ESTIMATION OF RANKED PARAMETERS: INSIGHTS AND ADVANCES

BY

EDWARD J. DUDEWICZ

TECHNICAL REPORT NO. 118
OCTOBER 27, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 75-09450
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATION OF RANKED PARAMETERS: INSIGHTS AND ADVANCES

BY

EDWARD J. DUDEWICZ

TECHNICAL REPORT NO. 118
OCTOBER 27, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 75-09450
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATION OF RANKED PARAMETERS: INSIGHTS AND ADVANCES

by Edward J. Dudewicz

Abstract

For many years statistics concerned itself to a large extent with problems in which the basic observations came from one source or population (one-population problems). Two-population problems were well-known (if unsolved, e.g. the Behrens-Fisher problem), but for the most part it was a one-population world until some time in the 1950's. In 1954 R. E. Bechhofer, by pioneering work in ranking and selection, brought the subject to full light of day with a context other than the type described by saying (as in classical ANOVA) "We have k populations, but would like to test the hypothesis that we really only have one." The effects of having k sources on estimation started to be investigated around 1967 in this author's Ph.D. work, and now a large and growing literature explores what has become a separate area of statistical inquiry. In this paper we review the history and applicability of this area from its inception to the present, give new results on ranked parameter estimation in the normal case, note how the results and principles of the area give us new insights on phenomena observed in statistics as early as 1927, and formulate a plan for future advances in the area.

Introduction

For many years, at first for good reasons as the field began to develop and face its first challenges but lately for questionable reasons as (pseudo?) statisticians develop ever more mathematical treatments of problems solved for all practical purposes years ago, statistics concerned itself to a large extent with problems in which the basic observations came from one source or population (we will call these problems one-population problems). Two-population problems were well-known, if unsolved, for example the Behrens-Fisher problem in the area of hypothesis testing, but for the most part it was a one-population world until some time around the 1950's when R. E. Bechhofer and others, by pioneering work in ranking and selection, brought the subject of k-populations to full light of day with a context other than the type described by saying (as in classical ANOVA) "We have k populations, but would like to test the hypothesis that we really only have one."

The effects of having k sources on estimation started to be investigated around 1966-67 in this author's Ph.D. work at Cornell University (though the problem was approached indirectly by others earlier, for example by Fraser (1952), and, in a statement of an open problem, by Bechhofer, Kiefer, and Sobel in earlier versions of their now-famous 1968 research monograph), and now a large and growing literature explores what has become a separate area of statistical inquiry.

\footnote{For example see Bechhofer (1954), which dealt with ranking and selection of normal population means (when variances are equal and known).}
I am happy to be able to use this occasion: to review the history and applicability of this area from its inception 10 years ago to the present including its relation to ranking and selection (Section 1), though I will refrain from presenting a complete review or bibliography, preferring instead to give indicative results and principles with typical sources or references; to give new results on ranked parameter estimation, in the case of the normal distribution, which illustrate these principles in concrete detail (Section 2); to note how the results and principles of the area can be interpreted to give us new insights on seemingly-mysterious phenomena observed in statistics as early as 1927 (Section 3); and to formulate a plan for future advances in the area (Section 4). There is a saying that for many years people were afraid that Congress might let them down, but that now they just wish Congress would let them up; since most readers of this article are seated, I hope that when they finish it they will have been let up!
1. **History and Applicability**

Let \( \pi_1, \ldots, \pi_k \) be \( k \) given populations such that observations from \( \pi_i \) are normally distributed with unknown mean \( \mu_i \) and variance \( \sigma_i^2 \) (i.e. are \( \mathcal{N}(\mu_i, \sigma_i^2) \)), \( 1 \leq i \leq k \), and \( \sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2 \) (say) is known. A (single-stage) rule \( \mathcal{G} \) long-used in this (distributionally-normal) setting in various instances of statistical decision problems is: Take \( n \) independent vectors \( \mathbf{X}_j = (X_{1j}, \ldots, X_{kj}) \), \( j = 1, \ldots, n \), where \( X_{ij} \) denotes the \( j \)th observation from \( \pi_i \); form the sample means \( \overline{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij} / n \) (\( i = 1, \ldots, k \)), and base the terminal decision solely on the statistics \( \overline{X}_1, \ldots, \overline{X}_k \). (We will deviate from this setting and rule from time to time below, but will make it clear when we do so and hence will not need the (complex) notation which would be necessitated by discussing general settings and rules from the outset.)

The following notation will make our discussion of the history and applicability precise. Let \( \mu[1] \leq \cdots \leq \mu[k] \) denote the values of \( \mu_1, \ldots, \mu_k \) in ranked (numerical) order. (It is assumed that both the numerical values of \( \mu_1, \ldots, \mu_k \) and the pairings of the \( \mu[1], \ldots, \mu[k] \) with the populations \( \pi_1, \ldots, \pi_k \) are completely unknown.)

The fixed number \( n \) of vectors required by rule \( \mathcal{G} \) depends on the particular problem under consideration. In the literature of what are called "ranking and selection" problems several procedures use rule \( \mathcal{G} \) (e.g., those of Bechhofer (1954), Gupta (1956), (1965), and others). In these problems \( k \geq 2 \) (otherwise they make no sense), and a \{goal, probability requirement, procedure\} structure is specified. A simple
example of such a problem is that of selecting the population (or, one of the populations) associated with the $i^{th}$ smallest mean (where $1 \leq i \leq k$). This selection desire is called one's goal, and typically a probability requirement is made on the chances of achieving it and a procedure is given (which tells how to sample, when to stop sampling, and what terminal decision to make) which meets the probability requirement. For many ranking and selection structures rule $\mathcal{O}$ has, in fact, some optimality properties.\footnote{While not all structures use rule $\mathcal{O}$ -- e.g. the nonparametric procedure of Bechhofer and Sobel (1958), the closed sequential procedure of Paulson (1964), the open sequential procedure of Bechhofer, Kiefer, and Sobel (1968), and the two-stage procedures of Dudewicz and Dalal (1975) do not use rule $\mathcal{O}$ -- the considerations given below for rule $\mathcal{O}$ generalize easily to procedures which do not use rule $\mathcal{O}$.} While hypothesis-testing for $k$-populations has long been covered (at least for the case $k = 2$, and especially when the null hypothesis is that there is really only one population) in almost every statistics

\footnote{Much more general goals have also been considered but do not concern us here.}

\footnote{The probability requirement affects one's sample size in, since the more stringent one's probability requirement vis-a-vis achieving the goal, the more sampling one must perform.}

\footnote{See Hall (1958), (1959), Bahadur and Goodman (1952), Lehmann (1966), and Eaton (1967).}

\footnote{Note that in rule $\mathcal{O}$ only the fixed number $n$ of independent vectors required depends on the structure on hand. Of course the various structures use the statistics generated by rule $\mathcal{O}$ in quite different ways.}
course, it was not possible for most teachers of statistics to include an introduction to ranking and selection until recently, simply because textbooks (which tend to lag the field by 25 years) did not\(^1\) include material on this important part of the statistical triplet \{hypothesis testing, estimation, ranking and selection\}. Yet while hypothesis testing has (traditionally, at least) asked "Are \(\pi_1, \ldots, \pi_k\) actually one population?" and since 1954 ranking and selection has (e.g.) asked "Which of \(\pi_1, \ldots, \pi_k\) is the best?", the (estimation) question "How good is the best?" was not formally broached until 1967. This important question amounts (in the simple setting we have chosen above) to estimating \(\mu[k]\) based on \(\bar{X}_1, \ldots, \bar{X}_k\). Study of this and related problems was certainly hindered by naive interpretations of invariance principles (e.g. invariance of maximum likelihood estimators), which lead one quite "naturally" to estimate \(\mu[k] \equiv \max(\mu_1, \ldots, \mu_k)\) by \(\bar{X}[k] \equiv \max(\bar{X}_1, \ldots, \bar{X}_k)\). However not only is it dubious to call \(\bar{X}[k]\) the MLE of \(\mu[k]\) (see Dudewicz (1971a) for details for the normal case and Dudewicz (1971b) for some general principles), but

\[E\bar{X}[k] > \mu[k] \quad \text{for all possible } \mu = (\mu_1, \ldots, \mu_k) \text{ vectors; i.e.,} \]

\(\bar{X}[k]\) always overestimates \(\mu[k]\) (see Dudewicz (1972) for details).

