On indifference zone selection with a preference threshold

Citation for published version (APA):

Document status and date:
Published: 01/01/1996

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

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

Link to publication

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

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

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Memorandum COSOR 96-22

On Indifference Zone Selection with a Preference Threshold
(Revised version of 95-02)

F.P.A. Coolen
P. van der Laan

Eindhoven, May 1996
The Netherlands
On Indifference Zone Selection
with a Preference Threshold*

F.P.A. Coolen
University of Durham
Department of Mathematical Sciences
Science Laboratories, South Road
Durham, DH1 3LE, England

P. van der Laan
Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

This paper discusses a selection criterion that generalizes the well-known concept of indifference zone selection through a preference threshold. A population is preferred to another population if the difference in the sums of observed values exceeds a given non-negative threshold value. We present an argument for this selection rule by modelling preference by imprecise previsions. We aim at guidelines to design a selection experiment, which is characterized by two numbers: the number of necessary observations per population, and the preference threshold. Next to the probability of correct selection we also need a second specification. In this paper we consider a probability of false selection that is strongly related to the minimum probability of correct selection. Based on this model the outcome of an experiment may be 'no selection', at least not based on strong preference of a single population. The ideas are presented through a simple selection problem for normal populations with common known variance. Although the theory has a frequentist nature, the derivation and justification of the selection rule through imprecise previsions relies on Bayesian foundations, and via this route we gain more insight into the selection criterion.

Keywords Bayesian inference; Frequentist approach; Imprecise previsions; Indifference zone selection; Normal populations; Preference threshold; Probability of correct selection; Probability of false selection; Probability of no selection; Subset selection.

1 Introduction

In the practical situation we often have to make decisions on the basis of observations being made. In many of these cases a careful statistical analysis is necessary before a decision can be made. Statistical estimation and testing methods provide us a methodology which can help us to analyse the observations in a reasonable way. Let us consider the following problem with wide practical application. From a set of $k$ ($k \geq 2$) populations, described by qualitative variables, we want to select the "best" population, where "best" is associated with the maximal (or minimal) value of an unknown parameter of the population distribution. In such a case, statistical selection methodology enables us to realistically formulate the question concerning the best population and to solve it in an adequate way.

In this paper we shall concentrate on normal populations, $\pi_1, \ldots, \pi_k$, where the population mean is the parameter of interest and with common known variance. The response variables in this paper are supposed to be the $k$ sample sums $\sum_{j=1}^{n} Y_{ij}$, with $Y_{ij}$ the $j$-th observation in the sample of common size $n$ from population $\pi_i (i = 1, \ldots, k)$. The approach usually used in practice is to test the so-called homogeneity hypothesis $H_0 : \mu_1 = \ldots = \mu_k$, where $\mu_i (i = 1, \ldots, k)$ is the mean of population $\pi_i$, by analysis of variance techniques. Multiple comparisons and simultaneous confidence intervals can give additional information. Having a sufficiently large number of observations one would usually expect to reject the homogeneity hypothesis, which in general cannot be the final practical decision. Rejection of $H_0$ implies that the $k$ population means are not equal to each other, a conclusion which can in most cases be drawn even before the observations have been made.

Statistical selection procedures have been developed specifically to answer questions like "Which population is best?". There are two basic approaches developed and presented in the literature. One approach has been developed by Bechhofer [Bech54]. The second approach has been thoroughly investigated by Gupta [Gupt56, Gupt65]. The approaches initiated by Bechhofer and Gupta are indicated by indifference zone approach, and subset selection approach, respectively. The subset selection procedure selects a subset, non-empty and as small as possible, of the $k$ populations in order to include the best population into the subset with a certain confidence. The size of the subset is random and depends among other things on the common variance $\sigma^2$ and the common sample size $n$. In general, we desire a selection rule which makes the expected subset size as small as possible. Gupta's selection rule $R_G$ can be described as follows. Select $\pi_i, i = 1, \ldots, k$, in the subset if and only if

$$\sum_{j=1}^{n} Y_{ij} \geq \max_{1 \leq i \leq k} \sum_{j=1}^{n} Y_{ij} - \tau \sigma \sqrt{n},$$

where the selection constant $\tau > 0$ must be determined such that the probability requirement of a correct selection (CS) with $R_G$

$$P(CS|R_G) \geq P^*$$

is met for all possible values of $\mu_1, \ldots, \mu_k (k^{-1} < P^* < 1)$. In this context a correct selection means that the best population belongs to the selected subset. Even when the ultimate goal in practice is to choose the best, the subset selection approach can be applied to eliminate inferior populations. Another characteristic feature is that this approach can analyse the results after the experiments have already been completed. There are no requirements for the
common sample size $n$ in contrast with the indifference zone approach. It can be expected that small sample sizes will correspond with large subset sizes, and the other way around.

The goal of the indifference zone approach is to indicate the best population. The selection procedure $R_B$ is to select that population that resulted in the largest sample value. The probability requirement for $CS$, is

$$P(CS|R_B) \geq P^*$$

if $\mu[k] - \mu[k-1] \geq \delta^*$, where the ordered means are denoted by $\mu[1] \leq \mu[2] \leq \ldots \leq \mu[k]$ and with $\delta^* > 0$ ($k^{-1} < P^* < 1$). This probability requirement can only be guaranteed if the common sample size $n$ is large enough. The indifference zone approach is useful at the experimental design stage in order to determine the required common sample size $n$. The generalization presented in this paper also explicitly deals with the design of a selection experiment. For strong requirements, as large $P^*$ and small $\delta^*$, using the indifference zone approach one has to pay automatically with large sample sizes. Interesting research has been performed by Hsu [Hsu81, Hsu84, Ste88]. Hsu integrated the indifference zone and subset selection approaches. His method is known as “multiple comparisons with the best”. An alternative subset selection procedure based on ranges and $F$-ratio’s, respectively, has been presented by Somerville [Som84, Som85a, Som85b]. Monte Carlo methods were used to obtain estimates of the expected subset size. There is some evidence, based on computational results, that the procedure is at least as efficient as Gupta’s procedure in terms of expected subset size. Generalizations and modifications of the approaches of Bechhofer and Gupta, respectively, can be found in Gupta and Panchapakesan [GuPa79]. An overview of the literature is given by Van der Laan and Verdooren [Van89]. The results in this paper cannot be used in a sequential selection scheme. For sequential selection of the normal population with the largest mean some useful results have been derived by Paulson [Paul64], with a slight modification presented by Coolen [Cool95]. An overview of sequential selection procedures is given by Wetherill and Glazebrook [WeGl86, sect.12.5].

Using the indifference zone approach we have

$$P(FS|R_B) = 1 - P(CS|R_B),$$

where $FS$ indicates false selection, that is selection of a non-best population. A rather straightforward generalization of the standard indifference zone approach is derived by selecting the population with the largest sample mean only if there is some specified minimal difference to the second-largest sample mean. This has been discussed by several authors, however without considering the design of the selection experiment. We refer to Bofinger [Bofi94] for related work and some additional references. Gutmann and Maymin [GuMa87] use a similar selection rule, but their results are affected by a slight difference in the criterion for a population to be best. Chen [Chen85] also discusses selection of the population with the largest sample mean only if there is some specified minimal difference to the second-largest sample mean, also considering design issues. Chen explicitly combines selection with a preliminary test of a null hypothesis that $\mu[k-1] = \mu[k]$, and if this hypothesis is not rejected Chen aims at not making a selection, for further discussion of Chen’s method see the discussion added to his paper [Chen85].

