DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 60

February 1969

ESTIMATION OF ORDERED PARAMETERS

by

Edward J. Dudewicz

Prepared under Cornell University contracts

Nonr-401(53), Office of Naval Research, and

GP-7798, National Science Foundation,

and

University of Rochester contracts

N00014-68A-0091, Office of Naval Research, and

GP-8958, National Science Foundation

TABLE OF CONTENTS

---

* A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Field of Operations Research, Cornell University, January 1969.

CHAPTER 1. INTRODUCTION

1.1. ABSTRACT

Suppose there are given $k \geq 2$ populations $\pi_1,\ldots,\pi_k$; observations from population π_i are normally distributed with unknown mean μ_i and common (known or unknown) variance $\sigma_i^2 = \sigma^2$ $(i = 1,\ldots,k)$. Let $\mu_{[1]}\leq\cdots\leq\mu_{[k]}$ denote the ranked values of $\mu_1,\ldots,\mu_k$. In this thesis we assume throughout 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 (although we vary the distribution from normality) and consider the problem: estimate some (or all) of $\mu_{[1]},\ldots,\mu_{[k]}$ based on $\overline{X}_1,\ldots,\overline{X}_k$, where $\overline{X}_1,\ldots,\overline{X}_k$ come from use of the following single-stage rule: Take n independent vectors $\underline{X}_j = (X_{1j},\ldots,X_{kj})$, $j = 1,\ldots,n$ (X_{ij} denotes the j<u>th</u> observation from π_i); for each population compute $\overline{X}_i = \sum_{j=1}^{n} X_{ij}/n$ $(i = 1,\ldots,k)$, and base the terminal decision on $\overline{X}_1,\ldots,\overline{X}_k$. (The fixed number n of vectors required depends on the particular problem.) This rule has been used in many instances of statistical decision problems. Applications to ranking and selection problems are noted.

Let $\overline{X}_{[1]}\leq\cdots\leq\overline{X}_{[k]}$ denote the ranked $\overline{X}_1,\ldots,\overline{X}_k$. A natural point estimator of $\mu_{[i]}$ is $\overline{X}_{[i]}$ $(1 \leq i \leq k)$, and its bias is studied when observations from π_i have density $f(x - \theta_i)$, $x\varepsilon\mathcal{R}$, where the location parameter θ_i is unknown $(i = 1,\ldots,k)$ and $E_f \equiv \int_{-\infty}^{\infty} xf(x)dx<\infty$. Upper and lower bounds, U_i and L_i, are derived for $E_{\underline{\mu}}\overline{X}_{[i]}$ $(1 \leq i \leq k)$ ($\underline{\mu}$ denotes the vector $(\mu_1,\ldots,\mu_k)$), and condition S(i), sufficient to imply that $\overline{X}_{[i]}$ is asymptotically unbiased as $n \to \infty$, is obtained. When $i = k$

(i = 1), $U_i(L_i)$ is the supremum (infimum) of $E_\mu \overline{X}_{[i]}$. It is shown that uniform integrability condition $C_1(i)$ implies $S(i)$. Condition C_2 (which holds if, e.g., $\int_{-\infty}^{\infty} x^2 f(x)dx<\infty$) also implies $S(i)$. The relationship is $C_2 \Leftrightarrow \{C_1(1),\ldots,C_1(k)\}$. The minimax|bias|estimator of type $\overline{X}_{[i]} + a$ is found for certain cases. These results are applied to the case where $f(\circ)$ is the normal density, and a uniform integrability argument shows that U_i and L_i are the supremum and infimum. It is noted that, for the location parameter case, $\overline{X}_{[i]}$ is strongly consistent for $\mu_{[i]}$ $(1 \leq i \leq k)$; applications are noted. Bounds are obtained on the mean squared error $E_\mu(\overline{X}_{[i]} - \mu_{[i]})^2$ $(1 \leq i \leq k)$, also for the location parameter case. For the case when $f(\circ)$ is the normal density these bounds are evaluated, and intervals in which the supremum and infimum of the mean squared error lie are determined.

Maximum likelihood estimation of $(\mu_{[1]},\ldots,\mu_{[k]})$ based on $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$ is studied. It is shown that any critical point for this problem is a solution of a system with derivatives taken for $\mu \in \Omega(\neq) = \{\mu: \mu_{[1]} \neq \mu_{[2]} \neq \cdots \neq \mu_{[k]}\}$ if boundary points are considered solutions and that $(\overline{X},\ldots,\overline{X})$ with $\overline{X} = (\overline{X}_{[1]} + \cdots + \overline{X}_{[k]})/k$ is a critical point. The nature of $(\overline{X},\ldots,\overline{X})$ is completely determined, and w.p. $\to 1$ as $n \to \infty$ it is a saddle point (unless $\mu_{[1]} = \cdots = \mu_{[k]}$, in which case it may be a relative maximum). Some results on the form of the maximum likelihood estimator (MLE) for $k \geq 2$ are given, while for $k = 2$ the MLE is found explicitly. MLE's for non-1-1 functions are discussed, and a concept of iterated MLE's (IMLE's) is introduced and discussed. The generalized MLE (GMLE) introduced by Weiss and Wolfowitz, which has a certain optimality property, is found to be $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$, which has desirable large sample concentration. It is shown that there is not just one GMLE but rather a whole class of GMLE's, and for $k = 2$ the MLE is shown to be in this class along with $\overline{X}_{[1]}$, $\overline{X}_{[2]}$. It is shown that for our problem (and others) a GMLE (if one exists) is equivalent to the maximum probability estimator (MPE) introduced by Weiss and Wolfowitz, if the latter is "good."

Confidence interval estimation of $\mu_{[1]}, \ldots, \mu_{[k]}$ is discussed, and upper and lower intervals on $\mu_{[i]}$ $(1 \leq i \leq k)$ are found, along with their maximal overprotection, for location parameter populations. Generalizing a result of Fraser, it is shown that exact upper intervals satisfying mild conditions do not exist.

1.2. OUTLINE OF THE THESIS

In Section 1.1, we have given an overview of the problem considered below and of the results obtained, and in Section 1.3 we make specific definition of the problem considered and introduce various notations. In the present section we outline briefly the contents of the various chapters.

Chapter 2. The problem of point estimation is considered for a location parameter family, and the bias of certain natural estimators is studied; a minimax estimator is found for certain cases. These general results are examined in the normal density case, for which additional results are obtained.

Chapter 3. The problem of strong consistency is considered for a location parameter family, and applications to value-estimation and Bayesian statistics are noted.

Chapter 4. For a location parameter family, bounds are obtained on the mean squared error of certain natural estimators. These results are examined in the normal density case, and additional bounds on the infimum and supremum of the mean squared error lead to intervals on these two quantities.

Chapter 5. Maximum likelihood estimators are studied for the normal density case. A concept of iterated maximum likelihood estimators is introduced and discussed. Generalized maximum likelihood estimators and maximum probability estimators are found.

Chapter 6. The problem of interval estimation is formulated. For a location parameter family upper and lower intervals are found, and it is shown that exact upper intervals satisfying mild conditions do not exist.

1.3. PROBLEM DEFINITION AND NOTATION

Consider the set-up

(1.3.1) Given $k(\geq 2)$ populations $\pi_1,\ldots,\pi_k$ such that observations from population π_i are normally distributed with unknown mean μ_i and common (known or unknown) variance $\sigma_i^2 = \sigma^2$ $(i = 1,\ldots,k)$,

and the following rule.

(1.3.2) RULE: Take n independent vectors $\underline{X}_j = (X_{1j},\ldots,X_{kj})$, $j = 1,\ldots,n$, where X_{ij} denotes the jth observation from the ith population π_i. For each population form the sample mean

(1.3.3) $$\overline{X}_i = \sum_{j=1}^{n} X_{ij}/n \qquad (i = 1,\ldots,k),$$

and base the terminal decision solely on the statistics $\overline{X}_1,\ldots,\overline{X}_k$.

(This rule has been utilized under set-up (1.3.1) in many instances of statistical decision problems.) Make the

(1.3.4) DEFINITION: Let $\mu_{[1]}\leq\cdots\leq\mu_{[k]}$ denote the ranked values of $\mu_1,\ldots,\mu_k$.

We assume throughout 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 (although we vary the distributional requirements from those of set-up (1.3.1)) and consider the problem: estimate some (or all) of $\mu_{[1]},\ldots,\mu_{[k]}$ based on the statistics provided by the single-stage Rule (1.3.2).

Consideration has been devoted in the literature to what are called "ranking and selection" problems. Since several of the proposed procedures in that type of statistical decision problem use Rule (1.3.2) (e.g., those of Bechhofer (1954), Gupta (1956), (1965), and others), and since one will often wish to estimate as well as select, we will briefly describe such problems and will refer below to uses of our results in such problems.

A simple example of such a problem is that of selecting the population (or, one of the populations) associated with the ith smallest mean ($1 \leq i \leq k$); this is called one's goal. (Much more general goals have also been considered.) Typically, a probability requirement is made and a procedure is given (which tells how to sample, when to stop sampling, and what terminal decision to make). The probability requirement affects one's sample sizes, since the more stringent one's probability requirement vis-a-vis achieving the goal, the more sampling one must perform. In Rule (1.3.2), only the fixed number n of independent vectors required depends on the particular {goal, probability requirement, procedure} structure on hand. (We note that Rule (1.3.2) has some optimal properties. See Hall (1958), (1959); Bahadur and Goodman (1952); Lehmann (1966); and Eaton (1967).) Of course the various structures use the statistics in quite different manners, and not all structures use Rule (1.3.2); e.g., the nonparametric procedure of Bechhofer and Sobel (1958), the closed sequential procedure of Paulson (1964), and the open sequential procedure of Bechhofer, Kiefer, and Sobel (1968) do not.

We will make use of the following definitions and notation.

(1.3.5) DEFINITION: For any set S, let $\nu(S) \equiv$ cardinal number of S. (If S is a finite set, then $\nu(S)$ is the number of elements in S.)

(1.3.6) DEFINITION: Let $R = \{x: -\infty<x<\infty\}$ and let $R^+ = \{x: x\geq 0\}$.

DEFINITION: For $\delta \in R^+$, let $\Omega_\delta(a,b,c,\ldots) = \{(\mu_1,\ldots,\mu_k):$
(1.3.7) $\mu_{[k]} - \mu_{[k-1]} \geq \delta,\ \mu_i \in R\ (i = 1,\ldots,k)$, a,b,c,... are held fixed$\}$. (In general a,b,c,... will be several of $\mu_{[1]},\ldots,\mu_{[k]}$.)

(1.3.8) DEFINITION: Let $\omega_{LFC}(\delta) = (\mu_{[k]}-\delta,\ldots,\mu_{[k]}-\delta,\mu_{[k]})$ and $\omega_{EM}(\mu_{[k]}) = (\mu_{[k]},\ldots,\mu_{[k]})$ be vectors of k components.

DEFINITION: Let $\overline{X}_{[1]}\leq\ldots\leq\overline{X}_{[k]}$ denote the ordered $\overline{X}_i$
(1.3.9) $(i = 1,\ldots,k)$. (We disregard the possibility of ties, which occur w.p. 0 in the cases considered below.)

DEFINITION: If a random variable (r.v.) X is normally distributed with mean μ and variance σ^2, we shall say X is $N(\mu,\sigma^2)$.

Denote the N(0,1) distribution function (d.f.) and
(1.3.10) density function (fr.f.) by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively; i.e., let

$$\Phi(x) = \int_{-\infty}^{x}\phi(y)dy \qquad (x \in R),$$

$$\phi(y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2} \qquad (y \in R).$$

DEFINITION: Let F and f be the respective d.f. and fr.f. of observations from an arbitrary univariate location parameter family; i.e.,

(1.3.11) $$F(x) = \int_{-\infty}^{x} f(y-\theta)dy \qquad (x \in \mathcal{R}), \text{ and}$$

$$f \text{ has the form } f(y-\theta) \qquad (y \in \mathcal{R}),$$

where θ is fixed, $\theta \in \Theta \subseteq \mathcal{R}$.

(1.3.12) DEFINITION: $\Omega(\neq) = \{\mu\colon \mu_{[1]} \neq \mu_{[2]} \neq \cdots \neq \mu_{[k]}\}$.

(1.3.13) DEFINITION: If $\mu \in \Omega(\neq)$, let $\overline{X}_{(i)}$ denote the sample mean produced by the population associated with $\mu_{[i]}$ $(i = 1,\ldots,k)$.

(1.3.14) DEFINITION: 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

$$(0 \equiv i_o < 1 \leq i_1 < i_2 < \cdots < i_{\ell-1} \leq k-1 < i_\ell \equiv k),$$

and we let

$$\overline{X}_{(i_j+1)} \leq \overline{X}_{(i_j+2)} \leq \cdots \leq \overline{X}_{(i_{j+1}-1)} \leq \overline{X}_{(i_{j+1})}$$

be the ranked values of the sample means from the population(s) associated with parameter $\mu_{[i_{j+1}]}$ $(j = 0,\ldots,\ell-1)$.

(1.3.15) DEFINITION: Let S_k be the symmetric group on k elements, i.e., $\{\alpha\colon \alpha = (\alpha(1),\ldots,\alpha(k))$ is a permutation of $(1,\ldots,k)\}$.

CHAPTER 2. POINT ESTIMATION: BIAS

2.1. BIAS OF A NATURAL ESTIMATOR OF $\mu_{[i]}$ $(1 \leq i \leq k)$ FOR A LOCATION PARAMETER FAMILY

Consider the set-up

(2.1.1) Given $k(\geq 2)$ populations $\pi_1,\ldots,\pi_k$ such that observations from population π_i have fr.f. $f(x-\theta_i)$, $x \in R$, where the location parameter θ_i is unknown $(i = 1,\ldots,k)$.

We make the

(2.1.2) <u>ASSUMPTION</u>: The fr.f. f is such that $E_f \equiv \int_{-\infty}^{\infty} xf(x)dx < \infty$,

so that we may talk of $\mu_1,\ldots,\mu_k$ (or of $\mu_{[1]},\ldots,\mu_{[k]}$). Denote the ranked values of the location parameters $\theta_1,\ldots,\theta_k$ by $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$. Then since

$$(2.1.3)\qquad \begin{aligned} \int_{-\infty}^{\infty} xf(x-a)dx &= \int_{-\infty}^{\infty}(x+a-b)f(x-b)dx = \int_{-\infty}^{\infty} xf(x-b)dx-(b-a) \\ &< \int_{-\infty}^{\infty} xf(x-b)dx \qquad\qquad (a<b;\ a,b \in \Theta), \end{aligned}$$

the population associated with $\mu_{[i]}$ is precisely the population associated with $\theta_{[i]}$ $(i = 1,\ldots,k)$. Also,

$$(2.1.4)\qquad \int_{-\infty}^{\infty} xf(x-\theta)dx = \int_{-\infty}^{\infty} xf(x)dx + \theta = E_f + \theta$$

where E_f is the mean of f when $\theta = 0$.

We will now study estimation of $\mu_{[i]}$ $(1 \leq i \leq k)$ when set-up (2.1.1) obtains, Rule (1.3.2) is used, and the pairing of $\pi_1,\ldots,\pi_k$ with $\mu_{[1]},\ldots,\mu_{[k]}$ is completely unknown (see Chapter 1). Denote the densities of $X_{ij}-\theta_i$ and X_{ij} by $f_{X_{ij}-\theta_i}$ and $f_{X_{ij}}$, respectively. Since

$$(2.1.5) \qquad f_{X_{ij}-\theta_i}(y) = f_{X_{ij}}(y+\theta_i) = f\,((y+\theta_i)-\theta_i) = f(y),$$

it follows that $X_{ij}-\theta_i$ does not depend on θ_i $(i = 1,\ldots,k)$.

$$(2.1.6) \qquad \underline{\text{DEFINITION}}: \quad G_n(y|f) = P[\{(X_{i1}-\theta_i)+\ldots+(X_{in}-\theta_i)\}/n \le y],$$
$$g_n(y|f) = \frac{d}{dy} G_n(y|f).$$

For $i = 1,\ldots,k$,

$$(2.1.7) \qquad P[\overline{X}_i \le x] = P[X_{i1}+\ldots+X_{in} \le nx] = P[(X_{i1}-\theta_i)+\ldots+(X_{in}-\theta_i) \le n(x-\theta_i)]$$
$$= G_n(x-\theta_i|f).$$

We now determine several d.f.'s and fr.f.'s which we will use in later sections.

$$(2.1.8) \qquad \underline{\text{THEOREM}}: \quad F_{\overline{X}_{[k]}}(x) = \prod_{i=1}^{k} G_n(x-\theta_i|f) \qquad (x \in \mathcal{R}),$$
$$f_{\overline{X}_{[k]}}(x) = \sum_{j=1}^{k} \left[\left(\prod_{\substack{i=1 \\ i \ne j}}^{k} G_n(x-\theta_i|f) \right) g_n(x-\theta_j|f) \right] \quad (x \in \mathcal{R}).$$

Proof:

$$F_{\overline{X}_{[k]}}(x) = P[\max(\overline{X}_1,\ldots,\overline{X}_k) \le x] = P[\overline{X}_1 \le x,\ldots,\overline{X}_k \le x]$$
$$= P[\overline{X}_1 \le x]\ldots P[\overline{X}_k \le x] = \prod_{i=1}^{k} G_n(x-\theta_i|f).$$

The expression for $f_{\overline{X}_{[k]}}(\circ)$ follows upon differentiation of $F_{\overline{X}_{[k]}}(\circ)$, utilizing the chain rule (see, e.g., Kaplan (1952), p. 86, (2-26)) and the fact that $G_n'(y|f) = \frac{d}{dy} G_n(y|f) \equiv g_n(y|f)$ (see, e.g., Fisz (1963), p. 35; or Parzen (1960), p. 169).

COROLLARY: $E_\mu \overline{X}_{[k]} = \int_{-\infty}^{\infty} x f_{\overline{X}_{[k]}}(x)\,dx$

(2.1.9)
$$= \sum_{j=1}^{k} \int_{-\infty}^{\infty} x\, g_n(x-\theta_j|f) \left(\prod_{\substack{i=1\\ i\neq j}}^{k} G_n(x-\theta_i|f) \right) dx.$$

A possible estimator of $\mu_{[i]}$ when set-up (2.1.1) obtains and Rule (1.3.2) is used is $\overline{X}_{[i]}$ $(i = 1,\ldots,k)$; we now study its expectation and bias. (Although quantities such as $E\mu\overline{X}_{[k]}$ depend on the unknown $\mu \in \Omega_0$, this dependence will sometimes be suppressed; e.g., we will write $E\overline{X}_{[k]}$ for $E\mu\overline{X}_{[k]}$.)

LEMMA: If X and Y are independent r.v.'s with

(2.1.10) $$F_X(x) = P[X \le x] \le P[Y \le x] = F_Y(x) \qquad (x \in R),$$

then $EX \geq EY$.

Proof: A geometrical proof of this lemma can easily be given using, e.g., Exercise 2.5 of Parzen (1960), pp. 211-212, "A geometrical interpretation of the mean of a probability law."

THEOREM: For $i = 1,\ldots,k$ and $x \in R$, $F_{\overline{X}_{[i]}}(x)\downarrow$ as $\mu_\ell\uparrow$

(2.1.11)

$(\ell = 1,\ldots,k)$.

Proof: Fix $\ell (1 \le \ell \le k)$. For $i = 1,\ldots,k$ and $x \in R$,

$F_{\overline{X}_{[i]}}(x) = P_\mu[\overline{X}_{[i]} \le x] = P_\mu$[The ith smallest of $\overline{X}_1,\ldots,\overline{X}_k$ is $\le x$]

$= P_\mu$[At least i of $\overline{X}_1,\ldots,\overline{X}_k$ are $\le x$]

$= P_\mu$[$\overline{X}_\ell \le x$ and at least i-1 of $\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k$ are $\le x$]

$+ P_\mu$[$\overline{X}_\ell > x$ and at least i of $\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k$ are $\le x$]

$= P_\mu[\overline{X}_\ell \le x] P_\mu$[At least i-1 of $\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k$ are $\le x$]

$+ \left(1-P_\mu[\overline{X}_\ell \le x]\right) P_\mu$[At least i of $\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k$ are $\le x$]

$$= G_n(x-\theta_\ell|f)P_\mu[\text{At least } i\text{-}1 \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x]$$

$$+[1-G_n(x-\theta_\ell|f)]P_\mu[\text{At least } i \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x].$$

Therefore,

$$\frac{d}{d\mu_\ell} F_{\bar{X}_{[i]}}(x) = \frac{d}{d\theta_\ell} F_{\bar{X}_{[i]}}(x)\,\frac{d\theta_\ell}{d\mu_\ell}$$

$$= -g_n(x-\theta_\ell|f)P_\mu[\text{At least } i\text{-}1 \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x]$$

$$+g_n(x-\theta_\ell|f)P_\mu[\text{At least } i \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x],$$

which is ≤ 0 iff

$$P_\mu[\text{At least } i \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x]$$

$$\leq P_\mu[\text{At least } i\text{-}1 \text{ of } \bar{X}_1,\ldots,\bar{X}_{\ell-1},\bar{X}_{\ell+1},\ldots,\bar{X}_k \text{ are } \leq x].$$

(2.1.12) DEFINITION: For $\ell = 1,2,3,\ldots$ let $h_\ell(g_n)$ be the expectation of the maximum of ℓ independent r.v.'s each having fr.f. $g_n(x)$; and let $h'_\ell(g_n)$ be the expectation of the minimum of ℓ independent r.v.'s each having fr.f. $g_n(x)$, i.e.,

$$h_\ell(g_n) = \int_{-\infty}^{\infty} y\ell[G_n(y)]^{\ell-1} g_n(y)dy,$$

$$h'_\ell(g_n) = \int_{-\infty}^{\infty} y\ell[1-G_n(y)]^{\ell-1} g_n(y)dy.$$

The following is well-known:

(2.1.13) LEMMA: If $g_n(x)$ is symmetric about $x = 0$ then

$$h'_\ell(g_n) = -h_\ell(g_n).$$

(2.1.13a) THEOREM: If $G_n(x) < 1$ for all x, then $\lim_{\ell\to\infty} h_\ell(g_n) = +\infty$.

Proof: By (2.1.12),

$$\ell\int_0^\infty y[G_n(y)]^{\ell-1}g_n(y)dy \geq h_\ell(g_n) = \ell\int_{-\infty}^\infty y[G_n(y)]^{\ell-1}g_n(y)dy$$
$$= \ell\int_0^\infty y[G_n(y)]^{\ell-1}g_n(y)dy + \ell\int_{-\infty}^0 y[G_n(y)]^{\ell-1}g_n(y)dy$$
$$\geq \ell\int_0^\infty y[G_n(y)]^{\ell-1}g_n(y)dy + \ell[G_n(o)]^{\ell-1}\int_{-\infty}^0 yg_n(y)dy.$$

Thus, since $\int_{-\infty}^\infty yg_n(y)dy < \infty$ and $\lim_{\ell\to\infty} \ell a^\ell = 0$ $(0<a<1)$, by taking the limit as $\ell\to\infty$ we obtain

$$\lim_{\ell\to\infty} h_\ell(g_n) = \lim_{\ell\to\infty} \ell\int_0^\infty y[G_n(y)]^{\ell-1}g_n(y)dy.$$

However, for any $M > 0$,

$$0 \leq \ell\int_0^M y[G_n(y)]^{\ell-1}g_n(y)dy \leq \ell\int_0^M y[G_n(M)]^{\ell-1}g_n(y)dy$$
$$= \ell[G_n(M)]^{\ell-1}\int_0^M yg_n(y)dy \to 0 \quad \text{as } \ell\to\infty.$$

Choosing $M > 1$, and since $G_n(M) < 1$ for any M, we find that

$$\lim_{\ell\to\infty} h_\ell(g_n) = \lim_{\ell\to\infty} \ell\int_M^\infty y[G_n(y)]^{\ell-1}g_n(y)dy \geq \lim_{\ell\to\infty} M\ell\int_M^\infty [G_n(y)]^{\ell-1}g_n(y)dy$$
$$= \lim_{\ell\to\infty} M[G_n(y)]^\ell\Big|_M^\infty = \lim_{\ell\to\infty} M\{1-[G_n(M)]^\ell\} = M.$$

Since $M > 1$ was arbitrary, the theorem follows.