We will restrict our brief history to the problem of point estimation of \(\mu[k]\), though certainly interval estimation of \(\mu[k]\)

---

\(^1\)This was remedied for junior, senior, and graduate calculus-based courses by publication of Dudewicz (1976), and will soon be remedied even for non-calculus courses by Gibbons, Olkin, and Sobel (1975).

\(^2\)More generally we could seek to estimate some (or all) of \(\mu[1], \ldots, \mu[k]\). Estimating \(\mu[1]\) amounts to asking "How bad is the worst?"
is also of much interest.\(^1\) Here, the results of point estimation lead the way, and lead us to believe that \(\bar{X}_n\) will not be the center of reasonable intervals for \(\mu\), since \(\bar{X}_n\) "overestimates" \(\mu\). In fact, Dudewicz and Tong (1971) have shown that an optimal interval for \(\mu\) in some cases lies totally to the left of \(\bar{X}_n\). Another result pointed to by point estimation results (given below) is the fact proven by Dudewicz (1970) that under mild conditions no exact interval for \(\mu\) exists. For recent references and results on interval estimation, the interested reader can refer to Chen and Dudewicz (1976) or to Chen (1976).

For the problem of point estimation of \(\mu\) our knowledge now includes the following facts: \(\bar{X}_n\) is too large since (as discussed above) \(E\bar{X}_n > \mu\) for all \(\mu\): no unbiased estimator or \(\mu\) exists (see Blumenthal and Cohen (1968) for a sketch\(^2\) of a proof); and \(\bar{X}_n\) is asymptotically optimal in the sense of Weiss and Wolfowitz (1966), hence any "good" estimator of \(\mu\) must be close to \(\bar{X}_n\) for large \(n\) (see Dudewicz (1976a); for related results see Dudewicz (1973)).

\(^1\)For some novel applications, see Dudewicz (1973a).

\(^2\)This proof involves differentiability of the estimator and is somewhat unappealing. A nicer proof, involving completeness, was hinted at in Stanford University (1976), where it was stated "\(\mu\) is the smaller. Prove that there is no unbiased estimator of \(\theta = \min(\mu_1, \mu_2)\). Hint: You may find it helpful to use the fact that the family of normal distributions \(\{N(\mu, I): \mu \in \mathbb{R}\}\) is complete whenever \(I\) is any nondegenerate interval of real numbers."
We will illustrate the statistical application of these facts regarding estimation of $\mu_{[k]}$ with three specific examples, and will be non-technical in our presentation (for technical details see the references and Section 2 below). The first example involves what I will call the "3-way Bulb Problem", and illustrates a setting where it is appealing (but wrong) to estimate $\mu_{[1]}$. Here a three-way light bulb contains two filaments (one for low brightness, one for medium brightness), and the two combined produce high brightness. The question is how to state the average life of the bulb: do we state our estimate of $\mu_1$ (the mean life of the low brightness filament), our estimate of $\mu_2$ (the mean life of the medium brightness filament), $\mu_{[1]}$ (the smaller of $\mu_1$ and $\mu_2$), or some other quantity (such as $E \min(X_{1j}, X_{2j})$)? This depends on how the bulb will be used (e.g., for a user who will always use the bulb on high brightness $E \min(X_{1j}, X_{2j})$ is appropriate—though why such a user would buy a three-way bulb instead of a one-filament high-brightness bulb is unclear), but $\mu_{[1]}$ does not seem appropriate for any use situation (although the industry's own Better Light Better Sight Bureau (1971) indicates that this was being done in 1971; since then the Federal Trade Commission has taken action leading to estimates of all 3 of $\mu_1, \mu_2, E \min(X_{1j}, X_{2j})$ appearing on the bulb packages, while the estimate of $\mu_{[1]}$ has disappeared).

The second example involves the "Best 1 of k Report", and illustrates how the principles of estimating $\mu_{[k]}$ can be easily modified for application to related situations. In this situation, a manufacturer who wishes to portray his product in a favorable light tests
k batches of n units each (perhaps the testing is done in independent laboratories) and reports \( \bar{X}_{[k]} \) as an estimate of the product's quality.\(^1\) Here \( \mu_{[1]} = \ldots = \mu_{[k]} \) is known, and we wish to know (based on knowledge only of \( \bar{X}_{[k]} \) and \( k \), the other data being unavailable to us because it is "unsuitable") how to "cut down" \( \bar{X}_{[k]} \) to a reasonable estimator of the common mean \( \mu \). While we will not discuss the details of the solution, they should be fairly clear to readers who understand the principles stated above and the details of Section 2.

Our third example involves the numerous situations wherein one directly desires to estimate \( \mu_{[k]} \). For example, commonly an experiment is conducted to select the best of \( k \) populations,\(^2\) and \( \bar{X}_{[k]} \) is reported as an estimate of \( \mu_{[k]} \). We now know that \( \bar{X}_{[k]} \) is too large, and results on estimating \( \mu_{[k]} \) are directly applicable to better estimation of \( \mu_{[k]} \).

---

\(^1\)A similar example involving drugs was cited by Good (1971), p. 14.

\(^2\)For example, some common situations are: We select an agricultural variety and (prior to the growing season) desire to estimate total production; we select a medical treatment and wish to estimate cure rate; we select a type of equipment and wish to estimate its reliability; we select a weather forecasting method and wish to estimate its forecasting ability.
2. Point Estimation of \( \mu[i] \) in Normal Populations

In this section we use the notation developed at the beginning of Section 1, and give new results on point estimation of \( \mu[i] \) \((1 \leq i \leq k)\). While the cases \( i = 1 \) and \( i = k \) are the most commonly applicable, the results for general \( i \) are useful when costs associated with the various populations differ. To briefly summarize our results (a number of which come from applying the general theory of Dudewicz (1972) to the normal case), let \( \bar{X}_{[1]} \leq \cdots \leq \bar{X}_{[k]} \) denote the \( X_1, \ldots, X_k \) generated by procedure \( \mathcal{P} \) in numerical order. For the "natural" point estimator \( \bar{X}_{[i]} \) of \( \mu[i] \) we study the bias, obtaining upper and lower bounds \( \bar{U}_i \) and \( \bar{L}_i \) for \( \mathbb{E} \bar{X}_{[i]} \), and find that \( \bar{X}_{[i]} \) is asymptotically unbiased as \( n \to \infty \) and that \( \bar{U}_i \) \((\bar{L}_i)\) is the supremum \((\infimum)\) of \( \mathbb{E} \bar{X}_{[i]} \). We also find the minimax \( |\text{bias}| \) estimator of type \( \bar{X}_{[i]} + a_i \), note that \( \bar{X}_{[i]} \) is strongly consistent for \( \mu[i] \) \((\text{and some applications of this fact})\), give bounds on the mean squared error \( \mathbb{E} (\bar{X}_{[i]} - \mu[i])^2 \), and determine intervals in which the supremum and infimum of the mean squared error lie. We ultimately see that intuition is sufficiently far from reality that \( \bar{X}_{[i]} \) itself should not be seriously considered as an estimator of \( \mu[i] \); in the case \( i = k \) this becomes intuitively clear upon further thought, for \( \bar{X}_{[k]} = \max(\bar{X}_1, \ldots, \bar{X}_k) \) will tend to overestimate \( \mu[k] \) when \( \mu[1] = \cdots = \mu[k-1] = -\infty \) as well as when \( \mu[1] = \cdots = \mu[k-1] = \mu[k] \) \(\text{(in fact, for all } \mu)\). A refined intuition then grasps a similar result for \( \bar{X}_{[i]} \), and seeks \((\text{for example})\) to instead use \( \bar{X}_{[i]} + a_i \)
where $a_i$ has some desirable property, such as minimizing (over all constants $a_i$) the maximum (over all $y_i$) of $|E_1(x_i + a) - \mu_i|$. If we denote the $N(0,1)$ distribution function (d.f.) and density function (r.f.) by $\Phi(\cdot)$ and $\phi(\cdot)$ respectively, and denote $(1/\sigma)\phi(y/\sigma)$ by $\phi_\sigma(y)$, then the quantities defined in Dudewicz (1972) for a location parameter family are (for $i = 1, \ldots, k$) as follows in the case of normality:

\[
\begin{align*}
\Phi_{\sigma}(y) &= \int_{-\infty}^{\infty} \phi_\sigma(y) dy = 0 ; \\
G_n(y|\phi_\sigma) &= P[\bar{X}_i - \mu_i \leq y] = \Phi_\sigma \left( \frac{\bar{X}_i - \mu_i}{\sigma/\sqrt{n}} \right) = \Phi \left( \frac{y}{\sigma/\sqrt{n}} \right) ; \\
\phi_n(y|\phi_\sigma) &= \frac{1}{\sigma/\sqrt{n}} \phi_\sigma \left( \frac{y}{\sigma/\sqrt{n}} \right) = \phi_\sigma/\sqrt{n}(y) ; \\
h_n(g_n) &= E[\max \text{ of } \ell \text{ r.v.'s with r.f. } g_n(y|\phi_\sigma)] \\
&= E[\max \text{ of } \ell \text{ } N(0,\sigma^2/n) \text{ r.v.'s}] \\
&= (\sigma/\sqrt{n})E[\max \text{ of } \ell \text{ } N(0,1) \text{ r.v.'s}] = (\sigma/\sqrt{n})h_\phi(\ell) ; \\
h_n'(g_n) &= -h_n(g_n) = -((\sigma/\sqrt{n})h_\phi(\ell)) .
\end{align*}
\]

Note that in the normal case, since $h_n(g_n) = -h_n'(g_n) = (\sigma/\sqrt{n})h_\phi(\ell)$ ($\ell = 1, 2, \ldots$), only $h_\phi(\ell)$, and not $h_\phi'(\ell)$, need be tabulated. $h_\phi(\ell) > 0$ for $\ell \geq 2$ since $\int_{-\infty}^{\infty} x\phi(x)dx = 0$ and the positive
weighting function \( [\Phi(x)]^{\ell-1} \) assigns greater weight to \( +x \) than to \( -x \) for all \( x > 0 \).) Tables of quantities more general than \( h_\beta(\phi) \) have been computed by (e.g.) Teichroew (1956) where \( h_\beta(\phi) = E(x_1; \beta) \), and by Harter (1961) where \( h_\beta(\phi) = E(x_1 | \beta) \). Tables of \( h_\beta(\phi) \) have been computed by Tippett (1925). In Table I we present some values of \( h_\beta(\phi) \) obtained from Harter (1961) for \( \ell = 2(1)10(5) \) 25, 50, 100, and from Tippett (1925) for \( \ell = 500, 1000 \). (For further references, see Kendall and Stuart (1963), pp. 329, 336.)

Table I. Values of \( h_\beta(\phi) \)

<table>
<thead>
<tr>
<th>( \ell )</th>
<th>( h_\beta(\phi) )</th>
<th>( \ell )</th>
<th>( h_\beta(\phi) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.56419</td>
<td>50</td>
<td>2.24907</td>
</tr>
<tr>
<td>3</td>
<td>0.81628</td>
<td>100</td>
<td>2.50759</td>
</tr>
<tr>
<td>4</td>
<td>1.02938</td>
<td>150</td>
<td>2.64925</td>
</tr>
<tr>
<td>5</td>
<td>1.16296</td>
<td>200</td>
<td>2.74604</td>
</tr>
<tr>
<td>6</td>
<td>1.26721</td>
<td>250</td>
<td>2.81918</td>
</tr>
<tr>
<td>7</td>
<td>1.35218</td>
<td>300</td>
<td>2.87777</td>
</tr>
<tr>
<td>8</td>
<td>1.42360</td>
<td>350</td>
<td>2.92651</td>
</tr>
<tr>
<td>9</td>
<td>1.48501</td>
<td>400</td>
<td>2.96818</td>
</tr>
<tr>
<td>10</td>
<td>1.53875</td>
<td>500</td>
<td>3.03670</td>
</tr>
<tr>
<td>15</td>
<td>1.73591</td>
<td>1000</td>
<td>3.24144</td>
</tr>
<tr>
<td>20</td>
<td>1.86748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.96531</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Corollary (2.18) and Theorem (2.23) of Dudewicz (1972) and the results of (2.1) above, the following theorem emerges for the normal case.
Theorem (2.2): For any \( i \) (1 \( \leq i \leq k \)),

\[
\mu_{[i]} - \left( \frac{\sigma}{\sqrt{n}} \right)^{h_{k-1}+1}(\phi) \leq E_{\mu} \bar{X}_{[i]} \leq \mu_{[i]} + \left( \frac{\sigma}{\sqrt{n}} \right)^{h_{i}}(\phi)
\]

and \( \bar{X}_{[i]} \) is asymptotically unbiased (as \( n \to \infty \)) as an estimator of \( \mu_{[i]} \).

Now, the bounds of Theorem (2.2) are actually the supremum and infimum. That is, let \( \mathbb{R} = \{x: -\infty < x < \infty\} \), \( \mathbb{R}^+ = \{x: x \geq 0\} \), and (for \( \delta \in \mathbb{R}^+ \))

\[
\Omega_{\delta}(a, b, c, \ldots) = \{(\mu_1, \ldots, \mu_k): \mu_{[k]} - \mu_{[k-1]} \geq \delta, \mu_i \in \mathbb{R}^+
\]

\( (i = 1, \ldots, k), a, b, c, \ldots \) are held fixed\} .

(In general \( a, b, c, \ldots \) will be several of \( \mu_{[1]}, \ldots, \mu_{[k]} \)). Then

Theorem (2.3): For any \( i \) (1 \( \leq i \leq k \)),

\[
\inf_{\mu} \left[ E_{\mu} \bar{X}_{[i]}: \mu \in \Omega_{0}(\mu_{[i]})) \right] = \mu_{[i]} - \left( \frac{\sigma}{\sqrt{n}} \right)^{h_{k-1}+1}(\phi)
\]

and

\[
\sup_{\mu} \left[ E_{\mu} \bar{X}_{[i]}: \mu \in \Omega_{0}(\mu_{[i]})) \right] = \mu_{[i]} + \left( \frac{\sigma}{\sqrt{n}} \right)^{h_{i}}(\phi)
\]

(The proof of Theorem (2.3), which in addition to results of Dudewicz (1972) for the general location parameter case uses special properties of the normal distribution, is given in an Appendix.)

Linear corrections to \( \bar{X}_{[i]} \) for minimum |bias| are worthy of
special note, since we may use them (in the normal case) to determine
the sample size \( n \) needed to satisfy several criteria (ranking and
selection, estimation, or both): (1) Set \( n \) as dictated by the
ranking and selection use of Rule \( \mathcal{O} \), say \( n_1 \); or (2) Set \( n \) to
make certain minimax \(|\text{bias}|\)'s suitably "small," say \( n_2 \); or (3) Set
\( n = \max(n_1, n_2) \). The following theorem follows from Dudewicz (1972)
and (2.1).

**Theorem (2.4):** For any \( i \) (1 \( \leq \) \( i \) \( \leq \) \( k \)) we minimize (over all constants
\( a \)) the maximum (over all \( \mu \)) of \( |E_{\mu}((\bar{X}_{[1]}+a)-\mu_{[1]}| \) by choosing

\[
a = \frac{\sigma}{2\sqrt{n}} \left( h_{k}(\phi) - h_{k-1}(\phi) \right) \]

Table I of values of \( h_{\ell}(\phi) \) indicates that for \( k \) in the range
in which Rule \( \mathcal{O} \) would usually be used (\( k \leq 10 \)) the factor \( h_{\ell}(\phi) \)
in the bias is not seriously detrimental, being only 1.5 for \( \ell = 10 \).
Even if \( \ell \) were of the size associated with large screening experiments,
the factor \( h_{\ell}(\phi) \) would still be only 3.0 for \( \ell = 500 \). As an
example, if one were setting \( n \) large enough to make the minimax
\(|\text{bias}| \) in \( \bar{X}_{[k]} + a \), as an estimator of \( \mu_{[k]} \), \( \leq \varepsilon \) \( (\varepsilon > 0) \), he
would find approximately that if \( n_0 \) sufficed for \( k = 2 \), \( 4n_0 \) would
suffice for \( k = 5 \); and that if \( n_0 \) sufficed for \( k = 9 \), \( 4n_0 \) would
suffice for \( k = 500 \), since by Theorem (2.4) the minimax \(|\text{bias}| \) is
0.5 (\( \sigma/\sqrt{n} \))\( h_{k}(\phi) \).