Although this generalized selection rule is intuitively logical, so far no fundamental arguments for it have been provided. In section 2 we use a concept of imprecise previsions to model preference, and we show how this leads to this generalized selection rule. This concept

3
is based on Bayesian foundations. In section 3 we explicitly aim at the use of this rule to design a selection experiment, generalizing Bechhofer's procedure. This leads to three possible outcomes of the experiment: $CS, FS$ and non-selection. A consequence of such an approach is that the model includes the possibility of no selection after an experiment, at least not based on strong preference of a single population. An advantage of this generalization is the greater flexibility of the selection procedure. To have Bechhofer's procedure as a special case of the generalized procedure, we start with a similar probability requirement for $CS$. However, next to the common sample size $n$ from each population, the design of the selection experiment needs a specified value for the minimal distance between the largest and second-largest sample means. We call this minimal distance the *preference threshold*, which fits well into the terminology in our derivation of the selection rule in section 2; Bofinger [Bofi94] calls it the 'least significant spacing'. Hence, a second requirement needs to be specified in the design phase. In section 3 we propose one possible second requirement, in terms of the probability of false selection $P(FS)$ for a parameter configuration that is directly related to the minimal probability of correct selection. In section 4 an example of a possible application is given. In section 5 we return to the concepts presented in section 2, which enable us to gain more insight into the selection criterion. At the end of the paper a table is provided giving selection constants for the method proposed in section 3.
2 Modelling preference by imprecise previsions

The concept of imprecise previsions [Wall91] provides a powerful expression of preference amongst random variables. Imprecise previsions are strongly related to previsions [DeFi74], and if the probability distribution of a random variable is known, its prevision is equal to the mean. If imprecise previsions are related analogously to imprecise probability distributions for a random variable, modelled by a set of probability distributions, the lower prevision of the random variable is equal to the infimum of the set of the mean values according to these distributions. The concept of imprecise previsions is more general than that of imprecise probabilities, since probabilities are just previsions for events, and we denote the lower prevision for a random variable $X$ by $\underline{E}(X)$, and the upper prevision for $X$ by $\overline{E}(X)$. Quite informally in the subjectivists language, $\underline{E}(X)$ can be interpreted as the supremum of all prices one is willing to pay for the yet uncertain reward $X$, whereas $\overline{E}(X)$ can be regarded as the infimum of all prices for which one would be willing to sell $X$ (where the price and reward should be expressed in the same unit of utility, see Walley [Wall91] for a comprehensive presentation and discussion of this concept). One is said to prefer $X_i$ to $X_j$ if $\underline{E}(X_i - X_j) > 0$, which can be interpreted as that you find it desirable to buy $X_i - X_j$ even for a positive price $\epsilon$, where $0 < \epsilon < \underline{E}(X_i - X_j)$. Since [Wall91] we can restrict the discussion to lower previsions, where it is to be noted that

$$\underline{E}(X_i - X_j) = -\overline{E}(X_j - X_i),$$

implies that no preference is explicitly expressed. Therefore, the preference as modelled by the lower prevision, used throughout this paper, could be called 'strong preference'. Obviously, this does not exclude some weaker form of preference one may have for the population related to the maximum sample mean, even if there is no strong preference based on the available information. If one is not used to two different concepts of preference, strong and weak, think about the difference between being allowed to make a choice (corresponding to strong preference) and being forced to make a choice (corresponding to weak preference).

In this section we show how a simple model for imprecise probabilities with updating in the Bayesian context [PeWa91] leads to selection rule (5), providing a foundational argument for this selection rule that has been missing in the related literature. Let $X_1$ and $X_2$ be independent normally distributed random variables, with means $\mu_1$ and $\mu_2$ respectively, and both with variance $\sigma^2$. Hence $D = X_1 - X_2 \sim N(\mu_d, \sigma_d^2)$, with $\mu_d = \mu_1 - \mu_2$ and $\sigma_d^2 = 2\sigma^2$. In a Bayesian framework with imprecise priors, a model with the prior class consisting of conjugate priors [PeWa91, sect.3.3] is mathematically attractive since updating is very easy. Assuming $\sigma^2$ to be known, we define a class $M_{\mu_d}^0$ of prior distributions for $\mu_d$, with $f_{\tau,\nu}$ denoting the probability density function of $N(\frac{\tau}{\nu}, \frac{\sigma^2}{\nu^2})$, by the corresponding class of probability density functions

$$M_{\mu_d}^0(c, \nu) = \{ f_{\tau,\nu} | -c \leq \tau \leq c \},$$

where $0 < \nu < \infty$, $0 \leq c < \infty$. If a random variable $X$ has a known probability distribution, the prevision for $X$ can logically be regarded to be equal to the mean value of $X$ [DeFi74,
and by generalization to a known class of prior probability distributions (1) for \( \mu_d \) the lower prior prevision for \( \mu_d \) equals the infimum of all means with respect to the distributions in this class, hence

\[
E(\mu_d|c, \nu) = \frac{-c}{\nu}. \tag{2}
\]

Analogously, the upper prevision for \( \mu_d \) related to the class (1) is \( \overline{E}(\mu_d|c, \nu) = \frac{\xi}{\nu} \). Suppose \( n \) independent observations from both \( X_1 \) and \( X_2 \), say \( x_{1,j} \) and \( x_{2,j} \) for \( j = 1, \ldots, n \), become available, so \( n \) independent realisations of the random variable \( D \), say \( d_j = x_{1,j} - x_{2,j} \), can be used to update the prior class (1). It is easily seen that the posterior class \( M_n^{\mu_d} \) is

\[
M_n^{\mu_d}(c, \nu, \sum_{j=1}^{n} d_j, n) = \{ f_{r, \nu+n} \mid -c + \sum_{j=1}^{n} d_j \leq r \leq c + \sum_{j=1}^{n} d_j \}. \tag{3}
\]

The lower posterior prevision, related to the posterior class (3) is equal to

\[
\underline{E}(\mu_d|c, \nu, \sum_{j=1}^{n} d_j, n) = \frac{-c + \sum_{j=1}^{n} d_j}{\nu + n}. \tag{4}
\]

Again, the upper posterior prevision

\[
\overline{E}(\mu_d|c, \nu, \sum_{j=1}^{n} d_j, n) = \frac{c + \sum_{j=1}^{n} d_j}{\nu + n}
\]

does not add much to our presentation.