<u>LEMMA</u>: If $Y_1,\ldots,Y_\ell$ are independent r.v.'s each having d.f. $G_n(x-\theta)$, then

(2.1.14) $$E\max\{Y_1,\ldots,Y_\ell\} = \theta + h_\ell(g_n),$$
$$E\min\{Y_1,\ldots,Y_\ell\} = \theta + h'_\ell(g_n).$$

<u>Proof</u>: Since Y_i has d.f. $G_n(x-\theta)$, $Y_i - \theta$ has d.f. $G_n(x)$ $(i = 1,\ldots,\ell)$ by (2.1.7). Thus,

$$E\,{\max \atop \min}\{Y_1,\ldots,Y_\ell\} = E\,{\max \atop \min}\{(Y_1-\theta)+\theta,\ldots,(Y_\ell-\theta)+\theta\}$$
$$= \theta + E\,{\max \atop \min}\{Y_1-\theta,\ldots,Y_\ell-\theta\} = \theta + \begin{cases} h_\ell(g_n) \\ h'_\ell(g_m). \end{cases}$$

<u>THEOREM</u>: For any i $(1 \leq i \leq k)$

$$\sup\{E_\mu \overline{X}_{[i]} : \mu \in \Omega_o(\mu_{[i]})\}$$

$$\leq E_{(\mu_{[1]},\ldots,\mu_{[i-1]},\mu_{[i]},\mu_{[i+1]},\ldots,\mu_{[k]})} \overline{X}_{[i]}$$

(2.1.16)

$$= E_{(\underbrace{\mu_{[i]},\ldots,\mu_{[i]},\mu_{[i]}}_{i \text{ times}},+\infty,\ldots,+\infty)} \overline{X}_{[i]}$$

$$= \theta_{[i]} + h_i(g_n) = \mu_{[i]} - E_f + h_i(g_n),$$

and

(2.1.15)

$$\inf\{E_\mu \overline{X}_{[i]} : \mu \in \Omega_o(\mu_{[i]})\}$$

$$\geq E_{(\mu_{[1]},\ldots,\mu_{[i-1]},\mu_{[i]},\mu_{[i+1]},\ldots,\mu_{[k]})} \overline{X}_{[i]}$$

(2.1.17)

$$= E_{(-\infty,\ldots,-\infty,\underbrace{\mu_{[i]},\mu_{[i]},\ldots,\mu_{[i]}}_{k-i+1 \text{ times}})} \overline{X}_{[i]}$$

$$= \theta_{[i]} + h'_{k-i+1}(g_n) = \mu_{[i]} - E_f + h'_{k-i+1}(g_n),$$

where the configurations of the vector $(\mu_{[1]},\ldots,\mu_{[k]})$ which involve values $\pm\infty$ are viewed as a situation eliminating the populations with mean values $\pm\infty$ from contention for ith highest sample mean. (The case $i = k$ in (2.1.16) and the case $i = 1$ in (2.1.17) involve no such eliminations.)

<u>Proof</u>: By Lemma (2.1.10) and Theorem (2.1.11), we increase $E_\mu \overline{X}_{[i]}$ by raising μ_j $(i,j = 1,\ldots,k)$. Now,

$$\theta_{[i]} + h'_{k-i+1}(g_n)$$

$$= E_{\mu = (\mu_{[1]},\ldots,\mu_{[i-1]},\mu_{[i]},\mu_{[i]},\ldots,\mu_{[i]})} \{\text{Smallest of } (\overline{X}_{(i)} - \theta_{[i]}) + \theta_{[i]},\ldots,(\overline{X}_{(k)} - \theta_{[i]}) + \theta_{[i]}\}$$

$$\leq E_{\mu=(\mu_{[1]},\ldots,\mu_{[i-1]},\mu_{[i]},\mu_{[i]},\ldots,\mu_{[i]})}\{\underline{\text{ith}} \text{ smallest of } \overline{X}_{(1)},\ldots,\overline{X}_{(i-1)},\overline{X}_{(i)},\ldots,\overline{X}_{(k)}\}$$

$$\leq E_{\mu}\{\underline{\text{ith}} \text{ smallest of } \overline{X}_{(1)},\ldots,\overline{X}_{(i-1)},\overline{X}_{(i)},\ldots,\overline{X}_{(k)}\} = E_{\mu}\overline{X}_{[i]}$$

$$= E_{\mu}\{\underline{\text{ith}} \text{ smallest of } \overline{X}_{(1)},\ldots,\overline{X}_{(i)},\overline{X}_{(i+1)},\ldots,\overline{X}_{(k)}\}$$

$$\leq E_{\mu=(\mu_{[i]},\ldots,\mu_{[i]},\mu_{[i]},\mu_{[i+1]},\ldots,\mu_{[k]})}\{\underline{\text{ith}} \text{ smallest of } \overline{X}_{(1)},\ldots,\overline{X}_{(i)},\overline{X}_{(i+1)},\ldots,\overline{X}_{(k)}\}$$

$$\leq E_{\mu=(\mu_{[i]},\ldots,\mu_{[i]},\mu_{[i]},\mu_{[i+1]},\ldots,\mu_{[k]})}\{\text{Largest of } (\overline{X}_{(1)}-\theta_{[i]})+\theta_{[i]},\ldots,(\overline{X}_{(i)}-\theta_{[i]})+\theta_{[i]}\}$$

$$= \theta_{[i]}+h_i(g_n).$$

(Note that for our purposes here, the ties in Definition (1.3.14) should be broken in an arbitrary manner.) Upon taking the desired supremum and infimum, the theorem follows.

<u>COROLLARY</u>: For any i ($1 \leq i \leq k$)

(2.1.19) $\quad \mu_{[i]} + (h'_{k-i+1}(g_n)-E_f) \leq E_{\mu}\overline{X}_{[i]} \leq \mu_{[i]} + (h_i(g_n)-E_f).$

Thus, (1) $\overline{X}_{[i]}$ is asymptotically unbiased (as $n\to\infty$) as an estimator of $\mu_{[i]}$ if

(2.1.18) (2.1.20) $\quad \begin{cases} h_i(g_n)\to E_f & \text{as } n\to\infty \text{ , and} \\ h'_{k-i+1}(g_n)\to E_f & \text{as } n\to\infty \text{ ;} \end{cases}$

(2) if the left and right members of (2.1.19) are the infimum and supremum of $E_{\mu}\overline{X}_{[i]}$ (respectively) then $\overline{X}_{[i]}$ is asymptotically unbiased (as $n\to\infty$) iff (2.1.20) holds.

With Corollary (2.1.18) as motivation, we will now study the questions of (i) when (2.1.20) holds and (ii) when the inf and sup above achieve the bounds of (2.1.19).

THEOREM:

$$\mu_{[k]} \leq E_\mu \overline{X}_{[k]} \leq \sup\{E_\mu \overline{X}_{[k]} : \mu \in \Omega_o(\mu_{[k]})\} = \mu_{[k]} + (h_k(g_n)-E_f) \tag{2.1.21}$$

$$\mu_{[1]} + (h_k'(g_n)-E_f) = \inf\{E_\mu \overline{X}_{[1]} : \mu \in \Omega_o(\mu_{[1]})\} \leq E_\mu \overline{X}_{[1]} \leq \mu_{[1]}.$$

Proof: The lower bound for $E_\mu \overline{X}_{[k]}$ (the upper bound for $E_\mu \overline{X}_{[1]}$) follows from the fact that $h_1'(g_n) = E_f$ (that $h_1(g_n) = E_f$). The equality for the sup for $E_\mu \overline{X}_{[k]}$, and for the inf for $E_\mu \overline{X}_{[1]}$, follow easily from Theorem (2.1.15) and the first sentence of the proof of Theorem (2.1.15). Note that they are actually attained at $\omega_{EM}(\mu_{[k]})$ and $\omega_{EM}(\mu_{[1]})$, respectively.

From Assumption (2.1.2), it follows that independent r.v.'s with fr.f. f obey the Law of Large Numbers, so that (cf. (2.1.7)) as $n\to\infty$, for any i $(1 \leq i \leq k)$

$$\left.\begin{array}{l} G_n(y|f) \\ [G_n(y|f)]^i \\ 1-[1-G_n(y|f)]^{k-i+1} \end{array}\right\} \to G_\infty(y|f) \equiv \begin{cases} 0, y < E_f \\ 1, y \geq E_f \end{cases} \tag{2.1.22}$$

since (2.1.4) is true. Each of the convergences indicated in (2.1.22) is weak convergence; i.e., F_n converges weakly to F iff $F_n \to F$ on the continuity set of F. It is not obvious that it is then the case that (2.1.20) holds, i.e., that for any i $(1 \leq i \leq k)$,

$$\left.\begin{array}{l} h_i(g_n) = \int_{-\infty}^{\infty} y\, d_y\{[G_n(y|f)]^i\} \\ h_{k-i+1}'(g_n) = \int_{-\infty}^{\infty} y\, d_y\{1-[1-G_n(y|f)]^{k-i-1}\} \end{array}\right\} \to E_f \text{ as } n\to\infty . \tag{2.1.23}$$

If we make the following definition (cf. Loève (1963), p. 182)

DEFINITION: If $g(\cdot)$ is a continuous function and $F_n(\cdot)$ is a d.f. $(n \geq 1)$, we say $|g|$ is uniformly integrable in F_n if

(2.1.24) $\int_{|x|\geq c_m} |g| dF_n \to 0$ uniformly in n as $c_m \to \infty$ with $m \to \infty$; i.e., if (for any $\varepsilon > 0$) there is an m_o such that for $m \geq m_o$ we have

$\int_{|x|\geq c_m} |g| dF_n < \varepsilon$ for all n (where $c_m \to \infty$ as $m \to \infty$),

then we may use the following theorem (cf. Loève (1963), p. 183, Theorem A.(ii))

(2.1.25) THEOREM: If F_n converges weakly to F (a d.f.) and $|g|$ is uniformly integrable in F_n, then

$$\int g dF_n \to \int g dF$$

to immediately state the

(2.1.26) THEOREM: For any i $(1 \leq i \leq k)$, (2.1.20) holds if $|y|$ is uniformly integrable in $[G_n(y|f)]^i$ and $1-[1-G_n(y|f)]^{k-i+1}$.

Proof: This follows from (2.1.22), (2.1.23), and Theorem (2.1.25).

(2.1.27) THEOREM: If (2.1.20) holds, then $\int_{-\infty}^{\infty} y d_y\{[G_n(y|f)]^i\} \to E_f$, and then $\int_0^{\infty} y d_y\{[G_n(y|f)]^i\} \to E_f^+$, $\int_{-\infty}^{0} |y| d_y\{[G_n(y|f)]^i\} \to E_f^-$ with $E_f = E_f^+ - E_f^-$. ($E_f^- = \lim_{n\to\infty} \int_{-\infty}^{0} |y| d_y\{[G_n(y|f)]^i\}$; E_f^+ similarly.)

For any i $(1 \leq i \leq k)$, (2.1.20) holds only if (as $n \to \infty$)

$$\int_{|y|\geq M} |y| d_y\{[G_n(y|f)]^i\} \to \begin{cases} 2E_f^-, & 0 \leq E_f < M \\ 2E_f^+, & -M < E_f \leq 0. \end{cases}$$

Note that $|y|$ is uniformly integrable in $[G_n(y|f)]^i$ means

$E_f^- = 0$ if E_f is non-negative ($E_f^+ = 0$ if E_f is non-positive). A similar result holds with respect to $\{1-[1-G_n(y|f)]^{k-i+1}\}$. Note that E_f^+ and E_f^- may depend on i.

Proof: Suppose $0 < E_f < M$. By the Helly-Bray Lemma (see, e.g., Loève (1963), p. 180),

$$\left.\begin{array}{l} \int_0^M y d_y\{[G_n(y|f)]^i\} \to E_f \\ \int_{-M}^0 y d_y\{[G_n(y|f)]^i\} \to 0 \end{array}\right\} \text{ as } n\to\infty .$$

Now, letting $n\to\infty$ in

$$\int_{-\infty}^{\infty} y d_y\{[G_n(y|f)]^i\} = \int_{-\infty}^{0} y d_y\{[G_n(y|f)]^i\} + \int_{0}^{M} y d_y\{[G_n(y|f)]^i\} + \int_{M}^{\infty} y d_y\{[G_n(y|f)]^i\}$$

we obtain

$$E_f = -E_f^- + E_f + \lim_{n\to\infty} \int_M^{\infty} y d_y\{[G_n(y|f)]^i\},$$

so that $\int_M^{\infty} y d_y\{[G_n(y|f)]^i\} \to E_f^-$ as $n\to\infty$, and thus

$$\int_{|y|\geq M} |y| d_y\{[G_n(y|f)]^i\} = -\int_{-\infty}^{-M} y d_y\{[G_n(y|f)]^i\} + \int_M^{\infty} y d_y\{[G_n(y|f)]^i\}$$
$$\to (E_f^- - 0) + E_f^- = 2E_f^- \quad \text{as } n\to\infty .$$

The case $-M < E_f < 0$ follows in a similar manner. The result for $E_f = 0$ follows from the equation $E_f = E_f^+ - E_f^-$ and the Helly-Bray Lemma.

We have thus seen that although a certain uniform integrability condition is sufficient for (2.1.20) to hold (Theorem (2.1.26)), it is not clear that it is necessary for (2.1.20) to hold (Theorem (2.1.27)).

We will now exhibit a condition (simpler than that of Theorem (2.1.26)) under which (2.1.20) holds. Fix i ($1 \leq i \leq k$), let $\overline{Z}_j$ be the mean of n independent r.v.'s $Z_{j1},\ldots,Z_{jn}$ each with fr.f. $f(\cdot)$

$(j = 1,\ldots,i)$, and suppose $E\bar{Z}_{j\ell} = \mu$ (say) exists. We wish to know when (as $n\to\infty$)

$$E \begin{smallmatrix}\max\\ \min\end{smallmatrix} (\bar{Z}_1,\ldots,\bar{Z}_i) \to \mu .$$

(2.1.28) <u>THEOREM</u>: If $E|\bar{Z}_j-\mu|\to 0$ (as $n\to\infty$) $(j = 1,\ldots,i)$, then (as $n\to\infty$)

$$E \begin{smallmatrix}\max\\ \min\end{smallmatrix} (\bar{Z}_1,\ldots,\bar{Z}_i) \to \mu .$$

<u>Proof</u>:

$$\begin{smallmatrix}\max\\ \min\end{smallmatrix} (\bar{Z}_1,\bar{Z}_2) = \frac{\bar{Z}_1+\bar{Z}_2}{2} \pm \frac{|\bar{Z}_1-\bar{Z}_2|}{2},$$

so that

$$E \begin{smallmatrix}\max\\ \min\end{smallmatrix} (\bar{Z}_1,\bar{Z}_2) = \mu \pm \tfrac{1}{2}E|\bar{Z}_1-\bar{Z}_2|.$$

However (since $|a|-|b| \leq |a-b|$ for $a,b \in R$)

$$|\bar{Z}_1-\bar{Z}_2| \leq |\bar{Z}_1-\mu| + |\bar{Z}_2-\mu|,$$

and thus (as $n\to\infty$) by the hypotheses of the theorem $E|\bar{Z}_1-\bar{Z}_2|\to 0$. The result for $k > 2$ follows by induction.

Although it can be proven (see, e.g., Loève (1963), p. 157, d.) that $E|\bar{Z}_1-\mu|\to 0$ implies that $E|\bar{Z}_1|\to|\mu|$, it is not clear when the converse is true. In our situation, we would like to know when $E\bar{Z}_1=\mu$ implies $E|\bar{Z}_1-\mu|\to 0$ (i.e., for which $f(\cdot)$'s this is the case).

(2.1.29) <u>THEOREM</u>: If $\mathrm{var}(\bar{Z}_1)\to 0$ (as $n\to\infty$) then $E|\bar{Z}_1-\mu|\to 0$.

<u>Proof</u>: This follows directly from the fact that $(E|X|^r)^{1/r}$ is a non-decreasing function of $r > 0$ for any r.v. X (see, e.g., Loève (1963), p. 156, c.).

(2.1.30) <u>LEMMA</u>: $\mathrm{Var}(\bar{Z}_1)\to 0$ iff $\int_{-\infty}^{\infty} x^2 f(x)dx<\infty$.

<u>Proof</u>:

$$\mathrm{Var}(\bar{Z}_1) = \frac{1}{n}\ \mathrm{var}(Z_{11}) = \frac{1}{n}\{\int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} x f(x)dx)^2\}.$$

These results on the satisfaction of (2.1.20) of Corollary (2.1.18) may be summarized as follows.

SUMMARY: For any i ($1 \leq i \leq k$), $\overline{X}_{[i]}$ is asymptotically unbiased (as $n \to \infty$) as an estimator of $\mu_{[i]}$ if

(2.1.31)

(1) $|y|$ is uniformly integrable in $[G_n(y|f)]^i$ and $1-[1-G_n(y|f)]^{k-i+1}$,

or if

(2) $\int_{-\infty}^{\infty} x^2 f(x)dx < \infty$.

(Note that (1) holds if, as is often the case, $f(\cdot)$ is concentrated on a bounded set in R.)

For reasons noted above Lemma (2.1.10) it was reasonable to study the expectation and bias of $\overline{X}_{[i]}$ as an estimator of $\mu_{[i]}$ ($i = 1,\ldots,k$) in our context.. With Corollary (2.1.18) as motivation, we note that estimators

(2.1.32) $\overline{X}_{[i]} + a \qquad (h'_{k-i+1}(g_n) - E_f \leq a \leq h_i(g_n) - E_f)$

(correction of $\overline{X}_{[i]}$ by adding a constant) may be preferable to $\overline{X}_{[i]}$ in certain contexts. If positive (negative) bias is very undesirable, one may use $a = h_i(g_n) - E_f$ ($a = h'_{k-i+1}(g_n) - E_f$) and obviate its possibility. If one's preferences on bias are more complicated, one might even remove the restriction $h'_{k-i+1}(g_n) - E_f \leq a \leq h_i(g_n) - E_f$. (Note that this restriction "makes sense" since (see (2.1.14) for notation)

$$h'_{a_1}(g_n) = E\min(Y_1,\ldots,Y_{a_1}) \leq EY_1 \leq E\max(Y_1,\ldots,Y_{a_2}) = h_{a_2}(g_n).)$$

Note that, for certain $f(\cdot)$'s, information about the distribution $G_n(\cdot|f)$ will be available for use in determining $h_i(g_n)$ and $h'_{k-i+1}(g_n)$

$(1 \leq i \leq k)$. For information and references see Reitsma (1963).

THEOREM: Fix i $(1 \leq i \leq k)$. Suppose that the sup and inf of (2.1.19) achieve the bounds of (2.1.19). Then we minimize

$$(1) \quad \left| \begin{matrix}\max \\ \min\end{matrix}_{\mu \in \Omega_o(\mu_{[i]})} (E_\mu \overline{X}_{[i]} - a - \mu_{[i]}) \right|$$

(2.1.33) by choosing $a = \begin{cases} h_i(g_n) - E_f \\ h'_{k-i+1}(g_n) - E_f \end{cases}$, and we minimize

$$(2) \quad \max_{\mu \in \Omega_o(\mu_{[i]})} \left| E_\mu \overline{X}_{[i]} - a - \mu_{[i]} \right|$$

by choosing $a = [h_i(g_n) + h'_{k-i+1}(g_n)]/2 - E_f$.

Proof:

$$\min_{a \varepsilon(-\infty,\infty)} \left| \begin{matrix}\max \\ \min\end{matrix}_{\mu \in \Omega_o(\mu_{[i]})} (E_\mu \overline{X}_{[i]} - a - \mu_{[i]}) \right|$$

$$= \min_{a \varepsilon (-\infty,\infty)} \left| \begin{matrix} (h_i(g_n) - E_f) - a \\ (h'_{k-i+1}(g_n) - E_f) - a \end{matrix} \right| = 0 \text{ at } a = \begin{cases} h_i(g_n) - E_f \\ h'_{k-i+1}(g_n) - E_f. \end{cases}$$

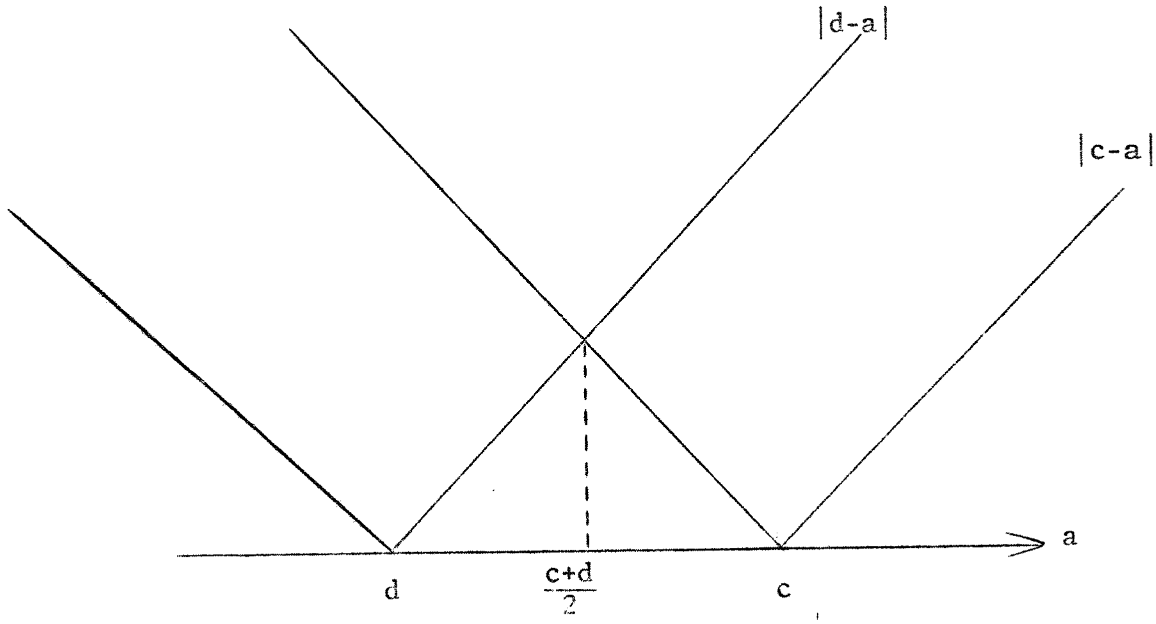
For (2),

$$\min_{a \varepsilon (-\infty,\infty)} \max_{\mu \in \Omega_o(\mu_{[i]})} \left| E_\mu \overline{X}_{[i]} - a - \mu_{[i]} \right|$$

$$= \min_{a\varepsilon(-\infty,\infty)} \max \left(|h_i(g_n) - E_f - a|, |h'_{k-i+1}(g_n) - E_f - a| \right)$$

$$= \frac{h_i(g_n) - h'_{k-i+1}(g_n)}{2} \text{ at } a = \frac{h_i(g_n) + h'_{k-i+1}(g_n)}{2} - E_f,$$

since (for $c \geq d$) $\min_a \max (|c-a|,|d-a|) = (c-d)/2$, as illustrated below.



It is of practical interest to know how any statistical procedure performs when the (distributional and other) assumptions under which it was derived are not met. We then say that (for deviations of a specified sort) the procedure is "robust" or "not robust," according to whether the goal(s) of the procedure are or are not met "well" under the deviations.

The question of how our procedure for estimating $\mu_{[i]}$ $(1 \leq i \leq k)$ performs when specific distributional assumptions are used to set n, but do not hold, is answered in part by our treatment of the estimation problem for a location parameter family in this section. (The question of robustness of Rule (1.3.2) is not our concern here; for some results on this see Dudewicz (1968).)