Note that if there are restrictions on the \( \mu_i \) (\( i = 1, \ldots, k \)) in
a practical case, then the $\inf$ and $\sup$ of Theorem (2.3) can be improved. For example, if $A \leq \mu \leq B (i = 1, \ldots, k)$, then "A" will replace "$-\infty$" and "B" will replace "$+\infty$" in that work. (A common case is $A = 0, B = +\infty$.) Such a process will result in a smaller $n_1$ being needed for estimation as in the previous paragraph.

If the $\sup$ and $\inf$ were desired over a more restricted set than $\mu \in \Omega_0(\mu[1])$, say $\mu \in \Omega_0(\mu[1])$, that $\sup$ and $\inf$ would also be attained by raising (lowering) the components of $\mu$ to the highest (lowest) possible values. Noting that this is somewhat analogous to the set over which a Probability Requirement is made in the "indifference zone" formulation of ranking and selection problems, one might at first think we would be interested in the $\sup$ ($\inf$) over $\mu \in \Omega_0(\mu[1])$. However, since our aim is good estimation of $\mu[1]$ regardless of $\mu$, the set used above ($\mu \in \Omega_0(\mu[1])$) will usually be the proper one. (For special uses of the estimate of $\mu[1]$ one may only "care" when, for some $\delta$, $\mu \in \Omega_0(\mu[1])$.)

We note that in passing that strong consistency of $\bar{X}[1]$ as an estimator of $\mu[1]$ follows from Section 3 of Dudewicz (1972). Hence obviously $g(\bar{X}[1])$ converges w.p. 1 to $g(\mu[1])$ for any continuous real-valued function $g(\cdot)$. Therefore $g(\bar{X}[k])$ may be used to yield an estimate of $g(\mu[k])$, where $g(\cdot)$ is a continuous function such that if we knew the mean of the selected population to be $\mu$, then we would know the expected worth to us (e.g., in dollars) of the selected population to be $g(\mu)$. Other applications might occur for a Bayesian taking $\mu[1]$ to be a r.v. ($1 \leq i \leq k$).
Note that strong consistency of $\bar{X}_{[1]}$ as an estimator of $\mu_1$ implies strong consistency of $\bar{X}_{[i]} + a_n$ where $\lim_{n \to \infty} a_n = 0$ $(i = 1, \ldots, k)$. (This, of course, was also the case for asymptotic unbiasedness.)

We next wish to consider the mean squared error of $\bar{X}_{[1]}$ as an estimator of $\mu_1$, applying the location parameter case results of Dudewicz (1972). Under normality, it can be seen that the functions

\[
H_\infty(x) = \Pr[\text{Minimum of } k-1+l \ N(\mu_{[1]}, \sigma^2/n) \text{ r.v.'s is } \leq x]
= \Pr\left[\text{Min of } k-1+l \ N(0,1) \text{ r.v.'s is } \leq \frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right]
= 1 - \left[1 - \Phi\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\right]^{k-1+l};
\]

\[
J_\infty(x) = \Pr[\text{Maximum of } i \ N(\mu_{[1]}, \sigma^2/n) \text{ r.v.'s is } \leq x]
= \Pr\left[\text{Max of } i \ N(0,1) \text{ r.v.'s is } \leq \frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right] = \left[\Phi\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\right]^i.
\]

Thus, we find

\[
\int_{\mu_{[1]}}^{\infty} (x-\mu_{[1]})^2 dH_\infty(x) = \int_{\mu_{[1]}}^{\infty} (x-\mu_{[1]})^2 d\left\{1 - \left[1 - \Phi\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\right]^{k-1+l}\right\}
= (\sigma^2/n) \int_{0}^{\infty} x^2 d\left\{-[1 - \Phi(x)]^{k-1+l}\right\} = (\sigma^2/n) \int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^{k-1+l}\right\};
\]

\[
\int_{-\infty}^{\mu_{[1]}} (x-\mu_{[1]})^2 dJ_\infty(x) = \int_{-\infty}^{\mu_{[1]}} (x-\mu_{[1]})^2 d\left\{\Phi\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\right\}^i = (\sigma^2/n) \int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^i\right\};
\]
and
\[
\int_{-\infty}^{\mu_{[i]}} (x - \mu_{[i]})^2 dH_\sigma(x) = \int_{-\infty}^{\mu_{[i]}} (x - \mu_{[i]})^2 d \left\{ 1 - \frac{\Phi^{\mu_{[i]}}}{\sigma/\sqrt{n}} \right\}^{k-i+1}
\]
\[= \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{0} x^2 d \left\{ [-1 - \Phi(x)]^{k-i+1} \right\} = \left( \frac{\sigma^2}{n} \right) \int_{0}^{\infty} x^2 d \left\{ [\Phi(x)]^{k-i+1} \right\} ;
\]
\[
\int_{\mu_{[i]}}^{\infty} (x - \mu_{[i]})^2 dJ_\sigma(x) = \int_{\mu_{[i]}}^{\infty} (x - \mu_{[i]})^2 d \left\{ \Phi \left( \frac{x - \mu_{[i]}}{\sigma/\sqrt{n}} \right) \right\}^{2} = \left( \frac{\sigma^2}{n} \right) \int_{0}^{\infty} x^2 d \left\{ [\Phi(x)]^{i} \right\} .
\]

It then follows from Theorem (4.5) of Dudewicz (1972) that we have

**Theorem (2.5):** For any \( i \ (1 \leq i \leq k) \) and any \( \mu \in \Omega_{o}(\mu_{[i]}) \),
\[
\left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{0} x^2 d \left\{ [\Phi(x)]^{k-i+1} \right\} + \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{0} x^2 d \left\{ [\Phi(x)]^{i} \right\} \leq E_{\mu}(\overline{X}_{[i]} - \mu_{[i]})^2
\]
\[
\leq \left( \frac{\sigma^2}{n} \right) \int_{0}^{\infty} x^2 d \left\{ [\Phi(x)]^{k-i+1} \right\} + \left( \frac{\sigma^2}{n} \right) \int_{0}^{\infty} x^2 d \left\{ [\Phi(x)]^{i} \right\} .
\]

In the case of normality, it is possible to further bound the supremum and infimum, thus obtaining an interval in which each must lie.

**Theorem (2.6):** For any \( i \ (1 \leq i \leq k) \), taking the inf and sup over \( \Omega_{o}(\mu_{[i]}) \),
\[
\inf_{\mu} E_{\mu}(\overline{X}_{[i]} - \mu_{[i]})^2 \leq \min \left( \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{\infty} x^2 d \left\{ [\Phi(x)]^{k-i+1} \right\}, \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{\infty} x^2 d \left\{ [\Phi(x)]^{i} \right\} \right),
\]
\[
\sup_{\mu} E_{\mu}(\overline{X}_{[i]} - \mu_{[i]})^2 \geq \max \left( \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{\infty} x^2 d \left\{ [\Phi(x)]^{k-i+1} \right\}, \left( \frac{\sigma^2}{n} \right) \int_{-\infty}^{\infty} x^2 d \left\{ [\Phi(x)]^{i} \right\} \right) .
\]
Proof: Since \( H_M(x) \) and \( J_M(x) \) converge weakly to \( H_\infty(x) \) and \( J_\infty(x) \) (respectively), it follows that, if \( x^2 \) is uniformly integrable in \( H_M \) and \( J_M \), then

\[
\lim_{M \to \infty} \int_{-\infty}^{\infty} x^2 dH_M(x) = \int_{-\infty}^{\infty} x^2 dH_\infty(x) = (\sigma^2/n) \int_{-\infty}^{\infty} x^2 d\{[\Phi(x)]^k\} = \int_{-\infty}^{\infty} x^2 d\{[\Phi(x)]^{k+1}\}
\]

\[
\lim_{M \to \infty} \int_{-\infty}^{\infty} x^2 dJ_M(x) = \int_{-\infty}^{\infty} x^2 dJ_\infty(x) = (\sigma^2/n) \int_{-\infty}^{\infty} x^2 d\{[\Phi(x)]^4\}.
\]