Using lower previsions to express preferences, the first population (related to \( X_1 \)) is preferred to the second (related to \( X_2 \)) if the lower prevision for \( \mu_d \) is greater than 0. The second population is preferred to the first if the upper prevision for \( \mu_d \) is less than zero. Because our prior class (1) is symmetric around 0, equation (2) implies that we do not explicitly prefer either population at the prior stage. In the posterior stage, however, the first population is preferred to the second if \( \sum_{j=1}^{n} d_j > c \), as follows from equation (4). The second population is preferred if \( \sum_{j=1}^{n} d_j < -c \). This leads to the selection rule (5), so an argument in favour of this selection rule is provided by the theory of imprecise previsions, which is an attractive concept for expressing preferences. However, it is not said that the mathematically attractive class of conjugate priors (1) is the one and only right model to be used, leaving an interesting subject for research. In section 5 we will briefly return to the concept used in this section, and we will discuss the role of \( \nu \), giving us more insight into the selection criterion via robustness arguments related to the prior in the Bayesian framework [Berg90].
3 Generalizing indifference zone selection

We present a generalization of the indifference zone approach to design selection experiments, as before we regard selection of the best of $k (\geq 2)$ normal populations with equal known variance, where 'best' relates to maximum population mean, with the $k$ independent populations denoted by $\pi_i, i = 1, \ldots, k$. Let $\bar{Y}_i$ denote the mean of a sample of $n_i (\geq 1)$ independent observations from $\pi_i$, with each single observation from population $\pi_i$ denoted by $Y_{ij}, j = 1, \ldots, n_i$, so $Y_{ij} \sim N(\mu_i, \sigma^2)$, with $\sigma^2 > 0$ assumed to be known. The ordered population means are $\mu[1] \leq \mu[2] \leq \ldots \leq \mu[k]$, and $\mu[i]$ is the mean of population $\pi(i)$. Obviously, the ordering of the $\mu_i$ is unknown, and we also need the notation $Y_{(i)j}$ and $\bar{Y}_{(i)}$ in relation to the observations from population $\pi(i)$. The goal of the selection experiment is selection of $\pi(k)$.

Following Bechhofer's indifference zone approach [Bech54], we introduce the parameter space $\Omega = \{\mu : \mu = (\mu_1, \ldots, \mu_k), \mu_i \in \mathbb{R}, i = 1, \ldots, k\}$. A subspace of $\Omega$ is defined by (with $\delta^* > 0$)

$$\Omega(\delta^*) = \{\mu \in \Omega : \mu[k] - \mu[k-1] \geq \delta^*\}.$$ 

The probability requirement in Bechhofer's indifference zone approach is that the probability of correct selection, $P(CS)$, is at least $P^*$ for $\mu \in \Omega(\delta^*)$, where $\delta^*$ and $P^* \in (\frac{1}{k}, 1)$ are to be specified. Assuming $n_i = n (\geq 1)$ for all $i = 1, \ldots, k$ the problem is to determine the smallest common sample size $n$ for which the following probability requirement holds

$$\inf_{\Omega(\delta^*)} P(CS) \geq P^*.$$ 

This condition is called the $P^*$-condition for the probability requirement of correct selection, $\Omega(\delta^*)$ is called the preference zone and $\Omega^c(\delta^*)$ is called the indifference zone. We apply the following selection rule $R_e$, as derived from the arguments in section 2: after $n$ observations from each population select $\pi_i$ if and only if

$$\sum_{j=1}^{n} Y_{ij} - \sum_{j=1}^{n} Y_{ij} > c$$ 

for $l = 1, \ldots, k; l \neq i$, with $c \geq 0$. Remark that $c = 0$ leads to the standard indifference zone approach. We refer to the constant $c$ as the preference threshold.

The infimum of $P(CS|R_e)$ over $\Omega(\delta^*)$ is attained for the so-called Least Favourable Configuration (LFC) in $\Omega(\delta^*)$ ([Bech54], see also Theorem 2 of [Chen85]), given by

$$\mu[1] = \ldots = \mu[k-1] = \mu[k] - \delta^*.$$ 

For the LFC the probability of correct selection is

$$P(\sum_{j=1}^{n} Y_{(k)j} - \sum_{j=1}^{n} Y_{(i)j} > c; i = 1, \ldots, k - 1 \mid LFC) =$$

$$P(\sum_{j=1}^{n} Y_{(i)} + c < n\bar{Y}_{(k)}; i = 1, \ldots, k - 1 \mid LFC) =$$

$$\int_{-\infty}^{\infty} \Phi^{k-1} \left(\frac{y-c-n\mu[1]}{\sqrt{n_\sigma}} \middle| \mu[1] = \mu[k] - \delta^*, \mu[1] = \mu[k] - \delta^*\right) d\Phi \left(\frac{y-n\mu[k]}{\sqrt{n_\sigma}}\right) =$$

$$\int_{-\infty}^{\infty} \Phi^{k-1} (z + c, k) d\Phi (z)$$
with
\[ \tau_{c,k} = \frac{n\delta^* - c}{\sqrt{n}\sigma}, \]
leading to \( P^* \)-condition
\[ \int_{-\infty}^{\infty} \Phi^{-1}(z + \tau_{c,k})d\Phi(z) = P^*. \] (7)

The results (6) and (7) were derived by Chen [Chen85, Th.2], but are given here for self-containment of this paper. The probability requirement (7) is used to determine \( \tau_{c,k} \) numerically (in our notation \( \tau_{c,k} \) the first index \( c \) refers to 'Correct selection', use of the more common notation \( \tau_{P^*,k} \) will be restricted to the standard situation). Equation (6) gives a relation between \( c \) and \( n \)
\[ c = n\delta^* - \tau_{c,k}\sqrt{n}\sigma. \] (8)

Although we will not explicitly use \( n \) as a function of \( c \), which would seem to suggest some sequential selection procedure, such a relation also follows easily from (6),
\[ n = \frac{1}{4}\delta^* - 2 \left\{ \tau_{c,k}\sigma + (\tau_{c,k}\sigma^2 + 4\delta^*c)^{-\frac{1}{2}} \right\}^2. \] (9)

To design the selection experiment we need to determine appropriate values of \( c \) and \( n \). So next to the probability requirement for \( P(CS|R_c) \), leading to relation (8), a second requirement is needed. This aspect has been discussed by Chen [Chen85] in an attempt to combine selection of the population with the largest population mean with testing of a null-hypothesis that \( \mu_{[k-1]} = \mu_{[k]} \), keeping the analogues of the power and the level of a test under control simultaneously. Designing selection experiments by determining appropriate values of \( c \) and \( n \) also seems to provide a nice topic for future research, especially since it may provide an opportunity to combine features of the indifference zone approach with properties of other approaches, for example Gupta's subset selection method (at the end of this section we briefly relate Gupta's method to our approach). Some related papers discuss the important but difficult problem about what actually can be said about a population once this is selected: is it actually the best population or not. Interesting contributions to this research are by Bofinger [Bofi85, Bofi86, Bofi94] and Gutmann and Maymin [GuMa87]. Their results discourage the use of a probability statement for an actually selected population as a second requirement in our approach.