The robustness interpretation of these results is large-sample. Small-sample robustness can be studied numerically for the $f(\cdot)$'s important in any particular problem, utilizing n. If one is considering a location parameter family other than the normal, results related to robustness can be used to help design "good" procedures, and to help compute the loss that would result from using sample means instead of the appropriate sufficient statistic. If this loss (measured perhaps in increments in n) were small enough, one might wish to use sample means since they might be more robust. (In any particular case this could be checked numerically.)

Examples of location parameter families where Assumption (2.1.2) holds but $\overline{X}_{[i]}$ is not an asymptotically unbiased estimator of $\mu_{[i]}$ $(1 \leq i \leq k)$ are presumed to exist. The case of Cauchy populations (excluded by (2.1.2)) may yield some insight. Here, $G_n(y|f_c)$ is independent of n (by a property of means of independent observations from f_c). (If Cauchy populations were being dealt with, Rule (1.3.2) would not be used. See Dudewicz (1966), pp. 39-45.)

The relationship between the uniform integrability condition of Theorem (2.1.26) and the condition of Theorem (2.1.28) (each of which is sufficient) is of interest. We first clarify the role of i $(1 \leq i \leq k)$ in Theorem (2.1.26).

(2.1.34) THEOREM: Fix i $(1 \leq i \leq k)$. If $|y|$ is uniformly integrable in $G_n(y|f)$, then it is uniformly integrable in $[G_n(y|f)]^i$ and $1-[1-G_n(y|f)]^{k-i+1}$.

Proof: For $-\infty \leq a < b \leq +\infty$,

$$\int_a^b |y| d_y\{[G_n(y|f)]^i\} = i\int_a^b |y|[G_n(y|f)]^{i-1} d_y G_n(y|f) \leq i\int_a^b |y| d_y G_n(y|f),$$

and (for $j \geq 1$)

$$d_y\{1-[1-G_n(y|f)]^j\} = +j[1-G_n(y|f)]^{j-1} d_y G_n(y|f) \leq j d_y G_n(y|f).$$

(2.1.35) THEOREM: $E|\overline{Z}_1-\mu|\to 0$ iff $|\overline{Z}_1|$ is uniformly integrable (i.e., $|y|$ is uniformly integrable in $G_n(y|f)$).

Proof: Since $E|\overline{Z}_1|<\infty$ (because $E\overline{Z}_1 = \mu$ exists) and since $\overline{Z}_1$ converges stochastically to μ, the result follows from the L_r-convergence theorem (see, e.g., Loève (1963), p. 163, c.).

2.2. THE NORMAL CASE

In this section we consider set-up (1.3.1), for which Rule (1.3.2) was originally suggested. The form of the location parameter family results of Section 2.1 is shown, and further results are provided for normal populations.

Denote $(1/\sigma)\phi(y/\sigma)$ by $\phi_\sigma(y)$. Then the quantities defined in Section 2.1 for a location parameter family are (for $i = 1,\ldots,k$) as follows in the case of normality.

$$f(x-\mu_i) = (1/\sigma)\phi((x-\mu_i)/\sigma) = \phi_\sigma(x-\mu_i);$$

$$E_{\phi_\sigma} = \int_{-\infty}^{\infty} y\phi_\sigma(y)dy = 0;$$

$$G_n(y|\phi_\sigma) = P[\overline{X}_i-\mu_i \le y] = P\left[\frac{\overline{X}_i-\mu_i}{\sigma/\sqrt{n}} \le \frac{y}{\sigma/\sqrt{n}}\right] = \Phi\left(\frac{y}{\sigma/\sqrt{n}}\right);$$

$$g_n(y|\phi_\sigma) = \frac{1}{\sigma/\sqrt{n}}\phi\left(\frac{y}{\sigma/\sqrt{n}}\right) = \phi_{\sigma/\sqrt{n}}(y); \tag{2.2.1}$$

$$h_\ell(g_n) = E[\text{max of } \ell \text{ r.v.'s with fr.f. } g_n(y|\phi_\sigma)]$$

$$= E[\text{max of } \ell\ N(0,\sigma^2/n) \text{ r.v.'s}]$$

$$= (\sigma/\sqrt{n})E[\text{max of } \ell\ N(0,1) \text{ r.v.'s}] = (\sigma/\sqrt{n})h_\ell(\phi);$$

$$h'_\ell(g_n) = -h_\ell(g_n) = -(\sigma/\sqrt{n})h_\ell(\phi) \text{ by Lemma (2.1.13).}$$

Note that in the normal case, since $h_\ell(g_n) = -h'_\ell(g_n) = (\sigma/\sqrt{n})h_\ell(\phi)$ $(\ell = 1,2,\ldots)$, only $h_\ell(\phi)$ need be tabulated. ($h_\ell(\phi) > 0$ 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 $h_\ell(\phi)$ have been computed by (e.g.) Teichroew (1956) where $h_\ell(\phi) = E(x_1;\ell)$, and by Harter (1961) where $h_\ell(\phi) = E(x_{1|\ell})$.

Tables of $h_\ell(\phi)$ have been computed by Tippett (1925). We now present some values of $h_\ell(\phi)$ obtained from Harter (1961) for ℓ = 2(1)10(5) 25(25)50(50)400, and from Tippett (1925) for ℓ = 500,1000. (For further references, see Kendall and Stuart (1963), pp. 329, 336.)

Table (2.2.2). Values of $h_\ell(\phi)$

ℓ	$h_\ell(\phi)$	ℓ	$h_\ell(\phi)$
2	.56419	50	2.24907
3	.84628	100	2.50759
4	1.02938	150	2.64925
5	1.16296	200	2.74604
6	1.26721	250	2.81918
7	1.35218	300	2.87777
8	1.42360	350	2.92651
9	1.48501	400	2.96818
10	1.53875	500	3.03670
15	1.73591	1000	3.24144
20	1.86748		
25	1.96531		

From Corollary (2.1.18), (2.1.31)(2), and (2.2.1), the following theorem emerges for the normal case.

THEOREM: For any i ($1 \leq i \leq k$),

(2.2.3) $\mu_{[i]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi) \leq E_\mu \overline{X}_{[i]} \leq \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi)$

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

The following theorem shows that the bounds of Theorem (2.2.3) are actually the sup and inf. (For the location parameter case, the inf for i = 1 and the sup for i = k were proven as Theorem (2.1.21).)

THEOREM: For any i ($1 \leq i \leq k$),

$$\inf\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} = \mu_{[i]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi)$$

(2.2.4) and

$$\sup\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} = \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi).$$

Proof: By Theorem (2.1.15), the infimum is $\geq \mu_{[i]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi)$ and the supremum is $\leq \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi)$. We will now show that

$$\inf\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} \leq \mu_{[i]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi)$$

$$\sup\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} \geq \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi).$$

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

$$\inf\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} \leq \inf\{E_\mu \overline{X}_{[i]}:$$

$$\mu = (\overbrace{\mu_{[1]},\ldots,\mu_{[1]}}^{i-1 \text{ terms}}, \overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{k-i+1 \text{ terms}}) \in \Omega_o(\mu_{[i]})\}$$

$$\sup\{E_\mu \overline{X}_{[i]}: \mu \in \Omega_o(\mu_{[i]})\} \geq \sup\{E_\mu \overline{X}_{[i]}:$$

$$\mu = (\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i \text{ terms}}, \overbrace{\mu_{[k]},\ldots,\mu_{[k]}}^{k-i \text{ terms}}) \in \Omega_o(\mu_{[i]})\}.$$

Case 1. The infimum. By Lemma (2.1.10) and Theorem (2.1.11), $E_\mu \overline{X}_{[i]}$, with $\mu = (-M,\ldots,-M,\mu_{[i]},\ldots,\mu_{[i]})$, decreases as $M\uparrow$. If we let $H_M(x)$ denote $F_{\overline{X}_{[i]}}(x)$ with $\mu = (-M,\ldots,-M,\mu_{[i]},\ldots,\mu_{[i]})$, the desired

$$\inf\{E_\mu \overline{X}_{[i]}: \mu = (-M,\ldots,-M,\mu_{[i]},\ldots,\mu_{[i]}) \in \Omega_o(\mu_{[i]})\} = \lim_{M\to\infty} \int_{-\infty}^{\infty} x\,dH_M(x).$$

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

$$H_M(x)\to H_\infty(x) \equiv F_{\overline{X}_{[i]}}(x) \text{ with } \mu=(\overbrace{-\infty,\ldots,-\infty}^{i-1 \text{ terms}},\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{k-i+1 \text{ terms}}).$$

Thus, by Theorem (2.1.25), if $|x|$ is uniformly integrable in H_M, then

$$\lim_{M\to\infty} \int_{-\infty}^{\infty} x\,dH_M(x) = \int_{-\infty}^{\infty} x\,dH_\infty(x) = \mu_{[i]} - (\sigma/\sqrt{n})h_{k-i+1}(\phi),$$

where the last equality uses Lemma (2.1.14) and (2.2.1). Since $|x|$ is uniformly integrable in H_M by Lemma (2.2.6), this part of the theorem is proven.

Case 2. The supremum. By Lemma (2.1.10) and Theorem (2.1.11), $E_\mu \overline{X}_{[i]}$, with $\mu=(\mu_{[i]},\ldots,\mu_{[i]},M,\ldots,M)$, increases as $M\uparrow$. If we let $J_M(x)$ denote $F_{\overline{X}_{[i]}}(x)$ with $\mu=(\mu_{[i]},\ldots,\mu_{[i]},M,\ldots,M)$, the desired

$$\sup\{E_\mu \overline{X}_{[i]} : \mu=(\mu_{[i]},\ldots,\mu_{[i]},M,\ldots,M) \in \Omega_o(\mu_{[i]})\} = \lim_{M\to\infty} \int_{-\infty}^{\infty} x\,dJ_M(x).$$

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

$$J_M(x)\to J_\infty(x) \equiv F_{\overline{X}_{[i]}}(x) \text{ with } \mu=(\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i \text{ terms}},\overbrace{+\infty,\ldots,+\infty}^{k-i \text{ terms}}).$$

The theorem follows as in Case 1, now using the fact that $|x|$ is uniformly integrable in J_M by Lemma (2.2.7).

LEMMA: For any $\mu \in \Omega_o(\mu_{[i]})$,

$$dF_{\overline{X}_{[i]}}(x) \le \sum_{\ell=i}^{k} \frac{\binom{k}{\ell}}{k!} \sum_{\beta\in S_k} \left\{ \sum_{j=1}^{\ell} f_{\overline{X}_{\beta(j)}}(x) \frac{F_{\overline{X}_{\beta(1)}}(x)\cdots F_{\overline{X}_{\beta(\ell)}}(x)}{F_{\overline{X}_{\beta(j)}}(x)} \right\} dx. \tag{2.2.5}$$

Proof:

$F_{\overline{X}_{[i]}}(x) = P[\text{At least } i \text{ of } \overline{X}_1,\ldots,\overline{X}_k \text{ are } \le x]$

$$= \sum_{\ell=i}^{k} P[\text{Exactly } \ell \text{ of } \overline{X}_1,\ldots,\overline{X}_k \text{ are } \le x]$$

$$= \sum_{\ell=i}^{k} \frac{1}{\ell!(k-\ell)!} \sum_{\beta\in S_k} P[\overline{X}_{\beta(1)}\le x,\ldots,\overline{X}_{\beta(\ell)}\le x \;\&\; \overline{X}_{\beta(\ell+1)}>x,\ldots,\overline{X}_{\beta(k)}>x]$$

$$= \sum_{\ell=i}^{k} \frac{\binom{k}{\ell}}{k!} \sum_{\beta \varepsilon S_k} F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x)[1-F_{\overline{X}_{\beta(\ell+1)}}(x)] \ldots [1-F_{\overline{X}_{\beta(k)}}(x)].$$

Thus,

$$dF_{\overline{X}_{[i]}}(x) = \sum_{\ell=i}^{k} \frac{\binom{k}{\ell}}{k!} \sum_{\beta \varepsilon S_k} \left\{ \sum_{j=1}^{\ell} f_{\overline{X}_{\beta(j)}}(x) \frac{F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x)}{F_{\overline{X}_{\beta(j)}}(x)} \circ \right.$$

$$\circ [1-F_{\overline{X}_{\beta(\ell+1)}}(x)] \ldots [1-F_{\overline{X}_{\beta(k)}}(x)] - \sum_{j=\ell+1}^{k} f_{\overline{X}_{\beta(j)}}(x) \circ$$

$$\left. \circ F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x) \frac{[1-F_{\overline{X}_{\beta(\ell+1)}}(x)] \ldots [1-F_{\overline{X}_{\beta(k)}}(x)]}{[1-F_{\overline{X}_{\beta(j)}}(x)]} \right\} dx$$

$$\leq \sum_{\ell=i}^{k} \frac{\binom{k}{\ell}}{k!} \sum_{\beta \varepsilon S_k} \left\{ \sum_{j=1}^{\ell} f_{\overline{X}_{\beta(j)}}(x) \frac{F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x)}{F_{\overline{X}_{\beta(j)}}(x)} \right\} dx.$$

LEMMA: $|x|$ is uniformly integrable in $H_M(x) = F_{\overline{X}_{[i]}}(x)$ with

(2.2.6) $$\mu = (\overbrace{-M, \ldots, -M}^{i-1 \text{ terms}}, \overbrace{\mu_{[i]}, \ldots, \mu_{[i]}}^{k-i+1 \text{ terms}}).$$

Proof: Let L be positive. Then, by Lemma (2.2.5),

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

$$\leq \sum_{\ell=i}^{k} \frac{\binom{k}{\ell}}{k!} \sum_{\beta \varepsilon S_k} \left\{ \sum_{j=1}^{\ell} \int_{|x| \geq L} |x| f_{\overline{X}_{\beta(j)}}(x) \frac{F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x)}{F_{\overline{X}_{\beta(j)}}(x)} dx \right\}.$$

Fix any $\varepsilon > 0$. We will now show that there is an $L = L(\varepsilon)$ such that the upper bound on $\int_{|x| \geq L} |x| dH_M(x)$ is $\leq \varepsilon$ regardless of the value of M. By

Definition (2.1.24), this will prove $|x|$ is uniformly integrable in $H_M(x)$.

Since $\ell = i, i+1, \ldots, k$, and since $i-1$ populations have means $-M$ while $k-i+1$ have means $\mu_{[i]}$, for any fixed ℓ and β at least one of $\overline{X}_{\beta(1)}, \ldots, \overline{X}_{\beta(\ell)}$ 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

$$T(\ell,\beta,j) = \frac{\binom{k}{\ell}}{k!} \int_{|x|\geq L} |x| f_{\overline{X}_{\beta(j)}}(x) \frac{F_{\overline{X}_{\beta(1)}}(x) \ldots F_{\overline{X}_{\beta(\ell)}}(x)}{F_{\overline{X}_{\beta(j)}}(x)} dx.$$

Case 1. $\overline{X}_{\beta(j)}$ comes from a population with mean $\mu_{[i]}$. Then

$$T(\ell,\beta,j) \leq \frac{\binom{k}{\ell}}{k!} \int_{|x|\geq L} |x| f_{\overline{X}_{\beta(j)}}(x) dx$$

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

Case 2. $\overline{X}_{\beta(j)}$ comes from a population with mean $-M$. Then one of $\overline{X}_{\beta(1)}, \ldots, \overline{X}_{\beta(\ell)}$ (but not $\overline{X}_{\beta(j)}$) comes from a population with mean $\mu_{[i]}$; call it $\overline{X}_{\beta_o}$. Then

$$T(\ell,\beta,j) \leq \frac{\binom{k}{\ell}}{k!} \int_{x\geq L} |x| f_{\overline{X}_{\beta(j)}}(x) dx + \frac{\binom{k}{\ell}}{k!} \int_{x\leq -L} |x| f_{\overline{X}_{\beta(j)}}(x) F_{\overline{X}_{\beta_o}}(x) dx.$$

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

Now, since $\overline{X}_{\beta_o}$ is $N(\mu_{[i]}, \sigma^2/n)$, for $x < -|\mu_{[i]}|$ (so that

$-x + \mu_{[i]} > 0)$

$$F_{\overline{X}_{\beta_o}}(x) = P[\overline{X}_{\beta_o} \leq x] = P\left[\frac{\overline{X}_{\beta_o} - \mu_{[i]}}{\sigma/\sqrt{n}} \leq \frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right] = \Phi\left(\frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right)$$

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

by the result (see, e.g., Feller (1957), p. 179) 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 $L \geq 2|\mu_{[i]}|$,

$$\int_{x \leq -L} |x| f_{\overline{X}_{\beta(j)}}(x) F_{\overline{X}_{\beta_o}}(x)dx$$

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

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

$$\leq \frac{\sigma/\sqrt{n}}{\sqrt{2\pi}} \int_{x \leq -L} 2 \frac{1}{\sqrt{2\pi}\ \sigma/\sqrt{n}} e^{-\frac{1}{2}\left(\frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right)^2} dx = \sigma\sqrt{2/(n\pi)}\ P[\overline{X}_{\beta_o} \leq -L].$$

Since $\overline{X}_{\beta_o}$ is $N(\mu_{[i]}, \sigma^2/n)$, it is clear that for $L \geq L_3(\ell,\beta,j,\mu_{[i]},\varepsilon)$

the second term of $T(\ell,\beta,j)$ is $< \frac{1}{2}\frac{\varepsilon}{(k-i+1)k!k}$, so that for

$L \geq L_4(\ell,\beta,j,\mu_{[i]},\varepsilon) = \max(L_2,L_3)$ we have $T(\ell,\beta,j) < \dfrac{\varepsilon}{(k-i+1)k!k}$

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

<u>LEMMA</u>: $|x|$ is uniformly integrable in $J_M(x) = F_{\overline{X}_{[i]}}(x)$ with

(2.2.7)
$$\mu = (\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i\text{ terms}},\overbrace{M,\ldots,M}^{k-i\text{ terms}}).$$

<u>Proof</u>: 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 (2.1.24), 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 $\int_{|x|\geq L} |x| dJ_M(x) \leq \varepsilon$ for all M.

For $M > |\mu_{[i]}|$, by Theorem (2.1.11),

$$\begin{aligned} J_M(x) &= F_{\overline{X}_{[i]}}(x) \text{ with } \mu=(\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i\text{ times}},\overbrace{M,\ldots,M}^{k-i\text{ times}}) \\ &\leq F_{\overline{X}_{[i]}}(x) \text{ with } \mu = (\overbrace{-M,\ldots,-M}^{i-1\text{ times}},\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{k-i+1\text{ times}}) \\ &= H_M(x). \end{aligned}$$

Define two d.f.'s

$$F(x) = \begin{cases} 1 & \text{if } x \geq -L \\ J_M(x) & \text{if } x < -L \end{cases}, \qquad G(x) = \begin{cases} 1 & \text{if } x \geq -L \\ H_M(x) & \text{if } x < -L. \end{cases}$$

Then by Lemma (2.1.10),

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

$$\int_{-\infty}^{-L} x dJ_M(x) - L(1-J_M(-L)) \geq \int_{-\infty}^{-L} x dH_M(x) - L(1-H_M(-L))$$

$$0 \geq \int_{-\infty}^{-L} x dJ_M(x) \geq \int_{-\infty}^{-L} x dH_M(x) + L\{H_M(-L)-J_M(-L)\}.$$

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 (2.2.6), we find that

$$0 \geq \int_{-\infty}^{-L} x dJ_M(x) \geq \int_{-\infty}^{-L} x dH_M(x) \to 0 \text{ uniformly in M.}$$

Thus, there is (for any fixed $\mu_{[i]}$) an $L_1(\varepsilon)$ such that for $L > L_1(\varepsilon)$ we have $\int_{-\infty}^{-L} |x| dJ_M(x) < \varepsilon/2$ uniformly in M.

Take $L > L_1(\varepsilon)$. By Theorem (2.2.3) and Theorem (2.2.4), we have

$$\begin{aligned}
\mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi) &\geq \sup\{E_\mu \overline{X}_{[i]}: \mu=(\mu_{[i]},\ldots,\mu_{[i]},M,\ldots,M) \in \Omega_o(\mu_{[i]})\} \\
&= \lim_{M\to\infty} \int_{-\infty}^{\infty} x dJ_M(x) = \lim_{M\to\infty} \{\int_{-\infty}^{-L} x dJ_M(x) + \int_{-L}^{L} x dJ_M(x) + \int_{L}^{\infty} x dJ_M(x)\} \\
&\geq -\varepsilon/2 + \lim_{M\to\infty} \int_{-L}^{L} x dJ_M(x) + \lim_{M\to\infty} \int_{L}^{\infty} x dJ_M(x) \\
&= -\varepsilon/2 + \int_{-L}^{L} x dJ_\infty(x) + \lim_{M\to\infty} \int_{L}^{\infty} x dJ_M(x).
\end{aligned}$$

The last step follows from the Helly-Bray Lemma (as in (2.1.27)). Since (as shown in Theorem (2.2.4))

$$\int_{-\infty}^{\infty} x dJ_\infty(x) = \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi),$$

for $L > L_2(\varepsilon)$ we have $\int_{-L}^{L} x dJ_\infty(x)$ within $\varepsilon/2$ of $\mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi)$.

Thus, if $L > \max(L_1,L_2)$ then

$$\mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi) \geq -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} + \mu_{[i]} + (\sigma/\sqrt{n})h_i(\phi) + \lim_{M\to\infty} \int_{L}^{\infty} x dJ_M(x)$$

$$\varepsilon \geq \lim_{M\to\infty} \int_L^\infty x dJ_M(x).$$

Thus, there is an $L = L_3(\varepsilon)$ such that $\int_{|x|\geq L} |x| dJ_M(x) \leq \varepsilon$ regardless of the value of M.

Among the results of Section 2.1 for a location parameter family which ergo hold for the normal family of the present section, the linear corrections for (e.g.) minimax|bias| at equation (2.1.32)ff are worthy of special note. We may then (in the normal case) readily 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 (1.3.2), say n_1. (2) Set n to make certain minimax|bias|'s suitably "small," say n_2. (3) Set $n=\max(n_1,n_2)$.

Table (2.2.2) of values of $h_\ell(\phi)$ indicates that for k in the range in which Rule (1.3.2) 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 ℓ 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 |bias| in $\bar{X}_{[k]} - a$, as an estimator of $\mu_{[k]}$, $\leq \varepsilon$ ($\varepsilon > 0$), he would find approximately that if n_o sufficed for k = 2, $4n_o$ would suffice for k = 5; and that if n_o sufficed for k = 9, $4n_o$ would suffice for k = 500, since by Theorem (2.1.33) the minimax|bias| is

$$\frac{h_i(g_n)-h'_{k-i+1}(g_n)}{2} = \frac{h_k(g_n)-h'_1(g_n)}{2} = \frac{1}{2}(\sigma/\sqrt{n})h_k(\phi).$$

Note that if there are restrictions on the μ_i (i = 1,...,k) in a practical case, then the inf and sup of Theorem (2.2.4) can be improved.

For example, if $A \leq \mu_i \leq B$ $(i = 1,\ldots,k)$, 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 $\mu \in \Omega_o(\mu_{[i]})$, say $\mu \in \Omega_\delta(\mu_{[i]})$, that sup and inf would also be attained by raising (lowering) the components of μ 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 $\mu \in \Omega_\delta(\mu_{[i]})$. However, since our aim is good estimation of $\mu_{[i]}$ <u>regardless</u> of μ, the set used above ($\mu \in \Omega_o(\mu_{[i]})$) will usually be the proper one. (For special uses of the estimate of $\mu_{[i]}$ one may only "care" when, for some δ, $\mu \in \Omega_\delta(\mu_{[i]})$.)

CHAPTER 3. POINT ESTIMATION: STRONG CONSISTENCY

3.1. STRONG (W.P. 1) CONSISTENCY OF A NATURAL ESTIMATOR OF $\mu_{[i]}$ ($1 \le i \le k$) FOR A LOCATION PARAMETER FAMILY

Consider $\bar{X}_{[i]}$ as an estimator of $\mu_{[i]}$ ($1 \le i \le k$) when Set-up (2.1.1) and Assumption (2.1.2) hold, i.e., when observations from population π_i have fr.f. $f(x-\theta_i)$, $X \varepsilon R$, $i = 1,\ldots,k$, and the mean of f exists. If Z is a constant (say θ) with probability one (w.p. 1), a sequence of estimators $\{Z_n;\ n \ge 1\}$ is said to be: <u>strongly consistent</u> (for θ) if Z_n converges to θ w.p. 1; <u>consistent</u> (for θ) if Z_n converges to θ in probability. Since convergence w.p. 1 implies convergence in probability, strong consistency implies consistency.