In this case it must be the case that the inf (sup) is less (greater) than or equal to each of these quantities. The fact that \( x^2 \) is uniformly integrable in \( H_M \) follows from a modification of the proof of Lemma A-2. The fact that \( x^2 \) is uniformly integrable in \( J_M \) requires major modification of the proof of Lemma A-3, as we will now note. Using the non-increasing function

\[
\psi(x) = \begin{cases} x^2, & x \leq -L \\ 0, & x > -L \end{cases}
\]

we find

\[
\int_{-\infty}^{\infty} \psi(x) dG(x) \geq \int_{-\infty}^{\infty} \psi(x) dF(x)
\]

\[
\int_{-\infty}^{-L} x^2 dH_M(x) + L^2 (1-H_M(-L)) \geq \int_{-\infty}^{-L} x^2 dJ_M(x) + L^2 (1-J_M(-L))
\]

\[
\int_{-\infty}^{-L} x^2 dH_M(x) \geq \int_{-\infty}^{-L} x^2 dJ_M(x) + L^2 \{H_M(-L)-J_M(-L)\}
\]
Now, since $H_{M}(-L) \geq J_{M}(-L)$ and since $\int_{-\infty}^{-L} x^2 dH_{M}(x) \to 0$ uniformly in $M$, we find that

$$0 \leq \int_{-\infty}^{-L} x^2 dJ_{M}(x) \leq \int_{-\infty}^{-L} x^2 dH_{M}(x) \to 0$$ uniformly in $M$.

Thus, there is (for any fixed $\mu_{[1]}$) an $L_{L}(\epsilon)$ such that for $L > L_{L}(\epsilon)$ we have $\int_{-\infty}^{-L} x^2 dJ_{M}(x) < \epsilon/2$ uniformly in $M$.

Since $J_{M}(x) \geq J_{\infty}(x)$, if we define

$$F(x) = \begin{cases} J_{M}(x) , & x \geq L \\ 0 , & x < L \end{cases}$$

$$G(x) = \begin{cases} J_{\infty}(x) , & x \geq L \\ 0 , & x < L \end{cases}$$

$$\psi(x) = \begin{cases} x^2 , & x \geq L \\ 0 , & x < L \end{cases}$$

then

$$\int_{-\infty}^{\infty} \psi(x) dF(x) \leq \int_{-\infty}^{\infty} \psi(x) dG(x)$$

$$\int_{L}^{\infty} x^2 dJ_{M}(x) + L^2 J_{M}(L) \leq \int_{L}^{\infty} x^2 dJ_{\infty}(x) + L^2 J_{\infty}(L)$$

$$0 \leq \int_{L}^{\infty} x^2 dJ_{M}(x) \leq L^2 \{ J_{\infty}(L) - J_{M}(L) \} + \int_{L}^{\infty} x^2 dJ_{\infty}(x) \leq \int_{L}^{\infty} x^2 dJ_{\infty}(x).$$

Now since $\int_{-\infty}^{\infty} x^2 dJ_{\infty}(x)$ exists, for $L > L_{2}(\epsilon)$ we have $\int_{L}^{\infty} x^2 dJ_{M}(x) \leq \epsilon/2$ uniformly in $M$. The result then follows as in Lemma A-3.

We now find the $\min$ and $\max$ needed in Theorem (2.6). This will
allow us to specify intervals in which the inf and sup must lie, and to study the lengths of these intervals.

**Lemma (2.7):** Let \( Z_1, \ldots, Z_n \) be independent r.v.'s, each with d.f. \( F \) such that \( F(z-) + F(-z) = 1 \) for all \( z \) (e.g., this occurs if \( F \) has a fr.f. which is symmetric about 0). Let \( G_n(z) \) be the d.f. of \( \max_{1 \leq i \leq n} Z_i \). Let \( h(u) \) be any non-decreasing function of \( u \geq 0 \) such that \( h(0) > -\infty \). Then \( \int_0^\infty h(u)dG_n(u) \) is non-decreasing in \( n \).

**Corollary (2.8):** \( \int_{-\infty}^\infty x^2 d\{\Phi(x)\}^2 = 1 \) for \( n = 1, 2 \) and is a strictly increasing function of \( n \) thereafter.

**Theorem (2.9):** For any \( i \) (\( 1 \leq i \leq k \)), \( \inf \{ E_\mu(\bar{X}_{[i]} - \mu_{[i]} \}^2 : \mu \in \Omega_0(\mu_{[i]}) \} \) is in the closed interval
\[
((\sigma^2/n) \int_{-\infty}^0 x^2 d\{\Phi(x)\}^{k-i+1}) + (\sigma^2/n) \int_{-\infty}^0 x^2 d\{\Phi(x)\}^{1} \\
(\sigma^2/n) \int_{-\infty}^\infty x^2 d\{\Phi(x)\}^{k-i+1}) \text{ if } i \geq \frac{k+1}{2}
\]
\[
((\sigma^2/n) \int_{-\infty}^0 x^2 d\{\Phi(x)\}^{k-i+1}) + (\sigma^2/n) \int_{-\infty}^0 x^2 d\{\Phi(x)\}^{1} \\
(\sigma^2/n) \int_{-\infty}^\infty x^2 d\{\Phi(x)\}^{1}) \text{ if } i < \frac{k+1}{2}
\]

and \( \sup \{ E_\mu(\bar{X}_{[i]} - \mu_{[i]} \}^2 : \mu \in \Omega_0(\mu_{[i]}) \} \) is in the closed interval
\[(\sigma^2/n) \int_{-\infty}^{\infty} x^2 d\{[\phi(x)]^{i+1}\},\]

\[= (\sigma^2/n) \int_{0}^{\infty} x^2 d\{[\phi(x)]^{k-i+1}\} + (\sigma^2/n) \int_{-\infty}^{0} x^2 d\{[\phi(x)]^{i}\}) \text{ if } i \geq \frac{k+1}{2},\]

\[(\sigma^2/n) \int_{-\infty}^{\infty} x^2 d\{[\phi(x)]^{k-i+1}\},\]

\[= (\sigma^2/n) \int_{0}^{\infty} x^2 d\{[\phi(x)]^{k-i+1}\} + (\sigma^2/n) \int_{-\infty}^{0} x^2 d\{[\phi(x)]^{i}\}) \text{ if } i < \frac{k+1}{2}.\]

Corollary (2.10): The inf and sup of Theorem (2.9) each lie in an interval of length

\[= (\sigma^2/n) \int_{0}^{\infty} x^2 d\{[\phi(x)]^{k-i+1}\} - \int_{-\infty}^{0} x^2 d\{[\phi(x)]^{i}\}) \text{ if } i \geq \frac{k+1}{2},\]

\[= (\sigma^2/n) \int_{0}^{\infty} x^2 d\{[\phi(x)]^{k-i+1}\} - \int_{-\infty}^{0} x^2 d\{[\phi(x)]^{k-i+1}\}) \text{ if } i < \frac{k+1}{2}.\]

By Corollary (2.8), the intervals of these lengths for the inf and sup fail to be disjoint iff \(i = \frac{k+1}{2}\), or \((i, k-i+1)\) is a permutation of \((1, 2)\)). In that case they have exactly one common point.
3. Insights on Statistical Phenomena

In this section we wish to note, by examples, how the results and principles of the area of estimation of ranked parameters can be interpreted to give us new insights on seemingly-mysterious phenomena observed in statistics as early as 1927, and to pose questions for future workers in statistical foundations. These phenomena (still under active investigation by statisticians) involve such facts as the following. First, in estimating \((\mu_1, \ldots, \mu_k, \bar{X}_1, \ldots, \bar{X}_k)\) has been shown by C. Stein and others to be inferior (for certain specific loss functions) to estimators which "shrink" \((\bar{X}_1, \ldots, \bar{X}_k)\). Second, in estimating \(\mu_1\) when \(\bar{\mu} = (\mu_1 + \cdots + \mu_k)/k\) is known, as often occurs in educational testing, it has been known since 1927 (see Kelley (1927)) that it is beneficial to move \(\bar{X}_1\) towards \(\bar{\mu}\); recently Bayesian interpretations and extensions of this fact have been given (see Novick, Jackson, Thayer, and Cole (1971)). While the results of the first type (by Stein and others) are based on solid mathematical statistics, they are sometimes attacked via their dependence on the specific loss function assumed. And, results of the second type are often questioned by those who scoff at educational statistics, as well as by anti-Bayesians. To others, both types of phenomena are simply mysterious. To us, it seems that the ranked parameter estimation results (which decrease \(\bar{X}_{[i]}\) if \(i > (k+1)/2\), leave \(\bar{X}_{[i]}\) unaltered if \(i = (k+1)/2\), and increase \(\bar{X}_{[i]}\) if \(i < (k+1)/2\), which are easy to understand once one's intuition has been "educated" by the
facts, cast an indirect light of implication (the full details of which have yet to be worked out by statisticians) which makes the "mysteries" stated seem a reasonable part of the field of statistics.