Since the design of a selection experiment according to Bechhofer’s approach only consists of determining \( n \), the minimal sample size for each population, related to chosen values of \( \delta^* \) and \( P^* \), no second probability requirement has been suggested yet, except for the approach of Chen [Chen85] where the second requirement is directly related to testing a homogeneity hypothesis. For our second probability requirement we consider the probability of false selection when applying selection rule \( R_c \), that is \( P(FS|R_c) \). In the standard indifference zone approach of Bechhofer we have \( P(FS|R_B) = 1 - P(CS|R_B) \), whereas application of selection rule \( R_c \), given by (5), leads to \( P(CS|R_c) + P(FS|R_c) \leq 1 \), with \( P(No\ Selection|R_c) = 1 - P(CS|R_c) - P(FS|R_c) \). It would be interesting to achieve an upper bound for \( P(FS|R_c) \). Unfortunately, \( P(FS|R_c) \) does not have a nice monotonic behaviour over \( \Omega(\delta^*) \). However, special attention to the LFC \( \mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta^* \) is justified by the results in Lemma 1 and Lemma 2. All the proofs are given in the appendix.
Lemma 1 \( \mu_{[k-1]} \) should be as large as possible to maximize \( P(FS) \), and \( P(FS) \) has a local maximum over \( \Omega(\delta^*) \) at the LFC \( \mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} = \delta^* \).

Lemma 2 is a generalization of Lemma 4.1 of Stefansson, Kim and Hsu [Ste88], which implies that the probability of selection for a fixed selection rule \( R_c \), so \( P(CS|R_c) + P(FS|R_c) \), is minimal over \( \Omega(\delta^*) \) at the LFC.

Lemma 2 Suppose that the joint density of \( Y_1, \ldots, Y_k \) is \( \prod_{i=1}^{k} f(y_i - \theta_i) \), where \( f(y - \theta) \) has monotone likelihood ratio in \( y \) and \( \theta \). Let \( \theta_{[1]} \leq \ldots \leq \theta_{[k]} \) be the ordered values of \( \theta_i, i = 1, \ldots, k \), and assume that \( \theta_{[1]} = \ldots = \theta_{[l]} =: \lambda \) for some \( l \in \{1, \ldots, k-1\} \). Then, for any fixed \( t > 0 \), the function \( \Psi(\lambda) := P_\theta(Y_{k:k} - Y_{k-1:k} \geq t) \) is monotone nonincreasing in \( \lambda \) over \( (-\infty, \theta_{[l+1]}) \), where \( Y_{1:k} \) denotes the \( i \)-th order statistic of \( Y_1, \ldots, Y_k \).

Note that Lemma 2 with \( l = 1 \) yields Lemma 4.1 of Stefansson, Kim and Hsu [Ste88].

Since our first probability requirement is equivalent to \( P(CS|R_c, LFC) = P^* \), we get the minimum probability of selection over \( \Omega(\delta^*) \) equal to \( P^* + Q^* \) by the second probability requirement

\[
P(FS|R_c, LFC) = Q^*, \quad Q^* \in (0, 1 - P^*].
\]

We call this second requirement the \( LFC - Q^* \)-condition. These two probability requirements imply that the probability of no selection does not exceed \( 1 - P^* - Q^* \) over \( \Omega(\delta^*) \), with this value attained at the LFC. Given \( R_c \), at the LFC the probability of false selection is

\[
\sum_{i=1}^{k-1} P(n\bar{Y}_{(i)} > n\bar{Y}_{(i)} + c; l = 1, \ldots, k; l \neq i \mid LFC) =
\sum_{i=1}^{k-1} \int_{-\infty}^{\infty} \Phi^{k-2} \left( \frac{y - c - n\mu_{[i]}}{\sqrt{n}\sigma} \right) \Phi \left( \frac{y - c - n\mu_{[i]}}{\sqrt{n}\sigma} \mid \mu_{[i]} = \mu_{[1]} + \delta^* \right) d\Phi \left( \frac{y - n\mu_{[i]}}{\sqrt{n}\sigma} \right) =
\sum_{i=1}^{k-1} \int_{-\infty}^{\infty} \Phi^{k-2} \left( z - \frac{c + n\delta^*}{\sqrt{n}\sigma} \right) \Phi \left( z - \frac{c + n\delta^*}{\sqrt{n}\sigma} \right) d\Phi(z) =
(k - 1) \int_{-\infty}^{\infty} \Phi^{k-2} \left( z - \frac{\tau_{f,k} - \tau_{c,k}}{2} \right) \Phi(z - \tau_{f,k})d\Phi(z)
\]

with

\[
\tau_{f,k} = \frac{n\delta^* + c}{\sqrt{n}\sigma}.
\]

This leads to the \( LFC - Q^* \)-condition

\[
(k - 1) \int_{-\infty}^{\infty} \Phi^{k-2} \left( z - \frac{\tau_{f,k} - \tau_{c,k}}{2} \right) \Phi(z - \tau_{f,k})d\Phi(z) = Q^*.
\]

Since \( \tau_{c,k} \) has already been determined, \( \tau_{f,k} \) (with index \( f \) related to 'False selection') follows from (11), and equation (10) gives the second relation between \( c \) and \( n \)

\[
c = -n\delta^* + \tau_{f,k}\sqrt{n}\sigma.
\]
Relations (8) and (12) lead to
\[ n = \left( \frac{\sigma(\tau_{c,k} + \tau_{f,k})}{2\delta^*} \right)^2 \]  
and
\[ c = \frac{\sigma^2}{4\delta^*} \left( \tau_{f,k}^2 - \tau_{c,k}^2 \right). \]  

These \( n \) and \( c \) are such that both probability requirements are satisfied for given \( \delta^* \), \( P^* \) and \( Q^* \), when using selection rule \( R_c \). Of course, the number of observations from each population should be equal to the smallest integer not less than the real-valued \( n \) of (13). Table 1, at the end of this paper, provides the selection constants \( \tau_{c,k}, \tau_{f,k} \) for several values of \( P^*, Q^*, k \).

Our approach is equal to the standard indifference zone procedure if and only if \( Q^* = 1 - P^* \), and it is easy to show that
\[ Q^* = 1 - P^* \iff c = 0 \iff \tau_{c,k} = \tau_{f,k}, \]  
with otherwise \( Q^* < 1 - P^*, c > 0 \) and \( \tau_{c,k} < \tau_{f,k} \).

We compare our approach to the standard indifference zone selection procedure, where after \( n \) observations from each population, that population is preferred which relates to the greatest observed mean value, with \( n \) again the smallest common sample size such that
\[ \inf_{\Omega(\delta^*)} P(CS) \geq \tilde{P}, \]  
with \( \tilde{P} \in (\frac{1}{2}, 1) \). Of course, the cases \( \tilde{P} = P^* \) and \( \tilde{P} = 1 - Q^* \) are of special interest. We denote the number of observations per population in the standard approach with \( \tilde{n} \) by \( n_{\tilde{P}} \). To fulfill the probability requirement, the probability of correct selection is derived analogously to equation (7) with \( c = 0 \), leading to
\[ \int_{-\infty}^{\infty} \Phi^{k-1}(z + \tau_{\tilde{P},k})d\Phi(z) = \tilde{P}, \]  
with
\[ \tau_{\tilde{P},k} = \frac{\sqrt{n_{\tilde{P}}\delta^*}}{\sigma}. \]  

After calculating \( \tau_{\tilde{P},k} \) by equation (16), relation (17) gives
\[ n_{\tilde{P}} = \left( \frac{\sigma\tau_{\tilde{P},k}}{\delta^*} \right)^2. \]  