(3.1.1) <u>LEMMA</u>: Let $T_1(n),\ldots,T_k(n)$ ($n \ge 1$) be r.v.'s which converge w.p. 1 to r.v.'s $T_1,\ldots,T_k$ (respectively). Suppose that $g(t_1,\ldots,t_k)$ is a continuous function of k real variables. Then

$$g(T_1(n),\ldots,T_k(n))$$

converges w.p. 1 to $g(T_1,\ldots,T_k)$.

<u>Proof</u>: Suppose that all r.v.'s involved are defined on a probability space $(\Omega,\mathcal{B},P)$. Then by a characterization of convergence w.p. 1 (see, e.g., Parzen (1960), p. 415), it suffices to prove that for every $\varepsilon > 0$, $\delta > 0$ there exists an integer $N_0 > 0$ such that

$$P[\sup_{n \ge N_0} |g(T_1(n),\ldots,T_k(n)) - g(T_1,\ldots,T_k)| > \varepsilon] < \delta.$$

However, by the continuity of $g(\cdot,\ldots,\cdot)$ and the convergence of $T_i(n)$ to T_i w.p. 1 ($1 \le i \le k$), this is clear.

(3.1.2) THEOREM: $\overline{X}_{[i]}$ is strongly consistent as an estimator of $\mu_{[i]}$ $(1 \leq i \leq k)$.

Proof: Since $\int_{-\infty}^{\infty} xf(x)dx$ is assumed to be a finite number, it follows by Kolmogorov's Strong Law of Large Numbers (see, e.g., Loève (1963), p. 239) that $\overline{X}_1,\ldots,\overline{X}_k$ converge w.p. 1 to $\mu_1,\ldots,\mu_k$ (respectively). Thus by Lemma (3.1.1) $\overline{X}_{[i]}$ converges w.p. 1 to $\mu_{[i]}$ $(i = 1,\ldots,k)$.

The stronger theorem, that $g(\overline{X}_{[i]})$ converges w.p. 1 to $g(\mu_{[i]})$ for any continuous real-valued function $g(\cdot)$ $(1 \leq i \leq k)$ is obvious. It can be used as follows: $g(\overline{X}_{[k]})$ 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 μ, 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 $\mu_{[i]}$ to be a r.v. $(1 \leq i \leq k)$.

Note that strong consistency of $\overline{X}_{[i]}$ as an estimator of $\mu_{[i]}$ implies strong consistency of $\overline{X}_{[i]} \pm a_n$ where $\lim_{n\to\infty} a_n = 0$ $(i = 1,\ldots,k)$.

(This, of course, was also the case for asymptotic unbiasedness.)

CHAPTER 4. POINT ESTIMATION: SQUARED ERROR

4.1. SQUARED ERROR OF A NATURAL ESTIMATOR OF $\mu_{[i]}$ $(1 \le i \le k)$ FOR A LOCATION PARAMETER FAMILY

In this section we consider the squared error of $\bar{X}_{[i]}$ as an estimator of $\mu_{[i]}$ $(1 \le i \le k)$ when Set-up (2.1.1) and Assumption (2.1.2) hold, i.e., when observations from population π_i have fr.f. $f(x-\theta_i)$, $x \in R$, $i = 1,\ldots,k$, and the mean of f exists. The expectation of this quantity, i.e.

(4.1.1) $$E_\mu(\bar{X}_{[i]}-\mu_{[i]})^2,$$

will be of special interest.

<u>LEMMA</u>: If $F(\cdot)$ and $G(\cdot)$ are d.f.'s with $F(x) \le G(x)$ $(x \in R)$, then for $\psi(x)$ any monotone non-decreasing function of x we have

(4.1.2) $$\int_{-\infty}^{\infty}\psi(x)dG(x) \le \int_{-\infty}^{\infty}\psi(x)dF(x),$$

with the inequality reversed if $\psi(x)$ is monotone non-increasing.

This lemma, which is a generalization of Lemma (2.1.10), has been essentially stated by Alam (1967), p. 283, who refers to Lehmann (1955) for the proof. That reference is concerned with more general questions (which makes it difficult to extract the needed proof). A simple proof (for the strictly monotone $\psi(\cdot)$ case) is possible using the inverse function. We omit this since Mahamunulu (1967), p. 1082, has recently published a reference on this result.

DEFINITION: For our location parameter family, let

(4.1.3) $$H_\infty(x) = F_{\overline{X}_{[i]}}(x) \text{ with } \mu = (\overbrace{-\infty,\ldots,-\infty}^{i-1 \text{ terms}},\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{k-i+1 \text{ terms}})$$

$$J_\infty(x) = F_{\overline{X}_{[i]}}(x) \text{ with } \mu = (\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i \text{ terms}},\overbrace{+\infty,\ldots,+\infty}^{k-i \text{ terms}}).$$

Although $H_\infty(\cdot)$ and $J_\infty(\cdot)$ depend on i $(1 \le i \le k)$, this dependence will be suppressed. (We used this notation for the normal case in Theorem (2.2.4).)

LEMMA: For any monotone non-decreasing function of x $\psi(x)$ and $\mu \in \Omega_o(\mu_{[i]})$,

(4.1.4) $$\int_{-\infty}^{\infty}\psi(x)dH_\infty(x) \le \int_{-\infty}^{\infty}\psi(x)dF_{\overline{X}_{[i]}}(x) \le \int_{-\infty}^{\infty}\psi(x)dJ_\infty(x)$$

$(i = 1,\ldots,k)$, with both inequalities reversed if $\psi(x)$ is monotone non-increasing.

Proof: This follows from Theorem (2.1.11) and Lemma (4.1.2).

THEOREM: For any i $(1 \le i \le k)$ and any $\mu \in \Omega_o(\mu_{[i]})$,

(4.1.5) $$\int_{\mu_{[i]}}^{\infty}(x-\mu_{[i]})^2dH_\infty(x) + \int_{-\infty}^{\mu_{[i]}}(x-\mu_{[i]})^2dJ_\infty(x) \le E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2$$

$$\le \int_{-\infty}^{\mu_{[i]}}(x-\mu_{[i]})^2dH_\infty(x) + \int_{\mu_{[i]}}^{\infty}(x-\mu_{[i]})^2dJ_\infty(x).$$

Proof: Define

$$\psi_1(x) = \begin{cases} (x-\mu_{[i]})^2 & \text{if } x-\mu_{[i]} > 0 \\ 0 & \text{if } x-\mu_{[i]} \le 0 \end{cases}$$

$$\psi_2(x) = \begin{cases} 0 & \text{if } x-\mu_{[i]} > 0 \\ (x-\mu_{[i]})^2 & \text{if } x-\mu_{[i]} \le 0. \end{cases}$$

Then by Lemma (4.1.4), since $\psi_1(x)$ is monotone non-decreasing in x and $\psi_2(x)$ is monotone non-increasing in x,

$$0 \le \int_{\mu_{[i]}}^{\infty} (x-\mu_{[i]})^2 dH_\infty(x) \le \int_{\mu_{[i]}}^{\infty} (x-\mu_{[i]})^2 dF_{\overline{X}_{[i]}}(x) \le \int_{\mu_{[i]}}^{\infty} (x-\mu_{[i]})^2 dJ_\infty(x),$$

$$\int_{-\infty}^{\mu_{[i]}} (x-\mu_{[i]})^2 dH_\infty(x) \ge \int_{-\infty}^{\mu_{[i]}} (x-\mu_{[i]})^2 dF_{\overline{X}_{[i]}}(x) \ge \int_{-\infty}^{\mu_{[i]}} (x-\mu_{[i]})^2 dJ_\infty(x) \ge 0,$$

from which the theorem follows easily.

Note that since (for any r.v. Z) $EZ^2-(EZ)^2 = \mathrm{Var}(Z)$ and since Corollary (2.1.18) gives us bounds on $E_\mu \overline{X}_{[i]}-\mu_{[i]}$, Theorem (4.1.5) can be used to obtain bounds on

$$\mathrm{Var}_\mu \overline{X}_{[i]} = \mathrm{Var}_\mu(\overline{X}_{[i]}-\mu_{[i]}) = E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2 - (E_\mu \overline{X}_{[i]}-\mu_{[i]})^2.$$

4.2. THE NORMAL CASE

In this section we first find the form of the results of Section 4.1 in the case of normality. Under normality,

$$H_\infty(x) = P[\text{Minimum of } k\text{-}i\text{+}1\ N(\mu_{[i]},\sigma^2/n) \text{ r.v.'s is } \leq x]$$

$$= P\left[\text{Min of } k\text{-}i\text{+}1\ N(0,1) \text{ r.v.'s is } \leq \frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right]$$

$$= 1-\left[1-\Phi\left(\frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right)\right]^{k-i+1};$$

$$J_\infty(x) = P[\text{Maximum of } i\ N(\mu_{[i]},\sigma^2/n) \text{ r.v.'s is } \leq x]$$

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

Thus,

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

$$= (\sigma^2/n)\int_{0}^{\infty} x^2 d\{-[1-\Phi(x)]^{k-i+1}\} = (\sigma^2/n)\int_{-\infty}^{0} x^2 d\{[\Phi(x)]^{k-i+1}\};$$

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

and

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

$$= (\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{-[1-\Phi(x)]^{k-i+1}\right\} = (\sigma^2/n)\int_{0}^{\infty} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\};$$

$$\int_{\mu_{[i]}}^{\infty}(x-\mu_{[i]})^2 dJ_\infty(x) = \int_{\mu_{[i]}}^{\infty}(x-\mu_{[i]})^2 d\left\{\left[\Phi\left(\frac{x-\mu_{[i]}}{\sigma/\sqrt{n}}\right)\right]^i\right\} = (\sigma^2/n)\int_0^\infty x^2 d\left\{\left[\Phi(x)\right]^i\right\}.$$

Thus, by specializing Theorem (4.1.5) to the case of normality and using the above results, we obtain the following theorem.

THEOREM: For any i $(1 \leq i \leq k)$ and any $\mu \in \Omega_o(\mu_{[i]})$,

(4.2.1)

$$(\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{\left[\Phi(x)\right]^{k-i+1}\right\} + (\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{\left[\Phi(x)\right]^{i}\right\} \leq E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2$$

$$\leq (\sigma^2/n)\int_0^\infty x^2 d\left\{\left[\Phi(x)\right]^{k-i+1}\right\} + (\sigma^2/n)\int_0^\infty x^2 d\left\{\left[\Phi(x)\right]^{i}\right\}.$$

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: For any i $(1 \leq i \leq k)$, taking the inf and sup over $\mu \in \Omega_o(\mu_{[i]})$,

$\inf E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2$

(4.2.2)

$$\leq \min\left((\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{\left[\Phi(x)\right]^{k-i+1}\right\},\ (\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{\left[\Phi(x)\right]^{i}\right\}\right)$$

$\sup E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2$

$$\geq \max\left((\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{\left[\Phi(x)\right]^{k-i+1}\right\},\ (\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{\left[\Phi(x)\right]^{i}\right\}\right).$$

Proof: Since (see Theorem (2.2.4)) $H_M(x)$ and $J_M(x)$ converge weakly to $H_\infty(x)$ and $J_\infty(x)$ (respectively), by Theorem (2.1.25) it follows that, if x^2 is uniformly integrable in H_M and J_M, then

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

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

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 (2.2.6).

The fact that x^2 is uniformly integrable in J_M requires major modification of the proof of Lemma (2.2.7), as will now be noted. Using Lemma (4.1.4) with the non-increasing function

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

(instead of Lemma (2.1.10)) 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

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

Thus, there is (for any fixed $\mu_{[i]}$) an $L_1(\varepsilon)$ such that for $L > L_1(\varepsilon)$ we have $\int_{-\infty}^{-L} x^2 dJ_M(x) < \varepsilon/2$ uniformly in M.

By Theorem (2.1.11), $J_M(x) \geq J_\infty(x)$. 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 by Lemma (4.1.2),

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

$$\int_L^{\infty} x^2 dJ_M(x) + L^2 J_M(L) \leq \int_L^{\infty} x^2 dJ_\infty(x) + L^2 J_\infty(L)$$

$$0 \leq \int_L^{\infty} x^2 dJ_M(x) \leq L^2\{J_\infty(L) - J_M(L)\} + \int_L^{\infty} x^2 dJ_\infty(x) \leq \int_L^{\infty} x^2 dJ_\infty(x).$$

Now since $\int_{-\infty}^{\infty} x^2 dJ_\infty(x)$ exists, for $L > L_2(\varepsilon)$ we have $\int_L^{\infty} x^2 dJ_M(x) \leq \varepsilon/2$ uniformly in M. The result then follows as in Lemma (2.2.7).

We now find the min and max needed in Theorem (4.2.2). This will allow us to specify intervals in which the inf and sup must lie, and to study the lengths of these intervals.

(4.2.3) LEMMA: 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 $\left|\max\limits_{1 \leq i \leq n} Z_i\right|$. 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.

Proof: For $u \geq 0$, $G_{n+1}(u) \leq G_n(u)$ $(n = 1,2,\ldots)$ since

$$G_n(u) = P\left[\left|\max_{1\leq i\leq n} X_i\right| \leq u\right] = P\left[-u \leq \max_{1\leq i\leq n} X_i \leq u\right]$$

$$= P\left[\max_{1\leq i\leq n} X_i \leq u\right] - P\left[\max_{1\leq i\leq n} X_i < -u\right] = F^n(u) - [1-F(u)]^n$$

implies that

$$G_n(u) - G_{n+1}(u) = \left.\begin{cases} 0 & \text{if } n = 1 \\ F(u)[1-F(u)][F^{n-1}(u) - F^{n-1}(-u-)] & \text{if } n > 0 \end{cases}\right\} \geq 0.$$

Hence the desired result follows from Lemma (4.1.2).

(4.2.4) COROLLARY: $\int_{-\infty}^{\infty} x^2 d\{[\Phi(x)]^n\} = 1$ for $n = 1,2$ and is a strictly increasing function of n thereafter.

Proof: Choosing $h(x) = x^2$ and $F = \Phi$, by Lemma (4.2.3)

$$\int_0^\infty x^2 dG_n(x) = \int_0^\infty x^2 d\{[\Phi(x)]^n\} - \int_0^\infty x^2 d\{[1-\Phi(x)]^n\}$$

$$= \int_0^\infty x^2 d\{[\Phi(x)]^n\} + \int_{-\infty}^0 x^2 d\{[\Phi(x)]^n\} = \int_{-\infty}^\infty x^2 d\{[\Phi(x)]^n\}$$

is non-decreasing in n.

THEOREM: For any i $(1 \leq i \leq k)$, $\inf\{E_\mu(\overline{X}_{[i]} - \mu_{[i]})^2 : \mu \in \Omega_0(\mu_{[i]})\}$ is in the closed interval

$$\left((\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\} + (\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^{i}\right\},\right.$$

$$\left.(\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\}\right) \text{ if } i \geq \frac{k+1}{2}$$

$$\left((\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\} + (\sigma^2/n)\int_{-\infty}^{0} x^2 d\left\{[\Phi(x)]^{i}\right\},\right.$$

(4.2.5)

$$\left.(\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{[\Phi(x)]^{i}\right\}\right) \quad \text{ if } i < \frac{k+1}{2}$$

and $\sup\left\{E_\mu(\overline{X}_{[i]}-\mu_{[i]})^2 : \mu \in \Omega_0(\mu_{[i]})\right\}$ is in the closed interval

$$\left((\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{[\Phi(x)]^{i}\right\},\right.$$

$$\left.(\sigma^2/n)\int_{0}^{\infty} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\} + (\sigma^2/n)\int_{0}^{\infty} x^2 d\left\{[\Phi(x)]^{i}\right\}\right) \text{ if } i \geq \frac{k+1}{2}$$

$$\left((\sigma^2/n)\int_{-\infty}^{\infty} x^2 d\left\{[\Phi(x)]^{k-i+1}\right\},\right.$$

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

Proof: See Theorem (4.2.1) for the lower (upper) end points on the inf (sup), and Theorem (4.2.2) with Corollary (4.2.4) for the other end points.

<u>COROLLARY</u>: The inf and sup of Theorem (4.2.5) each lie in an interval of length

$$(\sigma^2/n)\left(\int_0^{\infty} x^2 d\{[\Phi(x)]^{k-i+1}\} - \int_{-\infty}^{0} x^2 d\{[\Phi(x)]^{i}\}\right) \text{ if } i \geq \frac{k+1}{2}$$

(4.2.6)
$$(\sigma^2/n)\left(\int_0^{\infty} x^2 d\{[\Phi(x)]^{i}\} - \int_{-\infty}^{0} x^2 d\{[\Phi(x)]^{k-i+1}\}\right) \text{ if } i < \frac{k+1}{2}.$$

By Corollary (4.2.4), 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.

CHAPTER 5. POINT ESTIMATION: MAXIMUM LIKELIHOOD (ML) AND RELATED ESTIMATORS

5.1. MLE's FOR $\mu_{[1]},\ldots,\mu_{[k]}$

Consider first maximum likelihood estimation of $\mu_1,\ldots,\mu_k$; i.e., we seek the maximum likelihood estimators (MLE's), those functions $\hat{\mu}_1,\ldots,\hat{\mu}_k$ (if such exist) such that the density of the observed statistics (whatever they may be) is maximized by setting $\mu_1 = \hat{\mu}_1, \ldots, \mu_k = \hat{\mu}_k$.

Our observed statistics under Rule (1.3.2) are X_{ij} $(i = 1,\ldots,k;$ $j = 1,\ldots,n)$, but since $\overline{X}_1,\ldots,\overline{X}_k$ are sufficient statistics we may take them as fundamental. Then

$$(5.1.1) \qquad f_{\overline{X}_1,\ldots,\overline{X}_k}(x_1,\ldots,x_k) = (\sqrt{n}/\sigma)^k \phi\left(\frac{x_1-\mu_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_k-\mu_k}{\sigma/\sqrt{n}}\right)$$

and (if $\mu_i \neq \mu_j$; $i \neq j$; $i,j = 1,\ldots,k$) the MLE's of $\mu_1,\ldots,\mu_k$ based on $\overline{X}_1,\ldots,\overline{X}_k$ exist and are uniquely

$$(5.1.2) \qquad \hat{\mu}_1 = \overline{X}_1, \ \ldots \ ,\hat{\mu}_k = \overline{X}_k.$$

(The restriction to MLE's based on $\overline{X}_1,\ldots,\overline{X}_k$ is a consequence of the general result that MLE's are functions only of sufficient statistics for a problem; see, e.g., Hogg and Craig (1965), pp. 245-246.) The problem of possible equalities among $\mu_{[1]},\ldots,\mu_{[k]}$ is discussed below; similar results hold for the case of equalities among $\mu_1,\ldots,\mu_k$.

For the problem of finding an MLE of a 1-1 function $u(\mu_1,\ldots,\mu_k)$, it is well-known that (assuming the MLE of $\mu_1,\ldots,\mu_k$ exists) $u(\hat{\mu}_1,\ldots,\hat{\mu}_k) = \hat{u}$ (say) furnishes a solution, essentially because forcing $u=\hat{u}$ implies $\mu_1 = \hat{\mu}_1, \ldots, \mu_k = \hat{\mu}_k$. (See, e.g., Hogg and Craig (1965), p. 247.) If $u(\mu_1,\ldots,\mu_k)$ is not 1-1, i.e. if it is many-to-one, points other than $\mu_1 = \hat{\mu}_1, \ldots, \mu_k = \hat{\mu}_k$ may also be implied by $u = \hat{u}$. In this case Zehna (1966) was the first to state explicitly a reason for picking only the "right" point $\mu_1 = \hat{\mu}_1, \ldots, \mu_k = \hat{\mu}_k$ for attention (and thus for calling $\hat{u}$ an MLE). Berk (1967) gives a different justification for calling $\hat{u}$ an MLE.

From the above it is clear that, based on $\overline{X}_1,\ldots,\overline{X}_k$,

(5.1.3) $\hat{\mu}_{[i]} = \{i\underline{th} \text{ smallest of } \overline{X}_1,\ldots,\overline{X}_k\} = \overline{X}_{[i]}$ $\qquad (i = 1,\ldots,k)$

is the Berk-Zehna-MLE of $\mu_{[1]},\ldots,\mu_{[k]}$. Below we discuss the problem of MLE-type estimators of $(\mu_{[1]},\ldots,\mu_{[k]})$ from another point of view. This method, Iterated-MLE's, is discussed in Section 5.2.

Blumenthal and Cohen (1968a), (1968b) (who provided the author with preliminaries of their papers) studied, for a translation parameter family, (1) estimation of the pair $(\mu_{[1]},\mu_{[2]})$ for the sum of squared errors as loss function and (2) estimation of $\mu_{[2]}$ for a squared error loss function.

Other work on the case $k = 2$, in another formulation, was done by Katz (1963), who proposed to estimate $(\mu_{[1]},\mu_{[2]})$ when one knows that (e.g.) π_1 is associated with $\mu_{[1]}$ and π_2 is associated with $\mu_{[2]}$. This work was done for binomial probabilities and also for normal means, with (e.g.) sum of squared error losses. (The fact that $(\overline{X}_1,\overline{X}_2)$ is not a totally desirable estimator may be seen intuitively from the fact that, although $\mu_{[1]} \le \mu_{[2]}$, in general $\{\overline{X}_1 > \overline{X}_2\}$ can occur with positive probability.) In our work one does not know the association of the $\mu_{[i]}$ with the π_j $(i,j = 1,\ldots,k)$; see Robertson and Waltman (1968) for the case where one does.

Blumenthal and Cohen (1968), who utilize the MLE of $\mu_{[2]}$ found below, desired their estimate to be symmetric in $\overline{X}_1,\overline{X}_2$; in order to force this they based their estimate on the maximal invariant $\overline{X}_{[1]}$, $\overline{X}_{[2]}$. Note, however, that in order to obtain symmetry in $\overline{X}_1,\overline{X}_2$ (and certain other invariance conditions) in one's estimator, one <u>need</u> not go to $\overline{X}_{[1]},\overline{X}_{[2]}$ (at least for the normal case; see (5.1.3)). Note that (although the MLE of $\mu_{[2]}$ based on $\overline{X}_1,\overline{X}_2$ is $\overline{X}_{[2]}$) the MLE of $\mu_{[2]}$ based on $\overline{X}_{[1]},\overline{X}_{[2]}$ is not. In Section 5.2 we give additional justification for basing the MLE on $\overline{X}_{[1]},\overline{X}_{[2]}$.

We will now consider the general case in which it is desired to find the MLE's of $\mu_{[1]},\ldots,\mu_{[k]}$ based on $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$. The likelihood function is given in (B.1.1), and (due to its symmetry in $\mu_{[1]},\ldots,\mu_{[k]}$) if $\hat{\mu}_{[1]},\ldots,\hat{\mu}_{[k]}$ is an MLE then so is any permutation of it (so that it is not necessarily the case that $\hat{\mu}_{[1]}\le\cdots\le\hat{\mu}_{[k]}$). In order to eliminate such undesirable occurrences, we require a consistency condition.

<u>CONSISTENCY</u> <u>CRITERION</u>: Among the (at most $k!$) permutation

(5.1.4) MLE's which any $\hat{\mu}_{[1]},\ldots,\hat{\mu}_{[k]}$ which maximizes (B.1.1) provides, only the one with $\hat{\mu}_{[1]}\le\cdots\le\hat{\mu}_{[k]}$ will be called an MLE.