It remains to be seen whether a global theory can be constructed which formally incorporates all of these phenomena (including ranked parameter estimation) under "one roof." (Here we refer to a theory other than the well-known statistical decision theory.) This merits intensive investigation, since a satisfactory joint explanation of these phenomena can be expected to have important implications for other areas of statistics as well.
4. **A Plan for Future Advances**

In this section we give ideas on desirable future work in the area, note work in progress, and note some open problems, with the total comprising a plan for future advances in the area.

First, with regard to the "fringe area" noted in the second example of Section 1 (the "Best 1 of k Report") it should be noted that, although the problem is clear to those who understand this paper, it is worth having the details of point and interval estimation written down in detail (with associated high-quality tables needed for application) due to its practical importance. A plan for attacking this problem would be to first consider the case where \( (\bar{X}_{[1]}, k) \) alone are known and the variances of populations \( \pi_1, \ldots, \pi_k \) are equal and known \( (\sigma^2_1 = \ldots = \sigma^2_k = \sigma^2 \text{ known}) \), where the d.f. \( \Phi(\sqrt{n} x/\sigma) \) will play a large role. It would be very interesting to also compare the optimal solution there with that for the problem where all of the data are available and one uses \( \bar{X} = (\bar{X}_{[1]} + \ldots + \bar{X}_{[k]})/k \), to see how much better one could do it if one were able to obtain the rest of the data. Next in our plan would come the case of \( \sigma^2 \) unknown, where such distributions as the multivariate-t (in particular its equicoordinate points) arise, with various estimators of \( \sigma^2 \); again, a comparison with \( \bar{X} \) is of interest, as is a comparison with \( \bar{X}_{[1]} \). Finally, the case of \( \sigma^2_1 \)'s unknown and unequal could be considered (and procedures related to those of Dudewicz and Dalal (1975) may be of use here; a product of Student's-t distributions with unequal degrees of freedom may be expected to arise). While the attack plan seems straightforward, surprises may
arise which would be interesting in their own right (especially in the cases other than $\sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2$ known).

In the area of estimation of ranked parameters proper, we see a four-pronged attack being necessary to extend and develop the full theory needed to handle the various situations where the problem arises, and to handle them well. The four prongs are as follows.

First, in the case studied in this paper we need to know more about the best estimator of $\mu_{[k]}$, and how it depends on $\overline{X}_{[1]}, \ldots, \overline{X}_{[k-1]}$ as well as how it depends on $\overline{X}_{[k]}$. Currently we know no unbiased estimator of $\mu_{[k]}$ exists, but can achieve a bias which approaches zero as $n \to \infty$. We know that a good estimator should (if possible) collapse to $\overline{X}_{[k]}$ as $\mu_{[k-1]} \to -\infty$, and should (if possible) collapse to $\overline{X}$ when $\mu_{[1]} = \cdots = \mu_{[k]}$, though what it should look like in intermediate situations is largely unknown. We do know how to adjust via $\overline{X}_{[k]} - a_k$ optimally, some is known about estimators like $\overline{X}_{[k]} - g(\overline{X}_{[k]} - \overline{X}_{[k-1]})$, and little is known about estimators like

$$\overline{X}_{[k]} - g(\overline{X}_{[k]} - \overline{X}_{[1]}, \ldots, \overline{X}_{[k]} - \overline{X}_{[k-1]}) .$$

A hint for proper choice of $g(\cdot, \ldots, \cdot)$ would be yielded by solution for the optimal function in the case where $\mu_{[1]}, \ldots, \mu_{[k-1]}$ are known, since they can at least be naively estimated by $\overline{X}_{[1]}, \ldots, \overline{X}_{[k-1]}$ to obtain a first estimator to study, which hopefully may ultimately lead to an optimal estimator. Currently the author and Professor Saul Blumenthal are working on some aspects of this problem, especially on
second-order adjustments to existing estimators using mean squared error as a criterion. I have a suspicion that the area of ridge analysis (e.g., see McDonald (1975)) may hold some relationships to or hints for some of this work, though this is unconfirmed at the present time.

Second, still in the case studied in this paper, the maximum probability estimators (MPE's) investigated by Dudewicz (1973) bear further investigation. Can one obtain a series expression for the MPE? How does it compare with other estimators on (e.g.) mean squared error? How does the MPE depend on the interval \((-r,r)\) chosen in the Weiss-Wolfowitz theory?

Third, work is needed on cases which often arise in practice such as: What if the data arose from a split-plot, factorial, or other experimental design? How do experimental designs compare in efficiency regarding estimation of \(\mu[k]\) (e.g., this study might parallel one of V. Bawa for selection problems in factorial designs)? What new wrinkles emerge if one has correlated observations or time series data?

Finally, consideration of other settings is needed. While normal and location-parameter populations have been intensively investigated, as have scale-parameters populations (especially by H. J. Chen in the last several years), in simulation, reliability, and biostatistics non-normal non-location non-scale cases commonly arise. While one can still use sample means and (by central limit theorems) be assured of good large-sample properties, such procedures will be far from optimal, and exact theory is desirable.
It is our desire and hope that this summary of where we stand, and where we still need to go, will stimulate research in the area, directed towards the important problems outlined above.
Appendix: Proof of Theorem (2.3)

By Theorem (2.2), the infimum is 
\[ \mu_{[1]}^{-\frac{\sigma}{\sqrt{n}}}h_{k-i+1}(\phi) \]
and the supremum is 
\[ \mu_{[1]}^{+\frac{\sigma}{\sqrt{n}}}h_{i}(\phi). \]
We will now show that

\[ \inf \{ E_{\mu} \bar{X}[1]: \mu \in \Omega_{0}(\mu_{[1]}) \} \leq \mu_{[1]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi) \]

\[ \sup \{ E_{\mu} \bar{X}[1]: \mu \in \Omega_{0}(\mu_{[1]}) \} \geq \mu_{[1]} + (\sigma/\sqrt{n})h_{i}(\phi). \]

Now, since we are taking the inf and sup over more restricted sets,

\[ \inf \{ E_{\mu} \bar{X}[1]: \mu \in \Omega_{0}(\mu_{[1]}) \} \leq \inf \{ E_{\mu} \bar{X}[1]: \mu = (\mu_{[1]}, \ldots, \mu_{[1]}, \mu_{[1]}, \ldots, \mu_{[1]}), \mu \in \Omega_{0}(\mu_{[1]}) \} \]

\[ \sup \{ E_{\mu} \bar{X}[1]: \mu \in \Omega_{0}(\mu_{[1]}) \} \geq \sup \{ E_{\mu} \bar{X}[1]: \mu = (\mu_{[1]}, \ldots, \mu_{[1]}, \mu_{[1]}, \ldots, \mu_{[k]}), \mu \in \Omega_{0}(\mu_{[1]}) \} \} \]

**Case 1. The infimum.** First, 
\[ E_{\mu} \bar{X}[1], \mu = (-M, \ldots, -M, \mu_{[1]}, \ldots, \mu_{[1]}), \]
develops as \( M \). If we let \( H_{M}(x) \) denote \( F_{X}[1](-x) \) with \( \mu = (-M, \ldots, -M, \mu_{[1]}, \ldots, \mu_{[1]}), \) the desired
\[ \inf \{ E_{\mu} \bar{X}[1]: \mu = (-M, \ldots, -M, \mu_{[1]}, \ldots, \mu_{[1]}), \mu \in \Omega_{0}(\mu_{[1]}) \} = \lim_{M \to \infty} \int_{-\infty}^{\infty} x dH_{M}(x). \]
However, the following weak convergence holds as \( M \to \infty: \)
\[ H_M(x) \to H_\infty(x) \equiv \bar{F}_{\{x[i]\}}(x) \text{ with } \mu = (-\infty, \ldots, -\infty, \mu_{[1]}, \ldots, \mu_{[1]}). \]

Thus, if \(|x|\) is uniformly integrable in \(H_M\), then

\[
\lim_{M \to \infty} \int_{-\infty}^{\infty} xdH_M(x) = \int_{-\infty}^{\infty} xdH_\infty(x) = \mu_{[1]} - (c/\sqrt{n})h_{k-i+1}(\phi).
\]

Since \(|x|\) is uniformly integrable in \(H_M\) by Lemma A-2, this part of the theorem is proven.