Lemma 3 states a relation between \( n \) according to (13) and \( n_{P^*}, n_{1-Q^*} \), both according to (18) for \( \tilde{P} = P^* \) and \( \tilde{P} = 1 - Q^* \), respectively. The proof is given in the appendix.
Lemma 3

\[
Q^* < 1 - P^* \iff n_P^* < n < \frac{n_P^* + n_1 - Q^*}{2},
\]

\[
Q^* = 1 - P^* \iff n_P^* = n = n_1 - Q^*.
\]

The standard indifference zone procedure with \( n \) according to (13) would lead to

\[
\inf_{\Omega^{(\delta^*)}} P(CS) = \int_{-\infty}^{\infty} \Phi^{k-1} \left( z + \frac{\tau_{c,k} + \tau_{f,k}}{2} \right) d\Phi(z),
\]

which is, if \( Q^* < 1 - P^* \), greater than

\[
\int_{-\infty}^{\infty} \Phi^{k-1}(z + \tau_{c,k}) d\Phi(z) = P^*;
\]

and less than

\[
\int_{-\infty}^{\infty} \Phi^{k-1}(z + \tau_{f,k}) d\Phi(z) < 1 - Q^*.
\]

Hence, if you are, anyhow, forced to select a single population after the \( n \) observations from each population, even if you do not have a strong preference according to selection rule (5) with \( c > 0 \), then it is obviously best to choose that population which relates to the largest observed mean, and you will be on the safe side as the probability of a correct selection is greater than \( P^* \). This could be called a weak preference.

Finally, it is of interest to compare our procedure to Gupta’s subset selection approach [Gupta56, Gupta65]. Gupta’s approach gives false selection if the best population is not in the selected subset, which occurs if

\[
\max_{1 \leq i \leq k} \sum_{j=1}^{n} Y_{ij} - \sum_{j=1}^{n} Y_{(k)j} > \tau \sigma \sqrt{n},
\]

where \( Y_{(k)j} \) denotes the \( j \)-th observation form the best population. This corresponds to our approach in that there is strong preference for the population corresponding to \( \max_{1 \leq i \leq k} \sum_{j=1}^{n} Y_{ij} \) when compared, by regarding the sample statistics, to population \( \pi_{(k)} \). However, our approach is more closely related to Bechhofer’s indifference zone selection in that it explicitly aims at selection of a single population, and leads to the necessary minimum sample size. If the outcome of an experiment set up according to our approach is ‘no selection’, this corresponds to a subset which contains more than one population in Gupta’s approach. It may be possible to replace our second probability requirement by one closely related to Gupta’s approach, in an attempt to combine both classical procedures, leaving an interesting topic for future research.
4 Application

Let us now consider the following situation ([Van90]). In Papua New Guinea, oil palm cultivation started on a commercial scale in 1968. In 1976, about 12,000 ha were planted. To guide the oil palm cultivation the Dami Oil Palm Research Station has been founded at Kimbe, West New Britain, Papua New Guinea. At this station a \textit{dura} \times \textit{pisifera} progeny trial has been started in 1968. In this experiment nine ex-AVROS \textit{pisifera} with four selected Deli \textit{dura} palms have been crossed to get 15 families. These fifteen families were arranged in 5 randomized complete blocks with sixteen (4 \times 4) palms per plot with a 9 m triangular spacing.

For this example we have taken only ten families which remain in four complete blocks, the other families were discarded in several blocks due to diseases. The average fresh fruit bunch yield \( y \) (in kg/palm) over the years 1972-1977 of the four inner palms per plot has been analysed. Further, samples of Leaf 17 (this in year 1973) were taken from all inner palms, bulked per plot, and analysed at Banting Oil Palm Research Station (O.P.R.S.), former Harrisons Crosfield Research Station in Malaysia.

The percentage Magnesium content \( x \) (in \% Mg) in Leaf 17 has been determined. It turns out [Breu87] that this \% Mg has a good correlation \(( r = 0.70)\) with the yield of oil for the first 5 years of production (1972-1976); hence the \% Mg can be used to indicate good families for oil yield. Because the \% Mg determination has been done with the same procedure in Banting O.P.R.S. the standard deviation \( \sigma \) of the \% Mg determination is known to be 0.0186. The following average \% Mg for the 10 families over the 4 blocks has been found.

<table>
<thead>
<tr>
<th>family</th>
<th>( \bar{x}_i )</th>
<th>rank number</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i = 1 )</td>
<td>0.212</td>
<td>[5]</td>
</tr>
<tr>
<td>2</td>
<td>0.222</td>
<td>[7]</td>
</tr>
<tr>
<td>3</td>
<td>0.242</td>
<td>[8]</td>
</tr>
<tr>
<td>4</td>
<td>0.204</td>
<td>[3]</td>
</tr>
<tr>
<td>5</td>
<td>0.210</td>
<td>[4]</td>
</tr>
<tr>
<td>6</td>
<td>0.186</td>
<td>[2]</td>
</tr>
<tr>
<td>7</td>
<td>0.218</td>
<td>[6]</td>
</tr>
<tr>
<td>8</td>
<td>0.244</td>
<td>[9]</td>
</tr>
<tr>
<td>9</td>
<td>0.162</td>
<td>[1]</td>
</tr>
<tr>
<td>10</td>
<td>0.248</td>
<td>[10]</td>
</tr>
</tbody>
</table>

When we apply Bechhofer's selection procedure with \( P^* = 0.90 \) and \( n = 4 \), we find from table 1 the selection constant for \( k = 10 \), namely \( \tau_{0.90,10} = 2.98293 \) and thus

\[
\delta^* = \frac{0.0186}{\sqrt{4}} \cdot 2.98293 = 0.028.
\]

If we apply Bechhofer's procedure before the start of the experiment, we can determine the required minimum common sample size for \( P^* = 0.90 \) and \( \delta^* = 0.01 \). The result is

\[
n = \left( \frac{0.0186}{0.01} \cdot 2.98293 \right)^2 = 31 ,
\]
rounded to the nearest (larger) integer. For $\delta^* = 0.02$ the result is: $n = 8$. If we want to apply our procedure with second probability requirement $P(\overline{S}|R_c, LFC) = .05$ for $\Omega(0.01)$, then a selection experiment would be needed with common sample size

$$n = \left( \frac{0.0186 (2.98293 + 3.35582)}{2 \times 0.01} \right)^2 = 35,$$

rounded to the nearest (larger) integer, and selection rule $R_c$ with

$$c = \frac{0.0186^2}{4 \times 0.01} \left( 3.35582^2 - 2.98293^2 \right) = 0.020.$$

Now we want to indicate the best family with a minimum probability of correct selection $P^*$ for the average fresh fruit bunch yield $y$. The analysis of variance procedure gives mean square error of 6308.488 based on 27 degrees of freedom. For yield $y_{ij}$ of family $i$ in block $j$ we use the following model

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \text{ with } i = 1, 2, \cdots, 10 \text{ and } j = 1, 2, 3, 4.$$

The least squares mean $\bar{y}_i = \mu + \alpha_i + \bar{\beta}$ for the family $V_i$ is estimated by the family $V_i$ mean $\bar{y}_i$. The results were as follows:

<table>
<thead>
<tr>
<th>family $V_i$</th>
<th>$y_i$ kg/palm</th>
<th>rank number</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>999.97</td>
<td>[10]</td>
</tr>
<tr>
<td>2</td>
<td>890.50</td>
<td>[6]</td>
</tr>
<tr>
<td>3</td>
<td>927.24</td>
<td>[9]</td>
</tr>
<tr>
<td>4</td>
<td>888.60</td>
<td>[5]</td>
</tr>
<tr>
<td>5</td>
<td>697.90</td>
<td>[1]</td>
</tr>
<tr>
<td>6</td>
<td>745.83</td>
<td>[2]</td>
</tr>
<tr>
<td>7</td>
<td>863.19</td>
<td>[4]</td>
</tr>
<tr>
<td>8</td>
<td>916.06</td>
<td>[8]</td>
</tr>
<tr>
<td>9</td>
<td>915.62</td>
<td>[7]</td>
</tr>
<tr>
<td>10</td>
<td>854.94</td>
<td>[3]</td>
</tr>
</tbody>
</table>

Bechhofer’s selection procedure selects $V_1$. With $P^* = 0.90$ and $n = 4$, one finds

$$\delta^* = \left( \frac{6308.488}{4} \right)^{\frac{1}{2}} 2.98293 = 118.46.$$

For $P^* = 0.90$ and $\delta^* = 70$ the required minimum common sample size is

$$n = 6308.488 \left( \frac{2.98293}{70} \right)^2 = 12,$$

rounded to the nearest (larger) integer.

If, in our approach, we require $P(\overline{S}|R_c, LFC) = .01$ for $\Omega(70)$, then

$$n = 6308.488 \left( \frac{2.98293 + 4.11764}{2 \times 70} \right)^2 = 17,$$

rounded to the nearest (larger) integer, is needed, together with selection rule $R_c$ where

$$c = \frac{6308.488}{4 \times 70} \left( 4.11764^2 - 2.98293^2 \right) = 181.52.$$
5 Further analysis via imprecise previsions

We conclude the paper with a discussion of the remarkable role of the parameter $\nu$ in the classes (1) and (3). With preference expressed by

$$E(\mu_d|c, \nu, \sum_{j=1}^{n} d_j, n) > 0,$$

the value of $\nu$ does not seem to play any role for the selection procedure. One option is to set $\nu = 1$ and hence delete it in the notation used so far. But the presence of the hyperparameter $\nu$ enables us to choose the prior variance unequal to $\sigma_d^2$, and to analyse how this prior variance influences the selection scheme. Suppose we want all priors for $\mu_d$ in the class (1) to have variance $\sigma_0^2 > 0$. This is achieved by

$$\nu = \frac{\sigma_d^2}{\sigma_0^2}.$$

If we want the means of the prior class in between $\tau_l = -\tau_r < 0$ and $\tau_r > 0$, then we should set

$$c = \tau_r \nu = \sigma_d^2 \frac{\tau_r}{\sigma_0^2}.$$

Hence, population 1 is preferred to population 2 if

$$\sum_{j=1}^{n} d_j > \sigma_d^2 \frac{\tau_r}{\sigma_0^2}.$$

It may be attractive to reflect lack of prior knowledge in the model by choosing large bounds for the prior mean and/or high prior variance. If we let $\tau_r \to \infty$ and $\sigma_0^2 \to \infty$, three different cases may appear. If $\frac{\tau_r}{\sigma_0^2}$ is a constant, then the selection criterion is not influenced, and this directly relates to $\nu \downarrow 0$ and $c > 0$ a constant as before. If $\frac{\tau_r}{\sigma_0^2} \to \infty$, then $c \to \infty$ and no population will ever be preferred. Finally, if $\frac{\tau_r}{\sigma_0^2} \downarrow 0$, then we arrive at the standard indifference zone criterion of simply selecting the population with the greatest observed mean, since $c \downarrow 0$.

The influence of the hyperparameter $\nu$ can also be analysed through the imprecise predictive distributions for future observations of $D$, corresponding to the classes (1) and (3) for $\mu_d$. The prior class of imprecise predictive distributions, related to (1), is equal to

$$M_{D}^P(c, \nu) = \{\hat{f}_{r,\nu} \mid c \leq \tau \leq c\},$$

with $\hat{f}_{r,\nu}$ denoting the probability density function of $N\left(\frac{\tau}{\nu}, \frac{(\nu+1)\sigma_d^2}{\nu}\right)$. The lower prior predictive prevision for $D$ according to (21) is equal to

$$E(D|c, \nu) = \frac{-c}{\nu},$$

which is equal to the lower prior prevision for $\mu_d$. The posterior class of imprecise predictive distributions, related to (3), is equal to

$$M_{D}^{P}(c, \nu) = \{\hat{f}_{r,\nu} \mid c \leq \tau \leq c\},$$

with $\hat{f}_{r,\nu}$ denoting the probability density function of $N\left(\frac{\tau}{\nu}, \frac{(\nu+1)\sigma_d^2}{\nu}\right)$. The lower prior predictive prevision for $D$ according to (21) is equal to

$$E(D|c, \nu) = \frac{-c}{\nu},$$

which is equal to the lower prior prevision for $\mu_d$. The posterior class of imprecise predictive distributions, related to (3), is equal to

14
\[
M^0_D(c, \nu, \sum_{j=1}^n d_j, n) = \{ \bar{f}_{r, \nu+n} - c + \sum_{j=1}^n d_j \leq \tau \leq c + \sum_{j=1}^n d_j \}.
\]

(23)

The lower posterior predictive prevision for \( D \), related to (23), is also equal to the lower posterior prevision for \( \mu_d \), so

\[
\mathbb{E}(D|c, \nu, \sum_{j=1}^n d_j, n) = \frac{-c + \sum_{j=1}^n d_j}{\nu + n}.
\]

(24)

The corresponding upper previsions are easily derived. Regarding \( \nu \to \infty \) for the prior situation, both lower prior previsions (2) and (22) have limit 0, with the variance for the members in the prior class (1) becoming 0, and the variance in the prior predictive class (21) becoming \( \sigma_d^2 \). For the posteriors the same limits are reached if \( \nu \to \infty \), so the data do not have any influence in this limiting case. A more interesting limiting case is \( \nu \downarrow 0 \), leading to lower prior previsions (both for \( \mu_d \) as the predictive for \( D \)) decreasing towards \(-\infty \), and the variances of the members of the prior classes (1) and (21) increasing towards \( \infty \), which indeed seems to reflect prior ignorance, although there is some restriction in the fact that still only normal prior distributions are used, which is important with respect to tail behaviour. For \( \nu \downarrow 0 \) the lower posterior means, again equal for \( \mu_d \) and \( D \) according to (4) and (24), become \( \frac{1}{n}(-c + \sum_{j=1}^n d_j) \), whereas the variance of the members of the posterior class (3) for \( \mu_d \) becomes \( \sigma_d^2 \), and the variance of the members of the posterior predictive class (23) for \( D \) becomes \( \frac{n+1}{n} \sigma_d^2 \). Obviously, whatever \( \nu > 0 \) one chooses, the variance for the members of this posterior predictive class (23) is in between \( \sigma_d^2 \) and \( \frac{n+1}{n} \sigma_d^2 \), which differ only very little for \( n \) not too small.