From (B.1.1) and the form of $\phi(\cdot)$, it is clear that we may restrict our search for the maximum to $\mu_{[1]},\ldots,\mu_{[k]}$ such that $x_1 \le \{\mu_{[1]},\ldots,\mu_{[k]}\} \le x_k$. By (5.1.4) we need only consider the case $\mu_{[1]}\le\cdots\le\mu_{[k]}$, and not all $k!$ (fewer if there are any equalities) orderings. It is well-known (see, e.g., Hancock (1960), p. 80) that in such a case the maximum must occur at $\mu_{[1]},\ldots,\mu_{[k]}$ such that

(5.1.5) $$\frac{\partial f_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)}{\partial \mu_{[i]}} = 0 \qquad (i = 1,\ldots,k);$$

any point $\mu_{[1]},\ldots,\mu_{[k]}$ (which depends on the values of $x_1,\ldots,x_k$) where (5.1.5) holds is called a <u>critical point</u>.

In taking the derivatives (5.1.5), the results depend on how many of the k-1 inequalities $\mu_{[1]}\leq\cdots\leq\mu_{[k]}$ are equalities. There are thus 2^{k-1} mutually exclusive and exhaustive cases, say

(5.1.6) $$\Omega_o = \Omega_{(1)} + \Omega_{(2)} + \cdots + \Omega_{(2^{k-1})}$$

where the $\Omega_{(i)}$ are disjoint, $\Omega_{(1)} = \Omega(\neq)$ is defined in (1.3.12), and the $\Omega_{(i)}$ $(i = 2,\ldots,2^{k-1})$ are the other $2^{k-1} - 1$ cases in some order. Fix any i $(2 \leq i \leq 2^{k-1})$ and suppose that some $\mu^* \varepsilon\, \Omega_{(i)}$ solves the system (5.1.5) (i.e., is a critical point <u>when</u> the derivatives are taken for $\mu \varepsilon\, \Omega_{(i)}$). <u>Then</u> it is easy to verify (using (B.1.1)) that μ^* is a critical point of system (5.1.5) when derivatives are taken for $\mu \varepsilon\, \Omega_{(1)}$. We thus have the

(5.1.7) <u>THEOREM</u>: Any critical point for our problem is a solution of system (5.1.5) with derivatives taken for $\mu \varepsilon\, \Omega(\neq)$, provided only that we allow boundary points (i.e., points of $\Omega_{(2)} + \cdots + \Omega_{(2^{k-1})}$) to be considered solutions.

To completely justify calling the boundary points included in Theorem (5.1.7) critical points, one should show that any such point is a solution of system (5.1.5) when derivatives are taken for μ in its

$\Omega_{(i)}$; this is clear from the proof of Theorem (5.1.7).

Now (taking derivatives when $\mu_{[1]}<\ldots<\mu_{[k]}$) system (5.1.5) is

$$\text{(5.1.8)} \qquad \sum_{\beta\varepsilon S_k} (\sqrt{n}/\sigma)^k \phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)\frac{x_{\beta(i)}-\mu_{[i]}}{\sigma/\sqrt{n}}(\sqrt{n}/\sigma) = 0$$

$$(i = 1,\ldots,k),$$

or

$$\text{(5.1.9)} \qquad \mu_{[i]} = \frac{\sum_{\beta\varepsilon S_k} x_{\beta(i)}\phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)}{\sum_{\beta\varepsilon S_k} \phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)} \qquad (i=1,\ldots,k)$$

or

$$\text{(5.1.10)} \qquad \frac{\mu_{[j]}}{\mu_{[i]}} = \frac{\sum_{\beta\varepsilon S_k} x_{\beta(j)}\phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)}{\sum_{\beta\varepsilon S_k} x_{\beta(i)}\phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)} (i,j=1,\ldots,k;i<j).$$

(5.1.11) THEOREM: $(\hat{\mu}_{[1]},\ldots,\hat{\mu}_{[k]}) = (\bar{x},\ldots,\bar{x})$ with $\bar{x} = \frac{x_1+\ldots+x_k}{k}$ is a critical point.

Proof: It is clear that this is so from system (5.1.9).

We will now investigate the nature of this critical point. For $i,j = 1,\ldots,k$, for $x_1\leq\ldots\leq x_k$,

$$\frac{\partial^2}{\partial\mu_{[i]}\partial\mu_{[j]}} f_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)$$

(5.1.12)

$$= \begin{cases} \sum_{\beta\varepsilon S_k} \left(\frac{\sqrt{n}}{\sigma}\right)^{k+2} \phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)\cdot \\ \qquad \cdot \frac{x_{\beta(i)}-\mu_{[i]}}{\sigma/\sqrt{n}} \; \frac{x_{\beta(j)}-\mu_{[j]}}{\sigma/\sqrt{n}}, \; i \neq j \\ \sum_{\beta\varepsilon S_k} \left(\frac{\sqrt{n}}{\sigma}\right)^{k+2} \phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)\cdot \\ \qquad \cdot\left[\left(\frac{x_{\beta(i)}-\mu_{[i]}}{\sigma/\sqrt{n}}\right)^2-1\right], \; i = j. \end{cases}$$

Thus, for the matrix $Q = (d_{ij})$ of evaluations of (5.1.12) at $(\bar{x},\ldots,\bar{x})$ we find

$$d_{ij} = \left(\frac{\sqrt{n}}{\sigma}\right)^{k+4}\left[\prod_{\ell=1}^{k} \phi\left(\frac{x_\ell-\bar{x}}{\sigma/\sqrt{n}}\right)\right]\cdot\begin{cases} \sum_{\beta\varepsilon S_k} (x_{\beta(i)}-\bar{x})(x_{\beta(j)}-\bar{x}) & , \; i \neq j \\ \sum_{\beta\varepsilon S_k} \left[(x_{\beta(i)}-\bar{x})^2-\frac{\sigma^2}{n}\right] & , \; i = j \end{cases}$$

$$= (k-2)!\left(\frac{\sqrt{n}}{\sigma}\right)^{k+4}\left[\prod_{\ell=1}^{k} \phi\left(\frac{x_\ell-\bar{x}}{\sigma/\sqrt{n}}\right)\right]\cdot\begin{cases} \sum_{\substack{i,j=1 \\ i\neq j}}^{k} (x_i-\bar{x})(x_j-\bar{x}) & , \; i \neq j \\ (k-1)\sum_{i=1}^{k} (x_i-\bar{x})^2-k(k-1)\frac{\sigma^2}{n}, & i = j \end{cases}$$

(5.1.13)

$$= (k-2)!(\sqrt{n}/\sigma)^{k+4}\left[\prod_{\ell=1}^{k} \phi\left(\frac{x_\ell-\bar{x}}{\sigma/\sqrt{n}}\right)\right]\cdot\begin{cases} \mathrm{cov}(R,S)k(k-1) & , \; i \neq j \\ k(k-1)\mathrm{var}(R) - k(k-1)(\sigma^2/n), & i = j \end{cases}$$

$$= k!(\sqrt{n}/\sigma)^{k+4}\left[\prod_{\ell=1}^{k}\phi\left(\frac{x_\ell-\bar{x}}{\sigma/\sqrt{n}}\right)\right]\cdot\begin{cases}\mathrm{cov}(R,S) & , i \neq j\\ \mathrm{var}(R) - \sigma^2/n & , i = j,\end{cases}$$

where R and S are numbers selected at random (without replacement) from $\{x_1,\ldots,x_k\}$. If we let

$$\text{(5.1.14)}\qquad \begin{cases} c = c(x_1,\ldots,x_k) = k!(\sqrt{n}/\sigma)^{k+4}\left[\prod_{\ell=1}^{k}\phi\left(\frac{x_\ell-\bar{x}}{\sigma/\sqrt{n}}\right)\right]\\ d_1 = \mathrm{cov}(R,S)\cdot c\\ d_o = (\mathrm{var}(R) - \sigma^2/n)\cdot c,\end{cases}$$

then $d_{ij} = d_1$ $(i \neq j)$ and $d_{ij} = d_o$ $(i = j)$. Now, if we find the eigenvalues of Q we can utilize Theorems (A.2.1) and (A.2.2) to determine the nature of the critical point $(\bar{x},\ldots,\bar{x})$. Now

$$\text{(5.1.15)}\qquad |Q-\lambda I| = \det\begin{bmatrix} d_o-\lambda & d_1 & d_1 & \cdots & d_1 & d_1\\ d_1 & d_o-\lambda & d_1 & \cdots & d_1 & d_1\\ \cdot & & \ddots & & & \\ \cdot & & & \ddots & & \\ \cdot & & & & \ddots & \\ d_1 & d_1 & d_1 & \cdots & d_o-\lambda & d_1\\ d_1 & d_1 & d_1 & \cdots & d_1 & d_o-\lambda\end{bmatrix}$$

$$= (d_o-\lambda-d_1)^{k-1}(d_o-\lambda+(k-1)d_1)$$

where we have subtracted the last column from all others, added all rows to the last row, and taken minors. Thus, the k eigenvalues of Q are

$$\text{(5.1.16)}\qquad \begin{cases}\lambda_1 = \cdots = \lambda_{k-1} = d_o - d_1\\ \lambda_k = d_o + (k-1)d_1\end{cases}$$

and Theorems (A.2.1) and (A.2.2) give us the

<u>THEOREM</u>: The nature of the critical point $(\overline{x},\ldots,\overline{x})$ is:

(i) relative minimum if $-\frac{d_o}{k-1} < d_1 < d_o$

(ii) relative maximum if $d_o < d_1 < -\frac{d_o}{k-1}$

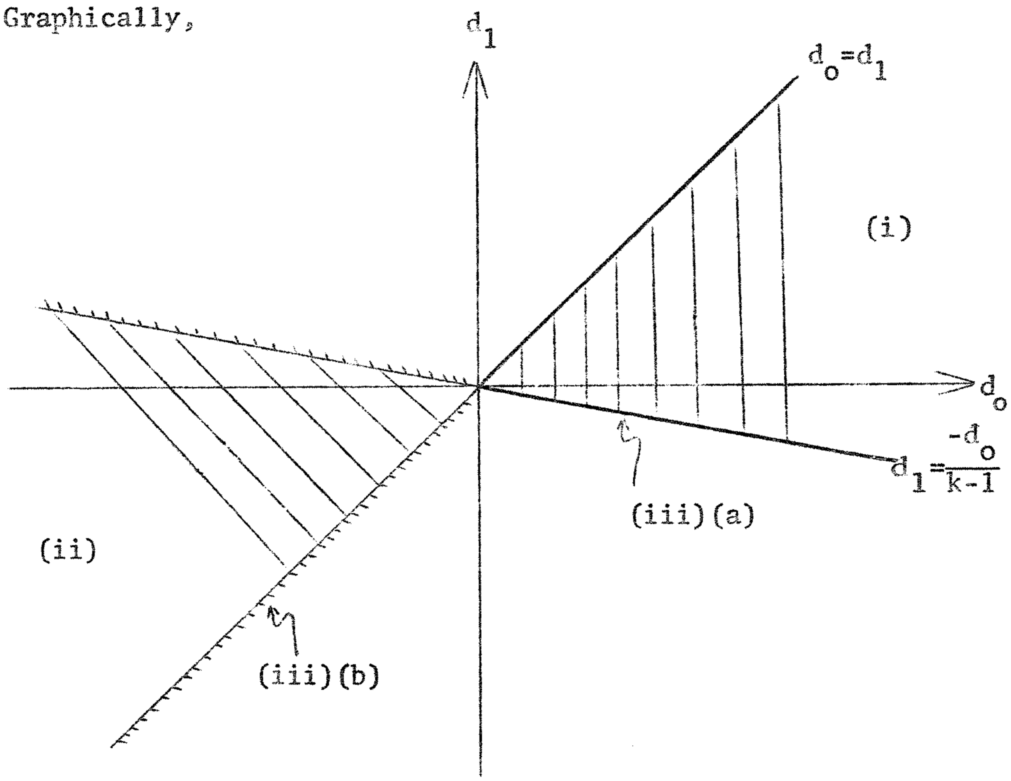
(iii) undecided if either: (a) $-\frac{d_o}{k-1} \leq d_1 = d_o$ or $-\frac{d_o}{k-1} = d_1 \leq d_o$

or: (b) $d_o = d_1 \leq \frac{-d_o}{k-1}$ or $d_o \leq d_1 = \frac{-d_o}{k-1}$

(iv) saddle point if $d_1 < \min\left(d_o, -\frac{d_o}{k-1}\right)$

(5.1.17) or if $d_1 > \max\left(d_o, -\frac{d_o}{k-1}\right)$.

Graphically,



The method of Theorem (A.2.3) can also be used to prove Theorem (5.1.17) (since the required determinants can be evaluated as in (5.1.15)), but is cumbersome.

We now wish to investigate the nature (asymptotic as $n\to\infty$ as well as small sample) of the critical point $(\bar{x},\ldots,\bar{x})$. Let $\chi^2_a(b)$ denote a non-central chi-square r.v. with "a" degrees of freedom and noncentrality "b".

THEOREM:

I. $P_\mu[(\bar{X},\ldots,\bar{X})$ is a relative minimum, or undecided$] = 0$.

II. $P_\mu[(\bar{X},\ldots,\bar{X})$ is a saddle point$] = P_\mu[\chi^2_{k-1}(\frac{1}{2}\frac{kn}{\sigma^2}\mathrm{Var}(M)) > k-1]$; otherwise $(\bar{X},\ldots,\bar{X})$ is a relative maximum. This

(5.1.18) probability does not depend on n if $\mu_{[1]}=\cdots=\mu_{[k]}$.

III. As $n\to\infty$, $P_\mu[(\bar{X},\ldots,\bar{X})$ is a saddle point$]\to 1$ unless $\mu_{[1]}=\cdots=\mu_{[k]}$ (in which case it is constant as given in II).

Proof: I. Case (i) or case (iii)(a) of Theorem (5.1.17) holds iff (see (5.1.13)) $-\frac{d_0}{k-1} \le d_1 \le d_0$, i.e. iff

$$-\frac{1}{k-1}(\mathrm{Var}(R) - \sigma^2/n) \le \mathrm{cov}(R,S) \le \mathrm{Var}(R) - \sigma^2/n,$$

i.e. iff (since $\mathrm{Var}(R) > 0$ w.p. 1)

$$(5.1.19) \qquad -\frac{1}{k-1} + \frac{\sigma^2/n}{(k-1)\mathrm{Var}(R)} \le \rho(R,S) \le 1 - \frac{\sigma^2/n}{\mathrm{Var}(R)}.$$

Since (w.p. 1) $\rho(R,S) = \frac{-1}{k-1}$, w.p. 1 equation (5.1.19) fails to hold. W.p. 1 case (iii)(b) fails to hold since (for it to hold) at least one of the inequalities in (5.1.19) must be an equality; this occurs w.p. 0.

II. As in I, it can be seen that case (ii) holds iff

$$(5.1.20) \qquad 1 - \frac{\sigma^2/n}{\operatorname{Var}(R)} < \rho(R,S) < \frac{-1}{k-1} + \frac{\sigma^2/n}{(k-1)\operatorname{Var}(R)}.$$

Since the r.h.s. of (5.1.20) holds w.p. 1, case (ii) holds iff

$$(5.1.21) \qquad 1 - \frac{\sigma^2/n}{\operatorname{Var}(R)} < \frac{-1}{k-1},$$

i.e. iff $\operatorname{Var}(R)\frac{k}{k-1} < \sigma^2/n$; otherwise (by I) case (iv) must hold. Now from Graybill (1961), p. 88 (Theorem 4.20), p. 91 (Problem 4.24),

$\operatorname{Var}(R) = (1/k)\sum_{i=1}^{k}(\overline{X}_i-\overline{X})^2$ is $(\sigma^2/(nk))\chi^2_{k-1}(\lambda)$ with

$$(5.1.22) \qquad \begin{aligned} \lambda &= \frac{1}{2}\frac{kn}{\sigma^2}\left(\frac{\Sigma\mu_i^{\,2}}{k} - \frac{(\Sigma\mu_i)^2}{k^2}\right) \\ &= \frac{1}{2}\frac{kn}{\sigma^2}\operatorname{Var}(M), \end{aligned}$$

where M is a number selected at random from $\{\mu_1,\ldots,\mu_k\}$. Thus,

$$(5.1.23) \qquad \begin{aligned} &P_\mu[(\overline{X},\ldots,\overline{X}) \text{ is a relative maximum}] = P_\mu[\operatorname{Var}(R) > \frac{k-1}{k}\frac{\sigma^2}{n}] \\ &= P_\mu[\frac{\sigma^2}{nk}\chi^2_{k-1}(\lambda) > \frac{k-1}{k}\frac{\sigma^2}{n}] = P_{\hat\mu}[\chi^2_{k-1}(\frac{1}{2}\frac{kn}{\sigma^2}\operatorname{Var}(M)) > k-1]. \end{aligned}$$

III. This follows from II.

Note that even when $(\overline{X},\ldots,\overline{X})$ is a relative maximum it is not necessarily an absolute one (which it would be if, e.g., the system had no other solution). Below we will find reason to believe that the maximum is "near" $(\hat\mu_{[1]},\ldots,\hat\mu_{[k]}) = (\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$.

For the case $k = 2$, Theorem (5.1.17) shows (after some reduction) that $(\bar{x},\bar{x})$ is

$$(5.1.24)\qquad \begin{cases} \text{a relative maximum} & \text{iff } (x_1-x_2)^2 < 2\sigma^2/n \\ \text{undecided (negative semi-definite)} & \text{iff } (x_1-x_2)^2 = 2\sigma^2/n \\ \text{a saddle point} & \text{iff } (x_1-x_2)^2 > 2\sigma^2/n. \end{cases}$$

Obtaining this result from Theorem (A.1.1) is interesting. The limiting results of Theorem (5.1.18) can, for the case $k = 2$, be obtained using (5.1.24).

We will now seek the MLE (for $k \geq 2$): We may (without loss) choose our estimator to be of the form

$$(5.1.25)\qquad \begin{cases} \hat{\mu}_{[1]} = x_1 + a_1(x_1,\ldots,x_k) \\ \vdots \\ \hat{\mu}_{[k]} = x_k + a_k(x_1,\ldots,x_k). \end{cases}$$

As noted following (5.1.4), we may restrict ourselves without loss to $x_1 \leq \{\hat{\mu}_{[1]},\ldots,\hat{\mu}_{[k]}\} \leq x_k$, from which it follows that we have

$$(5.1.26)\qquad \begin{cases} 0 \leq a_1 \\ -(x_i-x_1) \leq a_i \leq (x_k-x_i) \quad (i=1,\ldots,k) \\ a_k \leq 0. \end{cases}$$

Let (for $1 \leq \ell \leq k$; $i = 1,\ldots,k$)

$$(5.1.27)\qquad \begin{cases} A_\ell(i) = \displaystyle\sum_{\substack{\beta\varepsilon S_k \\ \beta(i)=\ell}} \phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-x_k-a_k}{\sigma/\sqrt{n}}\right) \\ A \quad\;\; = \displaystyle\sum_{\beta\varepsilon S_k} \phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-x_k-a_k}{\sigma/\sqrt{n}}\right). \end{cases}$$

Then (note that, for any $1 \le i \le k$, $A = A_1(i) + \ldots + A_k(i)$) from system (5.1.9) we find that $a_1,\ldots,a_k$ must satisfy the system

(5.1.28) $$(x_i + a_i)A = x_1A_1(i) + \ldots + x_kA_k(i) \qquad (i = 1,\ldots,k).$$

If we add the terms of (5.1.28) over $i = 1,\ldots,k$, we obtain (since $A = A_\ell(1) + \ldots + A_\ell(k)$ for $\ell = 1,\ldots,k$)

$$A(x_1+\ldots+x_k) + (a_1+\ldots+a_k)A = A(x_1+\ldots+x_k),$$

or (since $A > 0$) $a_1+\ldots+a_k = 0$. Thus, we have the

<u>THEOREM</u>: For $k \ge 2$, the MLE is given by $\hat{\mu}_{[1]} = \bar{X}_{[1]} + a_1(\bar{X}_{[1]},\ldots,\bar{X}_{[k]}), \ldots, \hat{\mu}_{[k]} = \bar{X}_{[k]} + a_k(\bar{X}_{[1]},\ldots,\bar{X}_{[k]})$,

(5.1.29) where $a_1,\ldots,a_k$ are some solution of system (5.1.28) and must satisfy

$$-(x_i-x_1) \le a_i \le (x_k-x_i) \qquad (i = 1,\ldots,k)$$

and

$$a_1+\ldots+a_k = 0.$$

<u>THEOREM</u>: For $i,j = 1,\ldots,k$, if $a_j \ne 0$ then

(5.1.30) $$a_i = a_j\frac{d_{1i}A_1(i) + \ldots + d_{ki}A_k(i)}{d_{1j}A_1(j) + \ldots + d_{kj}A_k(j)}$$

where $d_{ij} = x_i - x_j = -d_{ji} (i,j = 1,\ldots,k)$.

<u>Proof</u>: System (5.1.28) is equivalent to the system

$$a_i \sum_{\beta\in S_k} \phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-x_k-a_k}{\sigma/\sqrt{n}}\right)$$

$$= \sum_{\beta\varepsilon S_k} (x_{\beta(i)}-x_i)\phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-x_k-a_k}{\sigma/\sqrt{n}}\right) \qquad (i = 1,\ldots,k),$$

or (substituting the d_{ij}'s)

$$a_i(A_1(i)+\ldots+A_k(i)) = d_{1i}A_1(i) + \ldots + d_{ki}A_k(i) \qquad (i = 1,\ldots,k).$$

Thus, the theorem follows. (Note that the denominator $d_{1j}A_1(j) + \ldots + d_{kj}A_k(j)$ is zero iff $a_j = 0$.)

(5.1.31) LEMMA: For the case $k = 2$, $a_1 = -a_2$. Also, $0 \leq a_1 \leq x_2 - x_1$.

Proof: From Theorem (5.1.30),

$$a_1 = a_2\frac{d_{11}A_1(1) + d_{21}A_2(1)}{d_{12}A_1(2) + d_{22}A_2(2)} = a_2\frac{d_{21}A_2(1)}{d_{12}A_1(2)} = -a_2\frac{A_2(1)}{A_1(2)} = a_2.$$

The theorem follows from Theorem (5.1.29).

(5.1.32) LEMMA: Let $d = x_2 - x_1 \geq 0$. Then the MLE for $k = 2$ is given by $\hat{\mu}_{[1]} = \bar{X}_{[1]} + a_1(\bar{X}_{[1]},\bar{X}_{[2]})$, $\hat{\mu}_{[2]} = \bar{X}_{[2]} - a_1(\bar{X}_{[1]},\bar{X}_{[2]})$ where a_1 is some root of

$$d = a_1\left(1 + e^{\frac{d^2-2a_1d}{\sigma^2/n}}\right)$$

and $0 \leq a_1 \leq d$.

Proof: By Lemma (5.1.31) we must have $0 \leq a_1 = -a_2 \leq d$. Then by Theorem (5.1.29), the MLE must be of the form given where a_1 is some root of the system (5.1.28):

$$\begin{cases}(x_1 + a_1)A = x_1A_1(1) + x_2A_2(1)\\(x_2 - a_1)A = x_1A_1(2) + x_2A_2(2)\end{cases}$$

$$\begin{cases}x_1A_2(1) + a_1A = x_2A_2(1)\\x_2A_1(2) - a_1A = x_1A_1(2)\end{cases}$$

$$a_1A = dA_2(1) = dA_1(2)$$

$$a_1A = dA_1(2)$$

$$a_1 = d\frac{A_1(2)}{A_1(2) + A_2(2)}$$

$$a_1 = \frac{d}{1 + \frac{A_2(2)}{A_1(2)}}.$$

Now

$$A_1(2) = \sum_{\substack{\beta\varepsilon S_2\\ \beta(2)=1}} \phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\phi\left(\frac{x_{\beta(2)}-x_2-a_2}{\sigma/\sqrt{n}}\right) = \phi\left(\frac{d-a_1}{\sigma/\sqrt{n}}\right)\phi\left(\frac{a_1-d}{\sigma/\sqrt{n}}\right)$$

$$= \frac{1}{2\pi\sigma^2/n}\, e^{-\frac{(d-a_1)^2}{\sigma^2/n}};$$

$$A_2(2) = \sum_{\substack{\beta\varepsilon S_2\\ \beta(2)=2}} \phi\left(\frac{x_{\beta(1)}-x_1-a_1}{\sigma/\sqrt{n}}\right)\phi\left(\frac{x_{\beta(2)}-x_2-a_2}{\sigma/\sqrt{n}}\right) = \phi\left(\frac{-a_1}{\sigma/\sqrt{n}}\right)\phi\left(\frac{a_1}{\sigma/\sqrt{n}}\right)$$

$$= \frac{1}{2\pi\sigma^2/n}\, e^{-\frac{a_1^2}{\sigma^2/n}}.$$

Thus,

$$\frac{A_2(2)}{A_1(2)} = e^{-\frac{a_1^{\,2}}{\sigma^2/n} + \frac{(d-a_1)^2}{\sigma^2/n}} = e^{\frac{d^2-2a_1 d}{\sigma^2/n}}$$

and the lemma follows.