**Case 2. The supremum.** First, \(E_{\mu_{[1]}}\), with \(\mu = (\mu_{[1]}, \ldots, \mu_{[1]} M, \ldots, M)\), increases as \(M\). If we let \(J_M(x)\) denote \(F_{\{x[i]\}}(x)\) with \(\mu = (\mu_{[1]}, \ldots, \mu_{[1]}, M, \ldots, M)\), the desired

\[
\sup\{E_{\mu_{[1]}}: \mu = (\mu_{[1]}, \ldots, \mu_{[1]}, M, \ldots, M) \in \Omega(\mu_{[1]})\} = \lim_{M \to \infty} \int_{-\infty}^{\infty} xdJ_M(x).
\]

However, the following weak convergence holds as \(M \to \infty:\)

\[ J_M(x) \to J_\infty(x) \equiv \bar{F}_{\{x[i]\}}(x) \text{ with } \mu = (\mu_{[1]}, \ldots, \mu_{[1]}, +\infty, \ldots, +\infty). \]

The theorem follows as in Case 1, now using the fact that \(|x|\) is uniformly integrable in \(J_M\) by Lemma A-3.

We now need the following notation. Let \(\Omega(\not=) = \{\mu: \mu_{[1]} \not= \mu_{[2]} \not= \ldots \not= \mu_{[k]}\}\). If \(\mu \in \Omega(\not=)\), let \(\bar{x}_{(1)}\) denote the sample mean produced by the population
associated with \( \mu_{[i]} \) \( (i = 1, \ldots, k) \). If there is at least one break in the string of inequalities \( \mu_{[1]} \neq \cdots \neq \mu_{[k]} \), then the situation is that we have \( \ell (1 \leq \ell < k) \) groups of equal parameters

\[
\mu_{[1]} = \cdots = \mu_{[i_1]} \neq \mu_{[i_1+1]} = \cdots = \mu_{[i_2]}
\]

\[
\neq \cdots \neq \mu_{[i_{\ell-1}+1]} = \cdots = \mu_{[k]}
\]

with \( i_1, \ldots, i_{\ell-1} \) integers

\[
(0 = i_0 < i_1 < i_2 < \cdots < i_{\ell-1} \leq k-1 < i_\ell \equiv k)
\]

and we let

\[
\bar{X}_{(1)} \leq \bar{X}_{(2)} \leq \cdots \leq \bar{X}_{(\ell)} \leq \bar{X}_{(1)}
\]

be the ranked values of the sample means from the population(s) associated with parameter \( \mu_{[i_{\ell}+1]} \) \( (j = 0, \ldots, \ell-1) \). Finally, let \( S_k \) be the symmetric group on \( k \) elements, i.e.,

\[
\{\alpha : \alpha = (\alpha(1), \ldots, \alpha(k)) \text{ is a permutation of } (1, \ldots, k)\}.
\]

**Lemma A-1**: For any \( \mu \in \Omega_{\mu_{[1]}} \),

\[
\frac{dF}{dx} \left( \frac{x}{\bar{X}_{[1]}} \right) \leq \sum_{\ell=1}^{k} \frac{k!}{\ell!} \sum_{\beta \in S_k} \left( \sum_{j=1}^{\ell} \frac{F_{\beta(j)}(x)}{\bar{X}_{[\beta(j)]}} \right) \left( \frac{F_{\beta(1)}(x) \cdots F_{\beta(\ell)}(x)}{\bar{X}_{[\beta(1)]} \cdots \bar{X}_{[\beta(\ell)]}} \right) \right)
\]

30
Lemma A-2: \( |x| \) is uniformly integrable in \( H_M(x) = \frac{F}{X_{[i]}}(x) \) with

\[
\mu = (-M, \ldots, -M, \mu_{[i]}, \ldots, \mu_{[i]}).
\]

Proof: Let \( L \) be positive. Then, by Lemma A-1,

\[
0 < \int_{|x| \geq L} |x|dH_M(x) = \int_{|x| \geq L} |x|dF/(x) X_{[i]}
\]

\[
\leq \sum_{\ell = i}^{k} \left( \frac{k}{k+1} \right) \sum_{\beta \in S_k} \left\{ \int_{|x| \geq L} |x|dF/(x) \frac{F(\beta(1)) \cdots F(\beta(\ell))}{X_{[\beta]}(x)} dx \right\}.
\]

Fix any \( \epsilon > 0 \). We will now show that there is an \( L = L(\epsilon) \) such that

the upper bound on \( \int_{|x| \geq L} |x|dH_M(x) \) is \( \leq \epsilon \) regardless of the value of \( M \). By definition, this will prove \( |x| \) is uniformly integrable

in \( H_M(x) \).

Since \( \ell = i, i+1, \ldots, k \), and since the \( i-1 \) populations have means

\(-M\) while \( k-i+1 \) have means \( \mu_{[i]} \), for any fixed \( \ell \) and \( \beta \) at

least one of \( X_{[\beta]}(1), \ldots, X_{[\beta]}(\ell) \) is associated with a population with

mean \( \mu_{[i]} \).

Let us consider the terms which are summed in the upper bound on

\( \int_{|x| \geq L} |x|dH_M(x) \), a typical one of which is

31
\[
T(\ell, \beta, j) = \frac{(k)}{k!} \int_{|x| \geq L} |x|^k \frac{F_{\overline{X}}(x)}{F_{\overline{X}(j)}} \frac{F_{\overline{X}(l)}}{F_{\overline{X}(j)}} \ dx.
\]

**Case 1.** \(\overline{X}(j)\) comes from a population with mean \(\mu[i]\). Then

\[
T(\ell, \beta, j) \leq \frac{(k)}{k!} \int_{|x| \geq L} |x|^k \frac{F_{\overline{X}}(x)}{F_{\overline{X}(j)}} \ dx
\]

and, since \(\overline{X}(j)\) is \(N(\mu[i], \sigma^2/n)\), it is clear that for \(L \geq L_1(\ell, \beta, j, \varepsilon)\) we have \(T(\ell, \beta, j) < \frac{\varepsilon}{(k-i+1)k!} \).

**Case 2.** \(\overline{X}(j)\) comes from a population with mean \(-M\). Then one of \(\overline{X}(l), \ldots, \overline{X}(j)\) (but not \(\overline{X}(j)\)) comes from a population with mean \(\mu[i]\); call it \(\overline{X}_o\). Then

\[
T(\ell, \beta, j) \leq \frac{(k)}{k!} \int_{|x| \geq L} |x|^k \frac{F_{\overline{X}}(x)}{F_{\overline{X}(j)}} \ dx + \frac{(k)}{k!} \int_{|x| \leq -L} |x|^k \frac{F_{\overline{X}}(x)}{F_{\overline{X}(j)}} \ dx.
\]

Since \(\overline{X}(j)\) is \(N(-M, \sigma^2/n)\), it is clear that for \(L \geq L_2(\ell, \beta, j, \varepsilon)\) the first term is \(< \frac{\varepsilon}{2(k-i+1)k!} \) uniformly in \(M\).