Acknowledgements

Many thanks are due to Dr. Alessandro Di Bucchianico, Eindhoven University of Technology, for the computation of Table 1. We are indebted to an anonymous referee for important remarks on an earlier version of this paper which led to considerable improvements, as well as for providing the rigorous generalization of Lemma 4.1 of Stefansson, Kim and Hsu [Ste88] as presented in our Lemma 2 and for useful references to related literature.
References


Appendix: Proofs

Proof of Lemma 1
We first prove the second statement of Lemma 1: \( P(FS) \) has a local maximum over \( \Omega(\delta^*) \) at the LFC \( \mu[1] = \ldots = \mu[k-1] = \mu[k] - \delta^* \).

We consider the probability of False Selection,

\[
P(FS) = \sum_{i=1}^{k-1} P(nY(i) > nY(l) + c; l = 1, \ldots, k; l \neq i),
\]

at a general value of the parameter vector \( \mu \in \Omega(\delta^*) \). As before, let the ordered population means be \( \mu[1] \leq \ldots \leq \mu[k-1] \leq \mu[k] - \delta^* \), and define

\[
\begin{align*}
\epsilon_{k-1} &:= \mu[k] - \delta^* - \mu[k-1], \\
\epsilon_j &:= \mu[j+1] - \mu[j], \quad j = 1, \ldots, k - 2,
\end{align*}
\]

so \( \epsilon_j \geq 0, \quad j = 1, \ldots, k - 1 \) and

\[
\mu[l] = \mu[k] - \delta^* - \sum_{j=l}^{k-1} \epsilon_j, \quad l = 1, \ldots, k - 1.
\]

The LFC \( \mu[1] = \ldots = \mu[k-1] = \mu[k] - \delta^* \) is the case with \( \epsilon_j = 0, \quad j = 1, \ldots, k - 1 \).

Generally (for ease of notation we take products over an empty domain equal to 1), the probability of a False Selection can be written as

\[
P(FS) = \sum_{l=1}^{k-1} \int_{-\infty}^{\infty} \left[ \prod_{i=1, i \neq l}^{k} \Phi(z + \frac{n(\mu[l] - \mu[i] - c)}{\sqrt{n\sigma}}) \right] d\Phi(z)
\]

The partial derivative of this \( P(FS) \) with respect to \( \epsilon_j, \quad j = 1, \ldots, k - 1 \), is

\[
\frac{\partial P(FS)}{\partial \epsilon_j} = -\frac{\sqrt{n}}{\sigma} \sum_{l=1}^{k-1} \left\{ \int_{-\infty}^{\infty} \left[ \prod_{i=1, i \neq l}^{k} \Phi(z + \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{l-1} - c)}{\sqrt{n\sigma}}) \right] \right. \\

\left. \times \left[ \prod_{i=1, i \neq l}^{j-1} \Phi(z + \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{l-1} + \epsilon_{j} + c)}{\sqrt{n\sigma}}) \right] \\

\times \Phi(z - \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{j-1} + \epsilon_{j} + \delta^*)}{\sqrt{n\sigma}}) \right\} \phi(z) dz
\]

\[
\left. + \frac{\sqrt{n}}{\sigma} \int_{-\infty}^{\infty} \left[ \prod_{i=1}^{k-1} \Phi(z + \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{l-1} - c)}{\sqrt{n\sigma}}) \right] \right. \\

\left. \times \left[ \prod_{i=1, i \neq l}^{j} \Phi(z + \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{l-1} + \epsilon_{l} + c)}{\sqrt{n\sigma}}) \right] \right. \\

\left. \times \Phi(z - \frac{n(\epsilon_{\ell} + \ldots + \epsilon_{l-1} + \epsilon_{l} + \delta^*)}{\sqrt{n\sigma}}) \right\} \phi(z) dz
\]

In the LFC: \( \epsilon_j = 0, \quad j = 1, \ldots, k - 1 \) this partial derivative becomes

\[
\frac{\partial P(FS)}{\partial \epsilon_j} \bigg|_{\text{LFC}} = -\frac{\sqrt{n}}{\sigma} j \int_{-\infty}^{\infty} \Phi^{k-2}(z - \frac{c}{\sqrt{n\sigma}}) \phi(z - \frac{n\delta^* + c}{\sqrt{n\sigma}}) \phi(z) dz.
\]
Since the $\epsilon_j$, $j = 1, \ldots, k - 1$ are restricted to non-negative values, and the gradient of $P(F_S)$ with respect to $\epsilon := (\epsilon_1, \ldots, \epsilon_{k-1})$ has only negative components, the directional derivative in every reasonable direction out of the LFC is negative, which is a sufficient condition for the LFC to be a local maximum point of $P(F_S)$ restricted to $\Omega(\delta^*)$.

The first statement of Lemma 1: $\mu_{[k-1]}$ should be as large as possible to maximize $P(F_S)$, follows from the partial derivative of $P(F_S)$ with respect to $\epsilon_{k-1}$, as given above at any point in $\Omega(\delta^*)$:

$$\frac{\partial P(F_S)}{\partial \epsilon_{k-1}} = -\frac{\sqrt{2}}{\sigma} \sum_{j=1}^{k-1} \int_{-\infty}^{\infty} \left[ \prod_{i=1}^{k-2} \Phi(z - \frac{n(\epsilon_i + \ldots + \epsilon_{k-1} + \delta^*)}{\sqrt{n\sigma}}) \right] \times \phi(z - \frac{n(\epsilon_i + \ldots + \epsilon_{k-1} + \delta^*) + c}{\sqrt{n\sigma}}) \phi(z) dz.$$ 

At any point in $\Omega(\delta^*)$, so for any possible combination of $\epsilon_j$, $j = 1, \ldots, k - 1$, this partial derivative is negative, which implies that $\epsilon_{k-1} = 0$ is a necessary condition for a global maximum point of $P(F_S)$, so at the maximum point of $P(F_S)$ with $\mu$ restricted to $\Omega(\delta^*)$ we need to have $\mu_{[k-1]} = \mu_{[k]} - \delta^*$. \qed
Proof of Lemma 2
This proof is analogous to the proof of Lemma 4.1 of Stefansson, Kim and Hsu [Ste88]. By symmetry, assume \( \theta_1 = \ldots = \theta_l \equiv \lambda < \theta_{l+1} \leq \ldots \leq \theta_k \). Then, for any \( t > 0 \), we have

\[
\Psi(\lambda) = \sum_{j=1}^{k} \int_{-\infty}^{\infty} \left[ \prod_{i=1, i \neq j}^{k} F(y + \theta_j - \theta_i - t) \right] f(y) dy \\
= l \int_{-\infty}^{\infty} \left[ \prod_{i=1+1}^{k} F(y + \lambda - \theta_i - t) \right] F^{l-1}(y - t) f(y) dy \\
+ \sum_{j=l+1}^{k} \int_{-\infty}^{\infty} \left[ \prod_{i=l+1, i \neq j}^{k} F(y + \theta_j - \theta_i - t) \right] F^{l}(y + \theta_j - \lambda - t) f(y) dy
\]