(5.1.33) LEMMA: For fixed d and $0 \le a_1 \le d$, the roots of

$$d = a_1\left(1 + e^{\frac{d^2-2a_1 d}{\sigma^2/n}}\right) \tag{5.1.34}$$

are (1) $a_1 = d/2$, and (2) $a_1 = \frac{d}{2} + \frac{\varepsilon_o}{2d}\sigma^2/n$ if $d > \sqrt{2}\sigma/\sqrt{n}$.

Here ε_o is either of the two solutions of

$$d^2 n/\sigma^2 = \varepsilon \coth(\varepsilon/2). \tag{5.1.35}$$

Proof: First, $a_1 = d/2$ is seen to satisfy (5.1.34). Now, suppose there is another solution of (5.1.34), say (without loss of generality)

$$a_1 = d/2 + \frac{\varepsilon}{2d}\,\sigma^2/n$$

with $-d^2n/\sigma^2 \le \varepsilon \le d^2n/\sigma^2$ (since $0 \le a_1 \le d$), $\varepsilon \ne 0$. Substituting in (5.1.34), we find ε must satisfy

$$d = \left(\frac{d}{2} + \frac{\varepsilon}{2d}\frac{\sigma^2}{n}\right)\left(1 + e^{\frac{d^2-2\left(\frac{d}{2}+\frac{\varepsilon}{2d}\frac{\sigma^2}{n}\right)d}{\sigma^2/n}}\right) = \left(\frac{d}{2} + \frac{\varepsilon}{2d}\frac{\sigma^2}{n}\right)\left(1 + e^{-\varepsilon}\right)$$

$$= \frac{1}{2}\{d + de^{-\varepsilon} + \frac{\varepsilon}{d}\frac{\sigma^2}{n} + \frac{\varepsilon}{d}\frac{\sigma^2}{n}e^{-\varepsilon}\},$$

or

$$d^2 = d^2 e^{-\varepsilon} + \varepsilon\frac{\sigma^2}{n} + \varepsilon\frac{\sigma^2}{n}e^{-\varepsilon},$$

or (since $\varepsilon \ne 0 \Rightarrow 1 - e^{-\varepsilon} \ne 0$)

$$d^2 = \varepsilon\frac{\sigma^2}{n}\,\frac{1 + e^{-\varepsilon}}{1 - e^{-\varepsilon}} = \varepsilon\frac{\sigma^2}{n}\,\frac{e^{\varepsilon/2} + e^{-\varepsilon/2}}{e^{\varepsilon/2} - e^{-\varepsilon/2}} = \varepsilon\frac{\sigma^2}{n}\coth(\varepsilon/2).$$

(See, e.g., Hodgman (1959), pp. 281, 427, 431, 432.) Since $\coth(-z) = -\coth(z)$, $\varepsilon \coth(\varepsilon/2)$ is an even function. Now,

$$\lim_{\varepsilon\to o} \varepsilon\coth(\varepsilon/2) = \lim_{\varepsilon\to o}(1 + e^{-\varepsilon}) \cdot \lim_{\varepsilon\to o}\frac{\varepsilon}{1 - e^{-\varepsilon}} = 2\lim_{\varepsilon\to o}\frac{1}{e^{-\varepsilon}} = 2.$$

(See, e.g., Apostol (1957), p. 102.) Since

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}[\varepsilon\coth(\varepsilon/2)] &= \coth(\varepsilon/2) - (\varepsilon/2)\operatorname{csch}^2(\varepsilon/2)\\ &= \frac{\cosh(\varepsilon/2)}{\sinh(\varepsilon/2)} - (\varepsilon/2)\frac{1}{\sinh^2(\varepsilon/2)}\\ &= \frac{1}{\sinh(\varepsilon/2)}\left\{\cosh(\varepsilon/2) - \frac{\varepsilon/2}{\sinh(\varepsilon/2)}\right\},\end{aligned}$$

the facts $\sinh(\varepsilon/2) > 0$ if $\varepsilon > 0$ and

$$\begin{aligned}\cosh(\varepsilon/2) - \frac{\varepsilon/2}{\sinh(\varepsilon/2)} &= \frac{1}{\sinh(\varepsilon/2)}[\sinh(\varepsilon/2)\cosh(\varepsilon/2) - \varepsilon/2]\\ &= \frac{1}{\sinh(\varepsilon/2)}\left[\frac{\sinh(\varepsilon)}{2} - \varepsilon/2\right]\\ &= \frac{1}{2\sinh(\varepsilon/2)}\left[\varepsilon + \frac{\varepsilon^3}{3!} + \frac{\varepsilon^5}{5!} + \cdots - \varepsilon\right]\\ &= \frac{1}{2\sinh(\varepsilon/2)}\left[\frac{\varepsilon^3}{3!} + \frac{\varepsilon^5}{5!} + \frac{\varepsilon^7}{7!} + \cdots\right] > 0\end{aligned}$$

imply that $\frac{\partial}{\partial\varepsilon}[\varepsilon\coth(\varepsilon/2)] > 0$. Combining the above information, we may plot Figure (5.1.36).

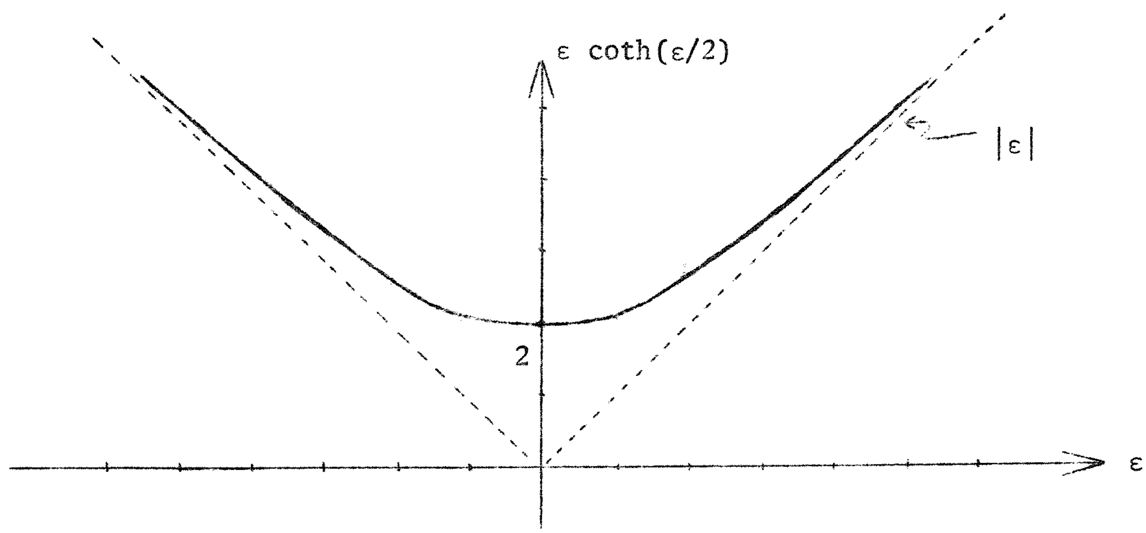


Figure (5.1.36).

Since $\coth(x) > 1$ for $x > 0$, the range of $\varepsilon \coth(\varepsilon/2)$ will be $[2, d^2\frac{n}{\sigma^2}{}^{+}]$ when ε is in $[-d^2n/\sigma^2, d^2n/\sigma^2]$. Thus, there will be two additional solutions if $d^2n/\sigma^2 > 2$ and none if $d^2n/\sigma^2 \leq 2$.

Note that $a_1 = 0$ corresponds to the estimator (x_1,x_2); $a_1 = d/2$ corresponds to $(\bar{x},\bar{x})$; and $a_1 = d$ corresponds to (x_2,x_1). Consistency Criterion (5.1.4) rules out values $a_1 > d/2$; thus, in seeking the MLE we only consider ε_o which is the negative solution of (5.1.35) in Theorem (5.1.33) (or, what is the same, $-\varepsilon_o$ where ε_o is the positive solution).

<u>THEOREM</u>: If $0 \leq d \leq \sqrt{2}\,\sigma/\sqrt{n}$, $(\bar{x},\bar{x})$ is the only critical point and is the MLE.

If $d > \sqrt{2}\,\sigma/\sqrt{n}$, there are two critical points. One

(5.1.37) yields $(\bar{x},\bar{x})$ and is a saddle point. The other yields the MLE

(5.1.38) $\left(\bar{x} - \frac{\varepsilon_o}{2d}\sigma^2/n,\ \bar{x} + \frac{\varepsilon_o}{2d}\sigma^2/n\right)$,

where ε_o is the positive solution of

(5.1.39) $d^2n/\sigma^2 = \varepsilon \coth(\varepsilon/2)$.

Theorem (5.1.37) follows from previous results, notably Lemma (5.1.32) for the form of the MLE, Lemma (5.1.33) for the solutions of a certain equation, and (5.1.24) for the nature of $(\bar{x},\bar{x})$. In obtaining the form of (5.1.38), relations such as

$$\hat{\mu}_{[1]} = x_1 + a_1 = x_1 + \frac{d}{2} - \frac{\varepsilon_o}{2d}\sigma^2/n$$

$$= \bar{x} - \frac{\varepsilon_o}{2d}\sigma^2/n$$

are used. Note that, for d^2n/σ^2 "large," $\varepsilon_o \approx d^2n/\sigma^2$, so that (5.1.38) is "close" to (x_1,x_2). The following lemma studies the approach of ε_o to $d^2\frac{n}{\sigma^2}$.

LEMMA: If ε_o is the positive solution of (5.1.39), then (with $o(n) \geq 0$)

(5.1.40) $$\varepsilon_o = \frac{d^2n}{\sigma^2} - o(n).$$

Proof: If we write $a = d^2/\sigma^2$, then we are interested in the positive solution of $\varepsilon \coth(\varepsilon/2) = a\cdot n$. Let us set this solution as $\varepsilon_o = a\cdot n - c_n$ and investigate the order of c_n. Substituting in the equation,

$$(a\cdot n - c_n)\coth\left(\frac{a\cdot n - c_n}{2}\right) = a\cdot n$$

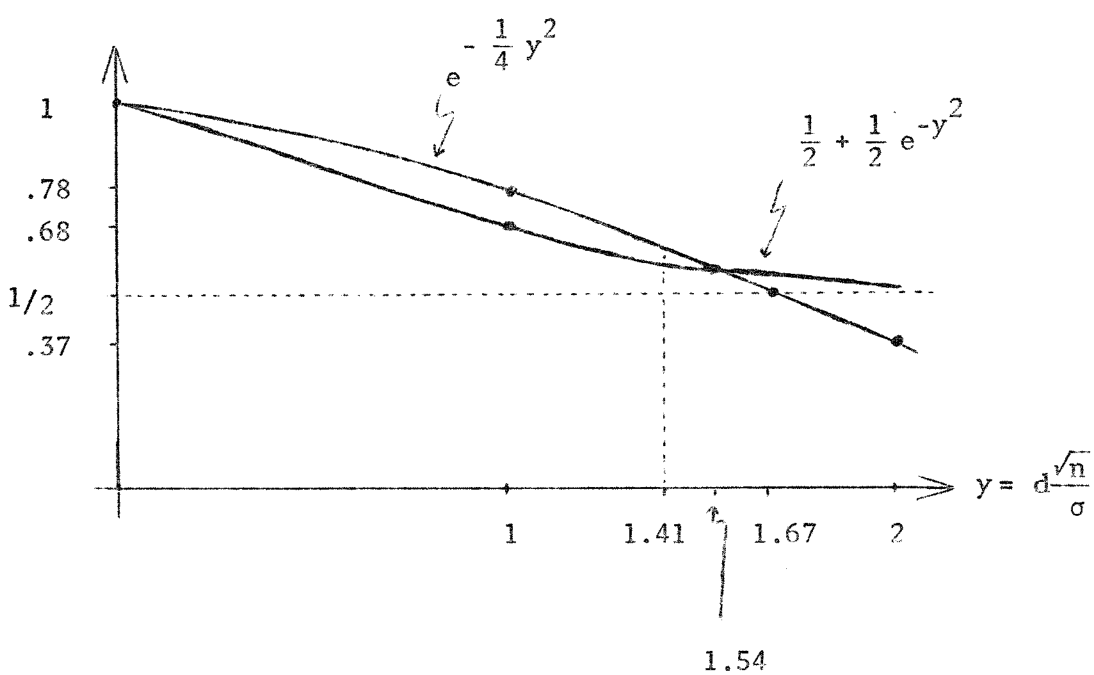
or

$$(5.1.41) \qquad \left(1 - \frac{c_n}{a\cdot n}\right)\coth\left(\frac{a\cdot n - c_n}{2}\right) = 1.$$

From Figure (5.1.36) we see that $\varepsilon_o \to \infty$ as $n\to\infty$, and since $\varepsilon_o > 0$ we have $c_n < a\cdot n$ or $\frac{c_n}{n} < a$. Since $\coth(x) > 1$ if $x > 0$, and since (5.1.41) must be satisfied, $\frac{c_n}{n} > 0$. Now, taking the limit of (5.1.41) as $n\to\infty$, we find that

$$(1 - b/a)\cdot 1 = 1$$

where $0 \le b = \lim_{n\to\infty} \frac{c_n}{n} \le a$. This is a contradiction unless $\lim_{n\to\infty} \frac{c_n}{n} = 0$, so that $c_n = o(n)$.



It is of interest to compare (for the case $k = 2$) the likelihoods of the three estimators $(\bar{X},\bar{X})$, $(\bar{X}_{[1]},\bar{X}_{[2]})$, and the MLE. With $d = x_2 - x_1$, we find (see (B.1.1))

$$\pi\frac{\sigma^2}{n} f_{\bar{X}_{[1]},\bar{X}_{[2]}}(x_1,x_2)$$

$$= \pi\phi\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\phi\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right) + \pi\phi\left(\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\phi\left(\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}\right)$$

(5.1.42)
$$= \begin{cases} e^{-\frac{d^2}{4\sigma^2/n}} & \text{if } (\mu_{[1]},\mu_{[2]}) = (\bar{x},\bar{x}), \text{ the MLE for } 0 \le d^2 \le 2\sigma^2/n \\[2ex] \frac{1}{2} + \frac{1}{2} e^{-\frac{d^2}{\sigma^2/n}} & \text{if } (\mu_{[1]},\mu_{[2]}) = (x_1,x_2) \\[2ex] \frac{1}{2} e^{-\frac{1}{4\sigma^2/n}\left(\frac{\varepsilon_o}{d}\frac{\sigma^2}{n} - d\right)^2} + \frac{1}{2} e^{-\frac{1}{4\sigma^2/n}\left(\frac{\varepsilon_o}{d}\frac{\sigma^2}{n} + d\right)^2} & \\ \quad \text{if } (\mu_{[1]},\mu_{[2]}) = \left(\bar{x} - \frac{\varepsilon_o}{2d}\frac{\sigma^2}{n}, \bar{x} + \frac{\varepsilon_o}{2d}\frac{\sigma^2}{n}\right), & \text{the MLE for } d^2 > 2\sigma^2/n. \end{cases}$$

If $0 \le d\sqrt{n}/\sigma \le \sqrt{2}$, $(\bar{X},\bar{X})$ is the MLE, and the curve of $(\bar{X},\bar{X})$ has ordinate 1/2 when $d\sqrt{n}/\sigma = 2\sqrt{\ell n 2} \simeq 1.67$. The curves of $(\bar{X},\bar{X})$ and $(\bar{X}_{[1]},\bar{X}_{[2]})$ cross at $d\sqrt{n}/\sigma \simeq 1.54$. At $d\sqrt{n}/\sigma = 2$, for $(\bar{X}_{[1]},\bar{X}_{[2]})$ we find $\frac{1}{2} + \frac{1}{2} e^{-y^2} = \frac{1}{2} + \frac{1}{2}(.01831) = .5092$, while for the MLE, a solution of

$4 = \varepsilon \coth(\varepsilon/2)$ is approximately $\varepsilon_o = 3.8$ (thus $\varepsilon_o/4 = .95$) and $1 - \varepsilon_o/4 = .05$. (See Abramowitz and Stegun (1964), p. 216.) Thus, for the MLE we find

$$\frac{1}{2} e^{-\frac{y^2}{4}\left(\frac{\varepsilon_o}{y^2} - 1\right)^2} + \frac{1}{2} e^{-\frac{y^2}{4}\left(\frac{\varepsilon_o}{y^2} + 1\right)^2} = \frac{1}{2}\left\{e^{-.0025} + e^{-3.8025}\right\}$$

$$\geq \frac{1}{2}\left\{e^{-.003} + e^{-3.81671}\right\} \simeq \frac{1}{2}(1.019) = .5095.$$

Note that Theorem (2.1.33) indicates the reasonableness of an estimator which compensates, as does the MLE = $(x_1 + a,\ x_2 - b)$, for under and over estimation with regard to expectation; the likelihood approach bears this out.

The above results indicate a weakness of taking a function of MLE's to estimate that function of the parameters for a problem (as discussed at (5.1.3)): namely, other methods yield <u>different</u> estimators with <u>higher</u> likelihoods. (In fact, with the other method the likelihood could never exceed $\frac{1}{2\pi}n/\sigma^2$; with our method it can never be less than $\frac{1}{2\pi}n/\sigma^2$.)

CHAPTER 5. POINT ESTIMATION: MAXIMUM LIKELIHOOD (ML) AND RELATED ESTIMATORS

5.2. MLE's FOR NON-1-1 FUNCTIONS: ITERATED MLE's (IMLE's)

At (5.1.3), we discussed the problem of providing maximum likelihood estimators (MLE's) for $\mu_{[1]},\ldots,\mu_{[k]}$, and noted the Berk-Zehna-MLE; most of the remainder of Section 5.1 was devoted to a study of another method of providing MLE's for $\mu_{[1]},\ldots,\mu_{[k]}$. We now formulate this latter method as a general inference principle and study it in some specific cases.

Suppose that θ (a parameter of interest) is in some space Θ and that we have a likelihood function $L(\theta)$ (from Θ to R). Assume that a unique MLE $\hat{\theta}$ of θ exists, i.e. $\hat{\theta} \varepsilon \Theta$ such that $L(\hat{\theta}) \geq L(\theta)$ for all $\theta \varepsilon \Theta$. Let $g(\circ)$ be some transformation of Θ, and suppose that $g(\Theta) = \Lambda$. Then if $g(\circ)$ is 1-1, $g(\hat{\theta})$ is clearly the MLE of θ. If $g(\circ)$ is not 1-1, Zehna (1966) and Berk (1967) both propose to employ the estimator $g(\hat{\theta})$, which we will call the Berk-Zehna-MLE.

Zehna proposes to use $g(\hat{\theta})$ since, if with $g(\theta)$ one associates the largest of the likelihoods of those θ' such that $g(\theta') = g(\theta)$, this "induced likelihood function" is maximized at $g(\hat{\theta})$. However, as Dr. Joseph Putter has pointed out in a personal communication, $g(\hat{\theta})$ may also be a minimum likelihood estimator. E.g., if (for some observations) we have the possibilities as given in Table (5.2.1),

Table (5.2.1)

θ	-2	-1	1	2
$L(\theta)$	.8	.7	.7	0

then $\hat{\theta}=-2$ is the MLE of θ, but if $g(\theta) = \theta^2$, then $g(\hat{\theta}) = 4$ corresponds to both a minimum likelihood estimator of θ and a maximum likelihood estimator of θ.

Berk proposes to use $g(\hat{\theta})$ since, if one simply adjoins to $g(\theta)$ another function $h(\theta)$ so that the mapping $\theta \to (g(\theta),h(\theta))$ is 1-1, then $(g(\hat{\theta}),h(\hat{\theta}))$ is the MLE of $(g(\theta),h(\theta))$. Berk states his belief that it is important that one's estimate maximize the likelihood function associated with some r.v.; and since it is not clear that Zehna's method does this, Zehna "misses the point." (Note that the Iterated MLE proposed below satisfies this criterion.) Berk's reasoning seems faulty in that, if one desires to estimate $g(\theta)$, there seems to be no reason to be concerned with any 1-1-izing function $h(\theta)$. Rather, $h(\theta)$ is added to preserve the status of $g(\hat{\theta})$ as an "MLE." (E.g., in Putter's example of Table (5.2.1), $h(\theta) = \mathrm{sgn}(\theta)$ will work but is irrelevant to the problem of estimating $g(\theta) = \theta^2$.)

Let θ, $\hat{\theta}$, Θ, Λ, $L(\theta)$, and $g(\theta)$ be as defined above. (In particular, we suppose that $\hat{\theta}$ exists and is unique.) We then propose the

(5.2.2) <u>DEFINITION:</u> Consider the likelihood function of the statistic $g(\hat{\theta})$, say L_g. If there is a $\bar{g} \in \Lambda$ such that $L_g(\bar{g}) \geq L_g(g')$ for all $g' \in \Lambda$, then $\bar{g}$ is called an <u>Iterated MLE</u> (IMLE) of $g(\theta)$.

Thus, the IMLE of $g(\theta)$ is the MLE of $g(\theta)$ based on $g(\hat{\theta})$ (if it exists and is unique).

<u>Example 1</u>. For the problem of estimating $g(\mu_1,\ldots,\mu_k) = (\mu_{[1]},\ldots,\mu_{[k]})$, the Berk-Zehna MLE is $\bar{X}_{[1]},\ldots,\bar{X}_{[k]}$, and in Section

5.1 we studied the IMLE (i.e., the MLE of $\mu_{[1]},\ldots,\mu_{[k]}$ based on $g(\overline{X}_1,\ldots,\overline{X}_k) = (\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$). For the case $k = 2$, Blumenthal and Cohen (1968) have compared the Berk-Zehna MLE of $\mu_{[2]}$ with our IMLE of $\mu_{[2]}$, with regard to mean squared error and bias. Let $\omega = (\mu_{[2]} - \mu_{[1]})/2$. They find that, for both mean squared error and for bias, the IMLE is better for ω small, and $\overline{X}_{[2]}$ is better for ω moderate.

Example 2. Let $Y_1,\ldots,Y_n$ be i.i.d. $N(\mu,\sigma^2)$ r.v.'s with μ and σ^2 both unknown ($-\infty < \mu < +\infty$, $\sigma^2 > 0$). The MLE of (μ,σ^2) is well-known: $(\overline{Y}, \sum_{i=1}^{n}(Y_i-\overline{Y})/n)$. Then for estimation of $g(\mu,\sigma^2) = \mu$, the Berk-Zehna MLE (which is $\overline{Y}$) and the IMLE (which is the MLE of μ based on $\overline{Y}$) coincide. Such coincidence occurs in many other cases, for example when our r.v.'s are uniform on $(0,\theta)$.

Example 3. Let $Y_1,\ldots,Y_n$ be i.i.d. $N(\mu,\sigma^2)$ r.v.'s with μ unknown ($-\infty < \mu < +\infty$) and σ^2 known ($\sigma^2 > 0$). The MLE of μ is well-known: $\overline{Y}$. Then for estimation of $g(\mu) = \mu^2$, the Berk-Zehna MLE is $\overline{Y}^2$. We will now study the IMLE (which is the MLE of μ^2 based on $\overline{Y}^2$).

