematical Statistics*, 28: 547–572.

[29] Lehmann, E.L. (1961). Some model problems of selection. *The Annals of Mathematical Statistics*, 32: 990–1012.

[30] Maturi, T. (2010). *Nonparametric predictive inference for Multiple Comparisons*. PhD thesis, Durham University, UK. `http://etheses.dur.ac.uk/230/`.

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

[32] Neuhauser, M. and Liu, P.Y. and Hothorn, L. A. (1998). Nonparametric tests for trend: Jonckheere's test, a modification and a maximum test. *Biometrical Journal*, 40: 899–909.

[33] Simpson, D.G. and Margolin, B.H. (1986). Recursive nonparametric testing for dose-response relaflonships subject to downturns at high doses. *Biometrika*, 73: 589–596.

[34] Swintek, J., Flynn, K., Haselman, J., and Sakurai, M. (2020). *StatCharrms: Statistical Analysis of Chemistry, Histopathology, and Reproduction Endpoints Including Repeated Measures and Multi-Generation Studies*. R package version 0.90.96.

[35] Wald, A. (1950). *Statistical Decision Functions*. John Wiley & Sons, New York.