where \( F \) is the cdf of \( f \). Therefore,

\[
\frac{d}{d\lambda} \Psi(\lambda) \\
= l \int_{-\infty}^{\infty} \sum_{j=l+1}^{k} \left[ \prod_{i=l+1, i \neq j}^{k} F(y + \lambda - \theta_i - t) \right] F^{l-1}(y - t) f(y + \lambda - \theta_j - t) f(y) dy \\
- \sum_{j=l+1}^{k} \int_{-\infty}^{\infty} \left[ \prod_{i=l+1, i \neq j}^{k} F(y + \theta_j - \theta_i - t) \right] \left[ F^{l-1}(y + \theta_j - \lambda - t) f(y + \theta_j - \lambda - t) f(y) dy \right] dy \\
= l \int_{-\infty}^{\infty} \sum_{j=l+1}^{k} \left[ \prod_{i=l+1, i \neq j}^{k} F(w - \theta_i - t) \right] F^{l-1}(w - \lambda - t) f(w - \theta_j - t) f(w - \lambda) dw \\
- l \sum_{j=l+1}^{k} \int_{-\infty}^{\infty} \left[ \prod_{i=l+1, i \neq j}^{k} F(w - \theta_i - t) \right] F^{l-1}(w - \lambda - t) f(w - \theta_j - t) f(w - \lambda) dw \\
= l \sum_{j=l+1}^{k} \int_{-\infty}^{\infty} \left[ \prod_{i=l+1, i \neq j}^{k} F(w - \theta_i - t) \right] F^{l-1}(w - \lambda - t) \times \\
[f(w - \theta_j - t) f(w - \lambda) - f(w - \lambda - t) f(w - \theta_j)] dw
\]

Because \( f \) has monotone likelihood ratio the last expression is nonpositive. Hence Lemma 2 follows. \( \square \)
Proof of Lemma 3

Let $Q^* < 1 - P^*$. This implies that $\tau_{c,k} < \tau_{f,k}$ which together with the fact that $\tau_{P^*,k} = \tau_{c,k}$ implies that $n_{P^*} < n$. A similar relation to (16) for determination of $\tau_{1-Q^*,k}$ is

$$(k - 1) \int_{-\infty}^{\infty} \Phi^{k-2}(z) \Phi(z - \tau_{1-Q^*,k}) d\Phi(z) = Q^*.$$ 

Comparing this to relation (11),

$$(k - 1) \int_{-\infty}^{\infty} \Phi^{k-2} \left( z - \frac{\tau_{f,k} - \tau_{c,k}}{2} \right) \Phi(z - \tau_{f,k}) d\Phi(z) = Q^*,$$

and using $\tau_{f,k} > \tau_{c,k}$ together with the strict monotonicity of $\Phi$, it follows that $\tau_{f,k} < \tau_{1-Q^*,k}$.

Together with $2\tau_{c,k} \tau_{f,k} < \tau_{c,k}^2 + \tau_{f,k}^2$ this leads to $n < \frac{n_{P^*} + n_{1-Q^*}}{2}$.

The proof of $n_{P^*} < n < \frac{n_{P^*} + n_{1-Q^*}}{2} \Rightarrow Q^* < 1 - P^*$, as well as the proof of (20) are now straightforward.

$\square$
<table>
<thead>
<tr>
<th>$k$</th>
<th>.80;10</th>
<th>.80;05</th>
<th>.80;01</th>
<th>.90;05</th>
<th>.90;01</th>
<th>.95;01</th>
<th>.95;005</th>
<th>.99;005</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.19023</td>
<td>1.19023</td>
<td>1.19023</td>
<td>1.81238</td>
<td>1.81238</td>
<td>2.32618</td>
<td>2.32618</td>
<td>3.28995</td>
</tr>
<tr>
<td></td>
<td>1.81238</td>
<td>2.32618</td>
<td>2.32618</td>
<td>3.28995</td>
<td>3.28995</td>
<td>3.64277</td>
<td>3.64277</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.65241</td>
<td>1.65241</td>
<td>1.65241</td>
<td>2.23020</td>
<td>2.23020</td>
<td>2.71011</td>
<td>2.71011</td>
<td>3.61730</td>
</tr>
<tr>
<td>4</td>
<td>1.89317</td>
<td>1.89317</td>
<td>1.89317</td>
<td>2.45157</td>
<td>2.45157</td>
<td>2.91623</td>
<td>2.91623</td>
<td>3.79694</td>
</tr>
<tr>
<td></td>
<td>2.38651</td>
<td>2.81929</td>
<td>2.81929</td>
<td>3.72416</td>
<td>3.72416</td>
<td>4.07434</td>
<td>4.11297</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.05280</td>
<td>2.05280</td>
<td>2.05280</td>
<td>2.59971</td>
<td>2.59971</td>
<td>3.05517</td>
<td>3.05517</td>
<td>3.91958</td>
</tr>
<tr>
<td></td>
<td>2.52317</td>
<td>2.94000</td>
<td>2.94000</td>
<td>3.83125</td>
<td>3.83125</td>
<td>4.18050</td>
<td>4.22779</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.17087</td>
<td>2.17087</td>
<td>2.17087</td>
<td>2.70996</td>
<td>2.70996</td>
<td>3.15909</td>
<td>3.15909</td>
<td>4.01209</td>
</tr>
<tr>
<td></td>
<td>2.62542</td>
<td>3.03099</td>
<td>3.03099</td>
<td>3.91301</td>
<td>3.91301</td>
<td>4.26085</td>
<td>4.31458</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.26391</td>
<td>2.26391</td>
<td>2.26391</td>
<td>2.79722</td>
<td>2.79722</td>
<td>3.24165</td>
<td>3.24165</td>
<td>4.08605</td>
</tr>
<tr>
<td>8</td>
<td>2.34032</td>
<td>2.34032</td>
<td>2.34032</td>
<td>2.86914</td>
<td>2.86914</td>
<td>3.30988</td>
<td>3.30988</td>
<td>4.14748</td>
</tr>
<tr>
<td>9</td>
<td>2.40493</td>
<td>2.40493</td>
<td>2.40493</td>
<td>2.93012</td>
<td>2.93012</td>
<td>3.36786</td>
<td>3.36786</td>
<td>4.19989</td>
</tr>
<tr>
<td>10</td>
<td>2.46075</td>
<td>2.46075</td>
<td>2.46075</td>
<td>2.98293</td>
<td>2.98293</td>
<td>3.41818</td>
<td>3.41818</td>
<td>4.24553</td>
</tr>
<tr>
<td>15</td>
<td>2.66114</td>
<td>2.66114</td>
<td>2.66114</td>
<td>3.17341</td>
<td>3.17341</td>
<td>3.60037</td>
<td>3.60037</td>
<td>4.42105</td>
</tr>
<tr>
<td>20</td>
<td>2.79209</td>
<td>2.79209</td>
<td>2.79209</td>
<td>3.29860</td>
<td>3.29860</td>
<td>3.72069</td>
<td>3.72069</td>
<td>4.52298</td>
</tr>
<tr>
<td>30</td>
<td>2.96444</td>
<td>2.96444</td>
<td>2.96444</td>
<td>3.46417</td>
<td>3.46417</td>
<td>3.88045</td>
<td>3.88045</td>
<td>4.67140</td>
</tr>
</tbody>
</table>

Table 1: Values of the selection constants $\tau_{c,k}$ and $\tau_{f,k}$. 

22