Since $(\sqrt{n}/\sigma)\overline{Y}$ is $N((\sqrt{n}/\sigma)\mu,1)$, $((\sqrt{n}/\sigma)\overline{Y})^2$ is (see, e.g., Fisz (1963), p. 343) a non-central chi-square r.v. with 1 d.f. and non-centrality $\lambda = \frac{n}{2\sigma^2}\mu^2$ say $\chi_1^2(\lambda)$, and has density (for $x \geq 0$)

$$f_1(x) = \frac{x^{-\frac{1}{2}}e^{-\lambda-\frac{x}{2}}}{\sqrt{2\pi}}\sum_{m=0}^{\infty}\frac{(2\lambda x)^m}{(2m)!} = \frac{x^{-\frac{1}{2}}e^{-\lambda-\frac{x}{2}}}{\sqrt{2\pi}}\sum_{m=0,2,4,\ldots}\frac{(\sqrt{2\lambda x})^m}{m!}$$

$$= \frac{x^{-\frac{1}{2}}e^{-\frac{x}{2}}}{\sqrt{2\pi}}e^{-\lambda}\cosh(\sqrt{2\lambda x}).$$

Thus, $\overline{Y}^2 = \frac{\sigma^2}{n}\left(\frac{\sqrt{n}}{\sigma}\overline{Y}\right)^2$ has density (for $y \geq 0$)

$$f_{\overline{Y}^2}(y) = \frac{1}{\sigma^2/n} f_1\left(\frac{y}{\sigma^2/n}\right) = \frac{y^{-\frac{1}{2}} e^{-\frac{n}{2\sigma^2}y}}{\sqrt{2\pi}\ \sigma/\sqrt{n}} e^{-\frac{n}{2\sigma^2}\mu^2} \cosh\left(\sqrt{\frac{n^2}{\sigma^4} y\mu^2}\right).$$

Hence (when $\overline{Y}^2 = y \geq 0$) the IMLE of μ^2 is the μ^2 which maximizes

$$e^{-\frac{n}{2\sigma^2}\mu^2} \cosh\left(\sqrt{\frac{n^2}{\sigma^4} y\mu^2}\right),$$

or $\widehat{\mu^2} = a^2(\overline{Y}^2)\overline{Y}^2$ where a^2 is the a^2 which maximizes

$$g(a^2) = e^{-\frac{n}{2\sigma^2}ya^2} \cosh\left(\frac{n}{\sigma^2} y\sqrt{a^2}\right). \tag{5.2.3}$$

Differentiating $g(a^2)$ with respect to a^2, we find

$$\frac{\partial g(a^2)}{\partial (a^2)} = -\frac{n}{2\sigma^2} y e^{-\frac{n}{2\sigma^2}ya^2} \cosh\left(\frac{n}{\sigma^2} y\sqrt{a^2}\right) + e^{-\frac{n}{2\sigma^2}ya^2} \sinh\left(\frac{n}{\sigma^2} y\sqrt{a^2}\right)\frac{n}{\sigma^2} y \frac{1}{2\sqrt{a^2}},$$

which is <0 iff $a > \tanh\left(\frac{n}{\sigma^2} ya\right)$. Since $\tanh(z) < 1$ for all z $(-\infty < z < \infty)$, the derivative is negative for all $a \geq 1$, so we may seek the maximum of (5.2.3) for $0 \leq a < 1$. Then, $a > \tanh\left(\frac{n}{\sigma^2} ya\right)$ iff $\tanh^{-1}(a) > \frac{n}{\sigma^2} ya$, which is so (see, e.g., Hodgman (1959), p. 431) iff

$$a + \frac{a^3}{3} + \frac{a^5}{5} + \ldots > \frac{n}{\sigma^2} ya,$$

i.e. iff

$$a^2\left[\frac{1}{3} + \frac{a^2}{5} + \frac{a^4}{7} + \ldots\right] > \frac{n}{\sigma^2} y - 1. \tag{5.2.4}$$

Since (5.2.4) holds for all a $(0 < a < 1)$ if $\frac{n}{\sigma^2} y - 1 \leq 0$, i.e. if $y \leq \frac{\sigma^2}{n}$, the IMLE of μ^2 is 0 if $\overline{Y}^2 \leq \sigma^2/n$. If $y > \sigma^2/n$, it is clear that there will be one critical point (corresponding to equality in (5.2.4)) and that it will be a maximum. Thus, the IMLE of μ^2 is

(5.2.5) $$\widehat{\mu^2} = \begin{cases} 0 & \text{if } \bar{Y}^2 \leq \sigma^2/n \\ a^2(\bar{Y}^2)\bar{Y}^2 & \text{if } \bar{Y}^2 > \sigma^2/n, \end{cases}$$

where a is the root of

(5.2.6) $$a^2\left[\frac{1}{3} + \frac{a^2}{5} + \frac{a^4}{7} + \ldots\right] = \frac{n}{\sigma^2}\bar{Y}^2 - 1.$$

5.3. GENERALIZED MLE's (GMLE's)

Generalized maximum likelihood estimators, introduced by Weiss and Wolfowitz (1966), provide (where available) asymptotically efficient estimators, whereas this is not always true for MLE's even if the latter can be found . As noted above, for the case of estimating $\mu_{[1]},\ldots,\mu_{[k]}$, what is meant by "the MLE" is not clear. One possibility, the IMLE, is difficult to compute and may or may not possess desirable properties. Most classical MLE theory assumes i.i.d. observations and is therefore not applicable in our case, since the IMLE is in this case the MLE based on non-i.i.d. observations: the ranked data. The theory of Weiss and Wolfowitz (1966) allows for more general situations, although most of their applications are to i.i.d. "non-regular" cases. (Corrections to Weiss and Wolfowitz (1966) are contained in Weiss and Wolfowitz (1967a), in Weiss and Wolfowitz (1967b), and below. An additional example is given in Weiss and Wolfowitz (1967c).)

We first summarize the results of Weiss and Wolfowitz (1966) for the case k = 2.

(5.3.1) DEFINITION: Let Θ be a closed region in R^2, $\Theta \subseteq \overline{\Theta}$ with $\overline{\Theta}$ a closed region such that every finite boundary point of Θ is an inner point of $\overline{\Theta}$.

(5.3.2) DEFINITION: For each n let X(n) denote the (finite) vector of r.v.'s of which the estimator is to be a function.

(5.3.3) DEFINITION: Let $K_n(x|\theta)$ be the density, with respect to a σ-finite measure μ_n, of $X(n)$ at the point x (of the appropriate space) when θ is the "true" value of the unknown parameter.

(5.3.4) DEFINITION: Let $r = (r_1, r_2)$ be fixed and positive. $\{Z_{n1}(X(n),r), Z_{n2}(X(n),r)\}$ is a sequence of GMLE's if, for each $\theta = (\theta_1,\theta_2) \in \Theta$, (A') and (B') below are satisfied.

CONDITION (A'): There is a sequence of positive constants $\{k_1(n), k_2(n)\}$ such that $k_1(n)\to\infty$, $k_2(n)\to\infty$, and a function $L(y_1,y_2|\theta)$ such that $L(\cdot|\theta)$ is a continuous d.f., and, for any $y = (y_1,y_2)$ and any integers h_1 and h_2

(5.3.5)

$$\lim_{n\to\infty} P_{\theta_1+\frac{h_1r_1}{k_1(n)},\,\theta_2+\frac{h_2r_2}{k_2(n)}}\left[k_1(n)\left(Z_{n1}-\theta_1-\frac{h_1r_1}{k_1(n)}\right) \le y_1,\right.$$
$$\left. k_2(n)\left(Z_{n2}-\theta_2-\frac{h_2r_2}{k_2(n)}\right) \le y_2\right] = L(y_1,y_2|\theta_1,\theta_2).$$

CONDITION (B'): For any integers h_1,h_2 there exists a set $S_n(\theta,h_1,h_2)$ in the space of $X(n)$ such that

(5.3.7) $$\lim P_{\alpha_{ij}}[X(n) \in S_n(\theta,h_1,h_2)] = 1 \qquad (i,j=0,1),$$

where

(5.3.8) $$\alpha_{ij} = \left(\theta_1 + \frac{(h_1+i)r_1}{k_1(n)},\ \theta_2 + \frac{(h_2+j)r_2}{k_2(n)}\right),$$

and there exist sequences

(5.3.9) $$\{a_{nij}(X(n),\theta,h_1,h_2)\} \qquad (i,j = 0,1)$$

of (two-dimensional) r.v.'s such that, as $n\to\infty$, $a_{nij} = (a_{nij1}, a_{nij2})$ converges stochastically to zero when α_{ij} is the parameter of the density of $X(n)$, and such that, whenever $X(n) \in S_n(\theta,h_1,h_2)$, we have the following: Let

(5.3.6) (5.3.10) $$M = \max\{K_n(X(n)\,|\,\alpha_{ij}),\ (i,j = 0,1)\},$$

(5.3.11) $$m = (m_1 m_2) = \left(\theta_1 + \frac{(h_1+1/2)r_1}{k_1(n)},\ \theta_2 + \frac{(h_2+1/2)r_2}{k_2(n)}\right).$$

Then, where "$(a<b,\ c<d)$" means "$(a \leq b,\ c<d)$ or $(a<b,\ c\leq d)$,"

(5.3.12a) $$M = K_n(X(n)\,|\,\alpha_{00}) \Rightarrow \left(Z_{n1} < m_1 + \frac{a_{n001}}{k_1(n)},\ Z_{n2} < m_2 + \frac{a_{n002}}{k_2(n)}\right),$$

(5.3.12b) $$M = K_n(X(n)\,|\,\alpha_{01}) \Rightarrow \left(Z_{n1} < m_1 + \frac{a_{n011}}{k_1(n)},\ Z_{n2} > m_2 + \frac{a_{n012}}{k_2(n)}\right),$$

(5.3.12c) $$M = K_n(X(n)\,|\,\alpha_{10}) \Rightarrow \left(Z_{n1} > m_1 + \frac{a_{n101}}{k_1(n)},\ Z_{n2} < m_2 + \frac{a_{n102}}{k_2(n)}\right),$$

(5.3.12d) $$M = K_n(X(n)\,|\,\alpha_{11}) \Rightarrow \left(Z_{n1} > m_1 + \frac{a_{n111}}{k_1(n)},\ Z_{n2} > m_2 + \frac{a_{n112}}{k_2(n)}\right).$$

THEOREM: (Weiss and Wolfowitz) Let $\{Z_{n1}(X(n),r), Z_{n2}(X(n),r)\}$ be a sequence of GMLE's. Let $\{T_n\}$ be any sequence of estimators of θ such that, for fixed $r = (r_1, r_2) > 0$ and all integers h_1, h_2

$$\lim P_{\theta_1,\theta_2}\left[-\frac{r_1}{2} < k_1(n)(T_{n1}-\theta_1) \leq \frac{r_1}{2},\ -\frac{r_2}{2} < k_2(n)(T_{n2}-\theta_2) \leq \frac{r_2}{2}\right]$$

(5.3.13)
$$= \lim P_{\theta_1+\frac{h_1 r_1}{k_1(n)},\theta_2+\frac{h_2 r_2}{k_2(n)}}\left[-\frac{r_1}{2} < k_1(n)\left(T_{n1}-\theta_1-\frac{h_1 r_1}{k_1(n)}\right) \leq \frac{r_1}{2},\right.$$
$$\left. -\frac{r_2}{2} < k_2(n)\left(T_{n2}-\theta_2-\frac{h_2 r_2}{k_2(n)}\right) \leq \frac{r_2}{2}\right]$$

for any $\theta \varepsilon \Theta$. Then

$$\lim P_\theta\left[-\frac{r_1}{2} < k_1(n)(Z_{n1}-\theta_1) < \frac{r_1}{2},\ -\frac{r_2}{2} < k_2(n)(Z_{n2}-\theta_2) < \frac{r_2}{2}\right]$$
$$\geq \limsup P_\theta\left[-\frac{r_1}{2} < k_1(n)(T_{n1}-\theta_1) \leq \frac{r_1}{2},\right.$$
$$\left. -\frac{r_2}{2} < k_2(n)(T_{n2}-\theta_2) \leq \frac{r_2}{2}\right].$$

Note that on p. 78 of Weiss and Wolfowitz (1966), condition (B') is mis-stated; therein, in (3.13) through (3.16) (corresponding to our (5.3.12a) through (5.3.12d) above)

$$\{a_{n001}, a_{n011}, a_{n101}, a_{n111}; a_{n002}, a_{n012}, a_{n102}, a_{n112}\}$$

should be

$$\left\{\frac{a_{n001}}{k_1(n)}, \frac{a_{n011}}{k_1(n)}, \frac{a_{n101}}{k_1(n)}, \frac{a_{n111}}{k_1(n)}; \frac{a_{n002}}{k_2(n)}, \frac{a_{n012}}{k_2(n)}, \frac{a_{n102}}{k_2(n)}, \frac{a_{n112}}{k_2(n)}\right\}.$$

Examination of the modification of the proof of pp. 73-74 of Weiss and Wolfowitz (1966) used for the proof of their Theorem 3.2 (Theorem (5.3.13) above) shows that without this change the quantities $a_{nij\ell}$ multiplied by the normalizing factors $k_1(n)$ and $k_2(n)$ would occur, and would not necessarily converge stochastically to zero (under the appropriate parameters). In their multi-parameter examples VI, VII, and VIII Weiss and Wolfowitz (1966) seem to satisfy the corrected (B'). (In example VIII this is not as clear as in examples VI and VII.)

We now investigate the application of these results to the estimation of $\mu_{[1]},\ldots,\mu_{[k]}$. For $k \geq 2$ we now choose

$$(5.3.14)\quad \begin{cases} X(n) = (\overline{X}_{[1]},\ldots,\overline{X}_{[k]}) \\ K_n(x|\theta) = K_n(x|\mu) = f^{(\mu)}_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k) \\ \qquad\qquad\qquad\qquad \equiv f_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k) \\ \mu_n = \text{Lebesgue measure on } R^k. \end{cases}$$

We would also like to choose $\Theta = \{\mu: \mu \in \Omega_o, \mu_1=\mu_{[1]},\ldots,\mu_k=\mu_{[k]}\}$, $\overline{\Theta} = R^k$ (which would satisfy (5.3.1)), but by Theorem (B.2.10) this would not allow satisfaction of condition (A') (essentially because $\mu \in \Theta_\cap[\Omega(\neq)]^c$ would not uniquely specify the limiting distribution). Thus, we fix $\eta^* > 0$ and choose

$$(5.3.15)\quad \begin{cases} \Theta(\eta^*) = \{\mu: \mu \in \Theta, \mu_k-\mu_{k-1} \geq \eta^*, \mu_{k-1}-\mu_{k-2} \geq \eta^*, \ldots, \mu_2-\mu_1 \geq \eta^*\} \\ \overline{\Theta} = \Theta(\eta^*/2). \end{cases}$$

(Although our results below would hold if we simply excluded the boundaries of our desired Θ, that set would not be closed.) Since our results lack real dependence on η^*, we have essentially only eliminated the boundary (where equalities exist).

For $k \geq 2$, consider the sequence

(5.3.16) $\{Z_{n1}(X(n),r),\ldots,Z_{nk}(X(n),r)\} = \{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}\}$

with $r = (r_1,\ldots,r_k)$ fixed and positive.

(5.3.17) THEOREM: For $k \geq 2$, condition (A') (or, more properly, its generalization to $k \geq 2$) holds for the sequence (5.3.16) for arbitrary $r > 0$, with $k_1(n) = k_2(n) = \sqrt{n}/\sigma$.

Proof: This follows from Theorem (B.2.8).

(5.3.18) LEMMA: Let h_1 and h_2 be any integers. Choose $S_n(\mu,h_1,h_2)$ $= R^k \cap \{\mu_{[1]}-\varepsilon_n \leq \overline{X}_{[1]} \leq \mu_{[1]}+\varepsilon_n,\ \mu_{[2]}-\varepsilon_n \leq \overline{X}_{[2]} \leq \mu_{[2]}+\varepsilon_n\}$, where $\varepsilon_n = \sigma/n^\delta$ ($0 < \delta < 1/2$ fixed). Then (for $i,j = 0,1$)

$$\lim P_{\alpha_{ij}}[X(n) \in S_n(\mu,h_1,h_2)] = 1.$$

Proof: By (5.3.8), here $\alpha_{ij} = (\mu_{[1]}+(h_1+i)r_1\sigma/\sqrt{n},\mu_{[2]}+(h_2+j)r_2\sigma/\sqrt{n})$, and (setting $a_1 = (h_1+i)r_1$, $a_2 = (h_2+j)r_2$)

$$P_{\alpha_{ij}}[X(n) \in S_n(\mu,h_1,h_2)] = P_{\mu+a\sigma/\sqrt{n}}\left[\mu_{[i]}-\sigma/n^\delta \leq \overline{X}_{[i]} \leq \mu_{[i]}+\sigma/n^\delta \ (i=1,2)\right]$$

(5.3.19)
$$= P_{\mu+a\sigma/\sqrt{n}}\left[-n^{\frac{1}{2}-\delta}-a_1 \leq \frac{\overline{X}_{[1]}-\mu_{[1]}-a_1\sigma/\sqrt{n}}{\sigma/\sqrt{n}} \leq n^{\frac{1}{2}-\delta}-a_1,\right.$$
$$\left.-n^{\frac{1}{2}-\delta}-a_2 \leq \frac{\overline{X}_{[2]}-\mu_{[2]}-a_2\sigma/\sqrt{n}}{\sigma/\sqrt{n}} \leq n^{\frac{1}{2}-\delta}-a_2\right].$$

However, by Theorem (B.2.8) the random quantities of (5.3.19) approach a joint limiting distribution, while the respective upper and lower limits on those quantities tend to $\pm\infty$. (In fact, the result is proven for any fixed $a = (a_1, a_2)$ and not just for $((h_1+i)r_1, (h_2+j)r_2)$.)

As noted in the proof of Lemma (5.3.18), for our case we have (for $i,j = 0,1$)

(5.3.20) $$\alpha_{ij} = (\mu_{[1]} + (h_1+i)r_1\sigma/\sqrt{n},\ \mu_{[2]} + (h_2+j)r_2\sigma/\sqrt{n}).$$

<u>LEMMA</u>: If $k = 2$, then (for $i,j = 0,1$)

(5.3.21)
$$K_n(x|\alpha_{ij})2\pi\,\frac{\sigma^2}{n}\,e^{\frac{r_1^2h_1^2}{2}+\frac{r_2^2h_2^2}{2}}$$
$$= a'e^{r_1 i\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}} - i(h_1+\frac{1}{2}i)r_1^2 + r_2 j\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}} - j(h_2+\frac{1}{2}j)r_2^2}$$
$$+ b'e^{r_2 j\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}} - j(h_2+\frac{1}{2}j)r_2^2 + r_1 i\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}} - i(h_1+\frac{1}{2}i)r_1^2},$$

where

$$\begin{cases} a' = e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2} e^{r_1h_1\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}} e^{r_2h_2\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}} \\ b' = e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2} e^{r_2h_2\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}} e^{r_1h_1\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}} \end{cases}.$$

<u>Proof</u>: (Note that $a' > 0$ and $b' > 0$ involve only σ, n, x_1, x_2, $\mu_{[1]}$, $\mu_{[2]}$, r_1, r_2, h_1, and h_2, and not i and j.) From (5.3.14), (5.3.20),

and (B.1.1),

$$K_n(x|\alpha_{ij})2\pi\ \sigma^2/n = e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[1]}-(h_1+i)r_1\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[2]}-(h_2+j)r_2\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right)^2}$$

$$+e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[2]}-(h_2+j)r_2\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[1]}-(h_1+i)r_1\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right)^2}$$

$$= e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2}\,\circ$$

$$\circ e^{-\frac{1}{2}\left[\frac{-2(x_1-\mu_{[1]})(h_1+i)r_1\sigma/\sqrt{n}}{\sigma^2/n}+(h_1+i)^2r_1^2\right]}\,\circ$$

$$\circ e^{-\frac{1}{2}\left[\frac{-2(x_2-\mu_{[2]})(h_2+j)r_2\sigma/\sqrt{n}}{\sigma^2/n}+(h_2+j)^2r_2^2\right]}$$

$$+e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2}\,\circ$$

$$\circ e^{-\frac{1}{2}\left[\frac{-2(x_1-\mu_{[2]})(h_2+j)r_2\sigma/\sqrt{n}}{\sigma^2/n}+(h_2+j)^2r_2^2\right]}\,\circ$$

$$\circ e^{-\frac{1}{2}\left[\frac{-2(x_2-\mu_{[1]})(h_1+i)r_1\sigma/\sqrt{n}}{\sigma^2/n}+(h_1+i)^2r_1^2\right]}$$

$$= e^{-\frac{1}{2}\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2}\,\circ$$

$$\cdot e^{\frac{x_1-\mu[1]}{\sigma/\sqrt{n}}(h_1+i)r_1-(h_1+i)^2\frac{r_1^2}{2}+\frac{x_2-\mu[2]}{\sigma/\sqrt{n}}(h_2+j)r_2-(h_2+j)^2\frac{r_2^2}{2}}$$

$$+\ e^{-\frac{1}{2}\left(\frac{x_1-\mu[2]}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu[1]}{\sigma/\sqrt{n}}\right)^2}\cdot$$

$$\cdot e^{\frac{x_1-\mu[2]}{\sigma/\sqrt{n}}(h_2+j)r_2-(h_2+j)^2\frac{r_2^2}{2}+\frac{x_2-\mu[1]}{\sigma/\sqrt{n}}(h_1+i)r_1-(h_1+i)^2\frac{r_1^2}{2}}$$

$$=\ e^{-\frac{r_1^2h_1^2}{2}-\frac{r_2^2h_2^2}{2}}\left\{e^{-\frac{1}{2}\left(\frac{x_1-\mu[1]}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu[2]}{\sigma/\sqrt{n}}\right)^2}\cdot\right.$$

$$\cdot e^{r_1(h_1+i)\frac{x_1-\mu[1]}{\sigma/\sqrt{n}}-i^2\frac{r_1^2}{2}-ih_1r_1^2}\, e^{r_2(h_2+j)\frac{x_2-\mu[2]}{\sigma/\sqrt{n}}-j^2\frac{r_2^2}{2}-jh_2r_2^2}$$

$$+\ e^{-\frac{1}{2}\left(\frac{x_1-\mu[2]}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu[1]}{\sigma/\sqrt{n}}\right)^2}\, e^{r_2(h_2+j)\frac{x_1-\mu[2]}{\sigma/\sqrt{n}}-j^2\frac{r_2^2}{2}-jh_2r_2^2}\cdot$$

$$\left.\cdot e^{r_1(h_1+i)\frac{x_2-\mu[1]}{\sigma/\sqrt{n}}-i^2\frac{r_1^2}{2}-ih_1r_1^2}\right\}$$

$$=\ e^{-\frac{r_1^2h_1^2}{2}-\frac{r_2^2h_2^2}{2}}\cdot$$

$$\cdot\left\{a'e^{r_1i\frac{x_1-\mu[1]}{\sigma/\sqrt{n}}-i^2\frac{r_1^2}{2}-ih_1r_1^2}\, e^{r_2j\frac{x_2-\mu[2]}{\sigma/\sqrt{n}}-j^2\frac{r_2^2}{2}-jh_2r_2^2}\right.$$

$$\left.+\ b'e^{r_2j\frac{x_1-\mu[2]}{\sigma/\sqrt{n}}-j^2\frac{r_2^2}{2}-jh_2r_2^2}\, e^{r_1i\frac{x_2-\mu[1]}{\sigma/\sqrt{n}}-i^2\frac{r_1^2}{2}-ih_1r_1^2}\right\}.$$

LEMMA: There exist a_{nij1} and a_{nij2} (which may depend on $X(n)$, μ, h_1, and h_2) which converge stochastically to zero when α_{ij} is the parameter of the density of $X(n)$ $(i,j=0,1)$ such that, if $X(n) \varepsilon\ S_n(\mu, h_1, h_2)$ and $M = K_n(X(n)|\alpha_{ij})$, then

(5.3.22)

(i) for i,j = 0,0

$$(5.3.23)\quad \begin{cases} \dfrac{\overline{X}_{[1]} - \mu_{[1]}}{\sigma/\sqrt{n}} < (h_1 + \frac{1}{2})r_1 + a_{n001} \\ \text{and} \\ \dfrac{\overline{X}_{[2]} - \mu_{[2]}}{\sigma/\sqrt{n}} < (h_2 + \frac{1}{2})r_2 + a_{n002} \end{cases}$$

(ii) for i,j = 0,1

$$(5.3.24)\quad \begin{cases} \dfrac{\overline{X}_{[1]} - \mu_{[1]}}{\sigma/\sqrt{n}} < (h_1 + \frac{1}{2})r_1 + a_{n011} \\ \text{and} \\ \dfrac{\overline{X}_{[2]} - \mu_{[2]}}{\sigma/\sqrt{n}} > (h_2 + \frac{1}{2})r_2 + a_{n012} \end{cases}$$

(iii) for i,j = 1,0

$$(5.3.25)\quad \begin{cases} \dfrac{\overline{X}_{[1]} - \mu_{[1]}}{\sigma/\sqrt{n}} > (h_1 + \frac{1}{2})r_1 + a_{n101} \\ \text{and} \\ \dfrac{\overline{X}_{[2]} - \mu_{[2]}}{\sigma/\sqrt{n}} < (h_2 + \frac{1}{2})r_2 + a_{n102} \end{cases}$$

(iv) <u>for i,j = 1,1</u>

$$(5.3.26)\quad \begin{cases} \dfrac{\overline{X}_{[1]}-\mu_{[1]}}{\sigma/\sqrt{n}} > (h_1+\frac{1}{2})r_1 + a_{n111} \\ \text{and} \\ \dfrac{\overline{X}_{[2]}-\mu_{[2]}}{\sigma/\sqrt{n}} > (h_2+\frac{1}{2})r_2 + a_{n112}. \end{cases}$$

<u>Proof</u>: <u>(i) Case i,j = 0,0</u>. For simplicity, write x for X(n), x_1 for $\overline{X}_{[1]}$, x_2 for $\overline{X}_{[2]}$, μ_1 for $\mu_{[1]}$, and μ_2 for $\mu_{[2]}$. Since $K_n(x|\alpha_{00}) \geq K_n(x|\alpha_{10})$, by Lemma (5.3.21),

$$(5.3.27)\quad \begin{aligned} a' + b' &\geq a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} + b'e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} \\ &\geq a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} + b'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} \end{aligned}$$

since $x_1 \leq x_2$. Thus, $1 \geq e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}$ and (taking logarithms)

$0 \geq r_1\left[\dfrac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1\right]$ so that (since $r_1 > 0$)

$$(5.3.28)\quad \frac{x_1-\mu_1}{\sigma/\sqrt{n}} \leq (h_1+\frac{1}{2})r_1.$$

We may (for example) take $a_{n001} = \dfrac{1}{n}$ and thereby satisfy the first part of (5.3.23).