Now, since \(\overline{X}_o\) is \(N(\mu[i], \sigma^2/n)\), for \(x < -|\mu[i]|\) (so that \(-x + \mu[i] > 0\))
\[ \mathbb{P}_{\bar{X}_{\beta_o}}(x) = \mathbb{P}(-\bar{X}_{\beta_o} - \mu [i] \leq x) = \left[ \frac{-\bar{X}_{\beta_o} - \mu [i]}{\sigma / \sqrt{n}} \leq \frac{x - \mu [i]}{\sigma / \sqrt{n}} \right] = \phi \left( \frac{x - \mu [i]}{\sigma / \sqrt{n}} \right) \]

\[ = 1 - \phi \left( \frac{-x - \mu [i]}{\sigma / \sqrt{n}} \right) \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{-x - \mu [i]}{\sigma / \sqrt{n}} \right)^2} \left( \frac{1}{\sigma / \sqrt{n}} \right) \frac{1}{\sigma / \sqrt{n}} \]

by the result that, for \( y \geq 0 \), \( 1 - \phi(y) \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y^2} \frac{1}{y} \). Thus, for \( L \geq 2 |\mu [i]| \),

\[ \int_{x \leq -L} |x| f_{\bar{X}_{\beta}(i)}(x) \mathbb{P}_{\bar{X}_{\beta_o}}(x) dx \]

\[ \leq \frac{\sigma / \sqrt{n}}{\sqrt{2\pi}} \int_{x \leq -L} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x + \mu [i]}{\sigma / \sqrt{n}} \right)^2} \frac{1}{\sigma / \sqrt{n}} e^{-\frac{1}{2} \left( \frac{-x - \mu [i]}{\sigma / \sqrt{n}} \right)^2} \]

\[ \leq \frac{\sigma / \sqrt{n}}{\sqrt{2\pi}} \int_{x \leq -L} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu [i]}{\sigma / \sqrt{n}} \right)^2} dx = \sigma \sqrt{2/(n\pi)} \mathbb{P}(-\bar{X}_{\beta_o} \leq -L) \]

Since \( \bar{X}_{\beta_o} \) is \( N(\mu [i], \sigma^2 / n) \), it is clear that for \( L \geq L_3(\ell, \beta, j, \mu [i], \epsilon) \), the second term of \( T(\ell, \beta, j) \) is \( \frac{1}{2} \frac{\epsilon}{(k-1)k!} \), so that for
$L \geq L_k(\ell, \beta, j, \mu[i], \varepsilon) = \max(L_2, L_3)$ we have $T(\ell, \beta, j) < \frac{\varepsilon}{(k-i+1)k!k}$ uniformly in $M$.

Using Case 1 and Case 2, since the bound on $\int_{|x| \geq L} |x| dH_M(x)$ involves $<(k-i+1)k!k$ terms, we have (uniformly in $M$)

$$|x| \int_{|x| \geq L} |x| dH_M(x) \leq \varepsilon.$$

**Lemma A-3:** $|x|$ is uniformly integrable in $J_M(x) = F_{\overline{X}[i]}(x)$ with $i$ terms $k-i$ terms

$$\mu = (\mu[i], \ldots, \mu[i], M, \ldots, M).$$

**Proof:** Let $L$ be positive. Now,

$$0 \leq \int_{|x| \geq L} |x| dJ_M(x) = \int_{|x| \geq L} |x| dF_{\overline{X}[i]}(x).$$

Fix $\varepsilon > 0$. By definition, to prove that $|x|$ is uniformly integrable in $J_M(x)$, it is sufficient to show that there exists an $L = L(\varepsilon)$ such that $|x| \int_{|x| \geq L} |x| dJ_M(x) \leq \varepsilon$ for all $M$.

For $M > |\mu[i]|$,

$$J_M(x) = F_{\overline{X}[i]}(x) \text{ with } \mu = (\mu[i], \ldots, \mu[i], M, \ldots, M)$$

$$\leq F_{\overline{X}[i]}(x) \text{ with } \mu = (-M, \ldots, -M, \mu[i], \ldots, \mu[i])$$

$$= H_M(x).$$
Define two d.f.'s
\[ F(x) = \begin{cases} 
1 & \text{if } x \geq -L \\
J_M(x) & \text{if } x < -L
\end{cases} \quad \text{and} \quad G(x) = \begin{cases} 
1 & \text{if } x \geq -L \\
H_M(x) & \text{if } x < -L
\end{cases} \]

Then
\[ \int_{-\infty}^{L} x dF(x) \geq \int_{-\infty}^{L} x dG(x) \]
\[ \int_{-\infty}^{-L} x dJ_M(x) + L(1-J_M(-L)) \geq \int_{-\infty}^{-L} x dH_M(x) - L(1-H_M(-L)) \]
\[ 0 \geq \int_{-\infty}^{-L} x dJ_M(x) \geq \int_{-\infty}^{-L} x dH_M(x) + L[H_M(-L)-J_M(-L)] . \]

Now, since \( H_M(-L) \geq J_M(-L) \) and since \( \int_{-\infty}^{-L} x dH_M(x) \to 0 \) uniformly in \( M \) by Lemma A-2, we find that
\[ 0 \geq \int_{-\infty}^{-L} x dJ_M(x) \to 0 \] uniformly in \( M \).

Thus, there is (for any fixed \( \mu_{[1]} \) and \( L_{1}(\epsilon) \)) such that for
\( L > L_{1}(\epsilon) \) we have \( \int_{-\infty}^{-L} x dJ_M(x) < \epsilon/2 \) uniformly in \( M \).

Take \( L > L_{1}(\epsilon) \). By Theorem (2.2) and Theorem (2.3), we have
\[
\mu_{[1]} + (\sigma/\sqrt{n})h_{L}(\psi) \geq \sup_{\{\hat{\mu}_{[1]} \in \Omega_{0}(\mu_{[1]}) : \mu = (\mu_{[1]}, \ldots, \mu_{[1]}, M, \ldots, M) \}} \lim_{M \to \infty} \int_{-\infty}^{L} x dJ_M(x) = \lim_{M \to \infty} \left( \int_{-\infty}^{L} x dJ_M(x) + \int_{-L}^{L} x dJ_M(x) + \int_{L}^{\infty} x dJ_M(x) \right) \]
\[ \geq -\epsilon/2 + \lim_{M \to \infty} \int_{-L}^{L} x dJ_M(x) + \lim_{M \to \infty} \int_{L}^{\infty} x dJ_M(x) \]
\[ = -\epsilon/2 + \int_{-L}^{L} x dJ_M(x) + \lim_{M \to \infty} \int_{L}^{\infty} x dJ_M(x) . \]
The last step follows from the Helly-Bray Lemma. Since (as shown in Theorem (2.3))

\[ \int_{-\infty}^{\infty} x \, d\bar{J}_M(x) = \mu_{[1]} + (\sigma/\sqrt{n}) h_1(\phi), \]

for \( L > L_2(\epsilon) \) we have \( \int_{-L}^{L} x \, d\bar{J}_M(x) \) within \( \epsilon/2 \) of \( \mu_{[1]} + (\sigma/\sqrt{n}) h_1(\phi) \). Thus, if \( L > \max(L_1, L_2) \) then

\[
\mu_{[1]} + (\sigma/\sqrt{n}) h_1(\phi) \geq -\frac{\epsilon}{2} - \frac{\epsilon}{2} + \mu_{[1]} + (\sigma/\sqrt{n}) h_1(\phi) + \lim_{M \to \infty} \int_{-L}^{L} x \, d\bar{J}_M(x).
\]

\[
\epsilon \geq \lim_{M \to \infty} \int_{-L}^{L} x \, d\bar{J}_M(x).
\]

Thus, there is an \( L = L_2(\epsilon) \) such that \( |x| \geq L \), \( |x|d\bar{J}_M(x) \leq \epsilon \) regardless of the value of \( M \).
References


35. Teichroew, D. (1956): Tables of expected values of order
statistics and products of order statistics for samples of
size twenty and less from the normal distribution. *Ann. Math.
Stat.*, 27, 410-426.

36. Tippett, L. H. C. (1925): On the extreme individuals and the
range of samples taken from a normal population. *Biometrika,
17*, 364-387.