Since $K_n(x|\alpha_{00}) \geq K_n(x|\alpha_{01})$, by Lemma (5.3.21),

(5.3.29)
$$a' + b' \geq a' e^{r_2 \frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2} + b' e^{r_2 \frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2}$$
$$\geq a' e^{r_2 \frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2}$$

or

(5.3.30)
$$1 + \frac{b'}{a'} \geq e^{r_2 \left[\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2\right]}.$$

Now, by the definitions of a' and b',

(5.3.31)
$$0 < \frac{b'}{a'} = \frac{e^{-\frac{1}{2}\left(\frac{x_1-\mu_2}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_1}{\sigma/\sqrt{n}}\right)^2}}{e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma/\sqrt{n}}\right)^2}} \cdot \frac{e^{r_2 h_2 \frac{x_1-\mu_2}{\sigma/\sqrt{n}}} e^{r_1 h_1 \frac{x_2-\mu_1}{\sigma/\sqrt{n}}}}{e^{r_1 h_1 \frac{x_1-\mu_1}{\sigma/\sqrt{n}}} e^{r_2 h_2 \frac{x_2-\mu_2}{\sigma/\sqrt{n}}}}$$
$$= \frac{e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1-\mu_2}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma/\sqrt{n}} + \frac{\mu_2-\mu_1}{\sigma/\sqrt{n}}\right)^2}}{e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma/\sqrt{n}}\right)^2}} \cdot e^{(r_1 h_1 - r_2 h_2)\frac{x_2-x_1}{\sigma/\sqrt{n}}}$$
$$= e^{-\frac{n}{\sigma^2}(\mu_2-\mu_1)(x_2-x_1)} e^{(r_1 h_1 - r_2 h_2)\frac{\sqrt{n}}{\sigma}(x_2-x_1)}$$
$$= \begin{cases} e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1) - \frac{\sigma}{\sqrt{n}}(r_1 h_1 - r_2 h_2)\}} & \text{if } \mu_2 > \mu_1 \\ e^{\frac{x_2-x_1}{\sigma/\sqrt{n}}(r_1 h_1 - r_2 h_2)} & \text{if } \mu_2 = \mu_1. \end{cases}$$

Since $\mu_2 > \mu_1$, from (5.3.30) and (5.3.31) we find (taking logarithms and simplifying) that

$$\frac{x_2-\mu_2}{\sigma/\sqrt{n}} \leq (h_2+\frac{1}{2})r_2 + \frac{1}{r_2}\ln(1+\frac{b'}{a'}). \tag{5.3.32}$$

We now wish to show that the choice $a_{n002} = \frac{1}{r_2}\ln(1+\frac{b'}{a'})$ is effective. (Here we use the fact that a_{n002} may depend on μ, as well as on $X(n)$, h_1, and h_2.) Since

$$\begin{aligned}|(\overline{X}_{[2]}-\overline{X}_{[1]})-(\mu_{[2]}-\mu_{[1]})| &= |(\overline{X}_{[2]}-\mu_{[2]})-(\overline{X}_{[1]}-\mu_{[1]})| \\ &\leq |\overline{X}_{[2]}-\mu_{[2]}| + |\overline{X}_{[1]}-\mu_{[1]}|,\end{aligned}$$

for any $\varepsilon > 0$, $\{|\overline{X}_{[2]}-\mu_{[2]}| < \frac{\varepsilon}{2}, |\overline{X}_{[1]}-\mu_{[1]}| < \frac{\varepsilon}{2}\} \Rightarrow$ $\{|(\overline{X}_{[2]}-\overline{X}_{[1]})-(\mu_{[2]}-\mu_{[1]})| < \varepsilon\}$, so that

$$\begin{aligned}&P_{\alpha_{00}}[|(\overline{X}_{[2]}-\overline{X}_{[1]})-(\mu_{[2]}-\mu_{[1]})| < \varepsilon] \\ &\geq P_{\alpha_{00}}[|\overline{X}_{[2]}-\mu_{[2]}| < \varepsilon/2, |\overline{X}_{[1]}-\mu_{[1]}| < \varepsilon/2].\end{aligned} \tag{5.3.33}$$

By Theorem (B.2.8), as $n\to\infty$

$$\begin{aligned}&P_{\alpha_{00}}[|\overline{X}_{[1]}-\mu_{[1]}| < \varepsilon/2] = P_{\alpha_{00}}[-\varepsilon/2 < \overline{X}_{[1]}-\mu_{[1]} < \varepsilon/2] \\ &= P_{\alpha_{00}}[-h_1r_1-\frac{\varepsilon}{2}\frac{\sqrt{n}}{\sigma} < \frac{\sqrt{n}}{\sigma}(\overline{X}_{[1]}-\mu_{[1]}-h_1r_1\sigma/\sqrt{n}) < \frac{\varepsilon}{2}\frac{\sqrt{n}}{\sigma} - h_1r_1]\to 1,\end{aligned} \tag{5.3.34}$$

a similar result holding for $\overline{X}_{[2]}$. By Lemma (B.2.1), the r.h.s. of (5.3.33) $\to 1$ as $n\to\infty$, so that the ℓ.h.s. must also $\to 1$ as $n\to\infty$. Taking $\varepsilon = \varepsilon'(\mu_{[2]}-\mu_{[1]})$ with $0 < \varepsilon' < 1$, this means that as $(n\to\infty)$

$$P_{\alpha_{00}}[(1-\varepsilon')(\mu_{[2]}-\mu_{[1]}) < \overline{X}_{[2]}-\overline{X}_{[1]} < (1+\varepsilon')(\mu_{[2]}-\mu_{[1]})]\to 1. \tag{5.3.35}$$

Using (5.3.35), noting that $x_2-x_1 > 0$, and taking $n \geq (r_1h_1-r_2h_2)^2\sigma^2 \cdot \delta/(\mu_2-\mu_1)^2$, it follows that the exponent a_n (say) of $b'/a' = e^{-a_n}$ in (5.3.31) is such that for all finite x we have $P_{\alpha_{00}}[a_n \leq x] \to 0$ as $n\to\infty$. Then it can be shown (successively) that

(5.3.36) $$P_{\alpha_{00}}\left[e^{-a_n} \leq x\right] \to \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases};$$

(5.3.37) $$P_{\alpha_{00}}\left[\ln\left(1+e^{-a_n}\right) \leq x\right] \to \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

From (5.3.37) it follows that our a_{n002} converges stochastically to zero under α_{00}.

<u>(ii)</u> <u>Case i,j = 0,1.</u> Since $K_n(x|\alpha_{00}) \leq K_n(x|\alpha_{01})$, by Lemma (5.3.21),

(5.3.38)
$$\begin{aligned} a' + b' &\leq a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} + b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} \\ &\leq a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} + b'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} \end{aligned}$$

since $x_1 \leq x_2$. Thus $e^{r_2\left[\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2\right]} \geq 1$ and (taking logarithms) $r_2\left[\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2\right] \geq 0$ so that (since $r_2 > 0$)

(5.3.39) $$\frac{x_2-\mu_2}{\sigma/\sqrt{n}} \geq (h_2+\frac{1}{2})r_2.$$

We may (for example) take $a_{n012} = -\frac{1}{n}$ and thereby satisfy the second part of (5.3.24).

Since $K_n(x|\alpha_{01}) \geq K_n(x|\alpha_{11})$, by Lemma (5.3.21) we have

$$a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} + b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$\geq \quad a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2 + r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$+ b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2 + r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}.$$

This can be reduced as follows:

$$a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} + b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}\left[1-e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\right]$$

$$\geq a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\, e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2};$$

$$1 + \frac{b'}{a'}e^{r_2\frac{x_1-x_2}{\sigma/\sqrt{n}}}\left[1-e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\right] \geq e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2};$$

$$\frac{x_1-\mu_1}{\sigma/\sqrt{n}} \leq (h_1+\frac{1}{2})r_1 + \frac{1}{r_1}\ell n\left\{1+ \frac{b'}{a'}e^{r_2\frac{x_1-x_2}{\sigma/\sqrt{n}}}\left[1-e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\right]\right\}.$$

Since $\mu_2 > \mu_1$, use of (5.3.31) reduces this inequality to

$$\frac{x_1-\mu_1}{\sigma/\sqrt{n}} \leq$$

$$\le (h_1+\tfrac{1}{2})r_1 + \frac{1}{r_1}\ln\left\{1+e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1)-\frac{\sigma}{\sqrt{n}}(r_1h_1-r_2h_2)\}}\right.\cdot$$

(5.3.40)
$$\left.\cdot e^{r_2\frac{\sqrt{n}}{\sigma}(x_1-x_2)}\left[1-e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}\right]\right\}$$

$$= (h_1+\tfrac{1}{2})r_1 + \frac{1}{r_1}\ln\left\{1+e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1)-\frac{\sigma}{\sqrt{n}}(r_1h_1-r_2(h_2+1))\}}\right.\cdot$$

$$\left.\cdot\left[1-e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}\right]\right\}.$$

In order to show that the choice of a_{n011} as the second term on the r.h.s. of (5.3.40) is effective, we will show that

(5.3.41)
$$P_{\alpha_{01}}[a_{n011}\le x]\to\begin{cases}1, & x>0\\ 0, & x\le 0.\end{cases}$$

This implies that a_{n011} converges stochastically to zero under α_{01}. (To show the inequality of (5.3.24), a_{n011} should actually be taken as (e.g.) the above plus $\frac{1}{n}$.) Now, if α_{00} is replaced by α_{01} and h_2+1 replaces h_2, then the same proof that yielded (5.3.35) yields (with $0<\varepsilon'<1$)

(5.3.42)
$$P_{\alpha_{01}}[(1-\varepsilon')(\mu_{[2]}-\mu_{[1]}) < \overline{X}_{[2]}-\overline{X}_{[1]} < (1+\varepsilon')(\mu_{[2]}-\mu_{[1]})]\to 1$$

as $n\to\infty$. Using (5.3.42), noting that $x_2-x_1>0$, and taking $n\ge(r_1h_1-r_2(h_2+1))^2\sigma^2\delta/(\mu_2-\mu_1)^2$ with $\delta>1$, the exponent of

(5.3.43) $$A_n \equiv e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1)-\frac{\sigma}{\sqrt{n}}(r_1h_1-r_2(h_2+1))\}}$$

is such that (as $n\to\infty$)

(5.3.44) $$P_{\alpha_{01}}[A_n \le x] \to \begin{cases} 1, & x > 0 \\ 0, & x \le 0 . \end{cases}$$

In

$$B_n \equiv -e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2} A_n = -e^{-(h_1+\frac{1}{2})r_1^2} e^{r_1\frac{\sqrt{n}}{\sigma}(x_2-x_1+x_1-\mu_1)} A_n$$

(5.3.45) $$= -e^{-(h_1+\frac{1}{2})r_1^2} e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1)-\frac{\sigma}{\sqrt{n}}(r_1(h_1+1)-r_2(h_2+1))\}} \cdot e^{r_1\frac{\sqrt{n}}{\sigma}(x_1-\mu_1)},$$

the middle exponential term tends stochastically to zero (under α_{01} as $n\to\infty$) as did A_n, since it is A_n with h_1 replaced by h_1+1, and the first exponential term is a constant. By Theorem (B.2.8), $\exp\{r_1\sqrt{n}(\overline{X}_{[1]}-\mu_{[1]})/\sigma\}$ has a non-degenerate limiting distribution since (for any $x > 0$)

(5.3.46) $$P_{\alpha_{01}}\left[e^{r_1\frac{\sqrt{n}}{\sigma}(\overline{X}_{[1]}-\mu_{[1]})} \le x\right] = P_{\alpha_{01}}\left[r_1\frac{\sqrt{n}}{\sigma}(\overline{X}_{[1]}-\mu_{[1]}) \le \ell n\ x\right]$$
$$= P_{\alpha_{01}}\left[\frac{\sqrt{n}}{\sigma}(\overline{X}_{[1]}-\mu_{[1]}-h_1r_1\frac{\sigma}{\sqrt{n}}) \le \ell n\ x^{1/r_1} - h_1r_1\right].$$

It then follows that, as $n\to\infty$,

(5.3.47) $$P_{\alpha_{01}}[B_n \le x] \to \begin{cases} 1, & x > 0 \\ 0, & x < 0 . \end{cases}$$

By (5.3.44) and (5.3.47), $1 + A_n + B_n$ converges stochastically to 1 under α_{01} as $n\to\infty$, and since $a_{n011} = \frac{1}{r_1}\ln\{1+A_n+B_n\}$ it follows that a_{n011} converges stochastically to zero under α_{01}.

<u>(iii)</u> <u>Case i,j = 1,0</u>. Since $K_n(x|\alpha_{10}) \ge K_n(x|\alpha_{11})$, by Lemma (5.3.21),

$$a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} + b'e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}$$

$$\ge a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2 + r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$+ b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2 + r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}$$

$$\ge a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\, e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} ;$$

$$1 + \frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}} \ge e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} ;$$

$$\frac{x_2-\mu_2}{\sigma/\sqrt{n}} \le (h_2+\frac{1}{2})r_2 + \frac{1}{r_2}\ln\left\{1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right\}.$$

We will now show that the choice $a_{n102} = \frac{1}{r_2}\ln\left\{1+ \frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right\}$

is effective. Since $\mu_2 > \mu_1$, by (5.3.31)

$$\text{(5.3.48)} \qquad \frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}} = e^{-\frac{n}{\sigma^2}(x_2-x_1)\{(\mu_2-\mu_1)-\frac{\sigma}{\sqrt{n}}(r_1(h_1+1)-r_2h_2)\}}$$

and the argument of (5.3.42) through (5.3.44) can be modified to show that this converges stochastically to zero under α_{10} as $n\to\infty$. The result for a_{n102} then follows.

Since $K_n(x|\alpha_{10}) \geq K_n(x|\alpha_{00})$, by Lemma (5.3.21)

$$a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2} + b'e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2} \geq a' + b';$$

$$e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}\left\{a'+b'e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right\} \geq a' + b';$$

$$\frac{x_1-\mu_1}{\sigma/\sqrt{n}} \geq (h_1+\frac{1}{2})r_1 + \frac{1}{r_1}\ln(1+\frac{b'}{a'}) - \frac{1}{r_1}\ln\left(1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right).$$

The efficacy of $a_{n101} = \frac{1}{r_1}\ln(1+\frac{b'}{a'}) - \frac{1}{r_1}\ln\left(1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right)$

is shown by a modification (allowing for α_{10}) of the proof for a_{n102} above.

(iv) Case i,j = 1,1. Since $K_n(x|\alpha_{01}) \leq K_n(x|\alpha_{11})$, by Lemma (5.3.21) and the fact that $x_1 \leq x_2$

$$a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$\leq a'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2} + b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

(5.3.49)
$$\leq a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2 + r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$+ b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2 + r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}$$

$$\leq a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2 + r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2}$$

$$+ b'e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}} - (h_2+\frac{1}{2})r_2^2 + r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2},$$

so that (utilizing the first and last lines above)

(5.3.50)
$$1 \leq e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2} + \frac{b'}{a'}e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}$$

$$= e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}} - (h_1+\frac{1}{2})r_1^2}\left\{1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right\};$$

(5.3.51)
$$\frac{x_1-\mu_1}{\sigma/\sqrt{n}} \geq (h_1+\frac{1}{2})r_1 - \frac{1}{r_1}\ln\left(1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right).$$

The efficacy of $a_{n111} = -\frac{1}{r_1}\ln\left(1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right)$ is shown by a modification (allowing for α_{11}) of the proof for a_{n102}.

Since $K_n(x|\alpha_{10}) \le K_n(x|\alpha_{11})$, we obtain (as with (5.3.49))

$$a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}$$

$$\le a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2} + b'e^{r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}$$

$$\le a'e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2+r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}}-(h_2+\frac{1}{2})r_2^2}$$

$$+ b'e^{r_2\frac{x_1-\mu_2}{\sigma/\sqrt{n}}-(h_2+\frac{1}{2})r_2^2+r_1\frac{x_2-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}$$

$$\le e^{r_1\frac{x_1-\mu_1}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}\, e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}}-(h_2+\frac{1}{2})r_2^2}\left(a'+b'e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right),$$

and (as with (5.3.50))

$$1 \le e^{r_2\frac{x_2-\mu_2}{\sigma/\sqrt{n}}-(h_2+\frac{1}{2})r_2^2}\left(1+\frac{b'}{a'}e^{r_1\frac{x_2-x_1}{\sigma/\sqrt{n}}}\right).$$

The rest of the proof is similar to that of the first part of case (iv) after (5.3.51).

(5.3.52) THEOREM: For $k \ge 2$, condition (B') (or, more properly, its generalization to $k \ge 2$) holds for the sequence (5.3.16) for arbitrary $r > 0$.

Proof: Condition (B') is given at (5.3.6). Its first requirement, (5.3.7), is satisfied by (the generalization to $k \geq 2$ of) Lemma (5.3.18). The remainder of its requirements are satisfied (for the case $k = 2$) by Lemma (5.3.22). We will now show that these remaining requirements are satisfied when $k > 2$.

As at (5.3.21) and (5.3.20), for $i_1,\ldots,i_k = 0,1$

$$K_n(x|\mu) = f^{(\mu)}_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)$$

(5.3.53)
$$= \sum_{\beta\varepsilon S_k} (\sqrt{n}/\sigma)^k \phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right);$$

(5.3.54)
$$\alpha_{i_1\ldots i_k} = (\mu_{[1]}+(h_1+i_1)r_1\sigma/\sqrt{n},\ldots,\mu_{[k]}+(h_k+i_k)r_k\sigma/\sqrt{n}).$$

Thus,

$$K_n\left(x\middle|\alpha_{i_1 i_2\ldots i_k}\right)(\sqrt{2\pi}\sigma/\sqrt{n})^k e^{\frac{r_1^2h_1^2}{2}+\cdots+\frac{r_k^2h_k^2}{2}}$$

$$= (\sqrt{2\pi})^k e^{\frac{r_1^2h_1^2}{2}+\cdots+\frac{r_k^2h_k^2}{2}}(1/\sqrt{2\pi})^k \cdot$$

$$\cdot \sum_{\beta\varepsilon S_k} e^{-\frac{1}{2}\sum_{j=1}^{k}\left(\frac{x_{\beta(j)}-\mu_{[j]}-(h_j+i_j)r_j\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right)^2}$$

$$= e^{\frac{r_1^2h_1^2}{2}+\cdots+\frac{r_k^2h_k^2}{2}}\sum_{\beta\varepsilon S_k}$$

$$e^{-\frac{1}{2}\sum_{j=1}^{k}\left\{\left(\frac{x_{\beta(j)}-\mu_{[j]}}{\sigma/\sqrt{n}}\right)^2 -2(x_{\beta(j)}-\mu_{[j]})(h_j+i_j)r_j\frac{\sqrt{n}}{\sigma}+(h_j+i_j)^2r_j^2\right\}}$$

(5.3.55)

$$= e^{\frac{r_1^2h_1^2}{2}+\dots+\frac{r_k^2h_k^2}{2}} \sum_{\beta\varepsilon S_k} e^{\sum_{j=1}^{k}\left\{-\frac{1}{2}\left(\frac{x_{\beta(j)}-\mu_{[j]}}{\sigma/\sqrt{n}}\right)^2 + h_jr_j\frac{(x_{\beta(j)}-\mu_{[j]})}{\sigma/\sqrt{n}}\right\}}\circ$$

$$\circ e^{\sum_{j=1}^{k}\left\{r_ji_j\frac{(x_{\beta(j)}-\mu_{[j]})}{\sigma/\sqrt{n}} - i_j(h_j+\frac{1}{2}i_j)r_j^2 - \frac{r_j^2h_j^2}{2}\right\}}$$

$$= \sum_{\beta\varepsilon S_k} a'(\beta)e^{\sum_{j=1}^{k}\left\{r_ji_j\frac{(x_{\beta(j)}-\mu_{[j]})}{\sigma/\sqrt{n}} - i_j(h_j+\frac{1}{2}i_j)r_j^2\right\}},$$

where

$$a'(\beta) = e^{\sum_{j=1}^{k}\left\{-\frac{1}{2}\left(\frac{x_{\beta(j)}-\mu_{[j]}}{\sigma/\sqrt{n}}\right)^2 + h_jr_j\frac{(x_{\beta(j)}-\mu_{[j]})}{\sigma/\sqrt{n}}\right\}}. \qquad (5.3.56)$$

While for the case $k = 2$ there were $2! = 2$ terms in the final summation, there are now $k!$ terms.

As there were $2^2 = 4$ parts to Lemma (5.3.22), there are 2^k parts here. We will give the proof for the part corresponding to (5.3.23), since it is indicative. I.e., in the case $i_1,\dots,i_k = 0,\dots,0$,

$$(5.3.57) \qquad \begin{cases} \dfrac{\overline{X}_{[1]}-\mu_{[1]}}{\sigma/\sqrt{n}} < (h_1+\frac{1}{2})r_1 + a_{n0\dots01} \\[2ex] \dfrac{\overline{X}_{[2]}-\mu_{[2]}}{\sigma/\sqrt{n}} < (h_2+\frac{1}{2})r_2 + a_{n0\dots02} \\ \vdots \end{cases}$$

$$\left| \frac{\overline{X}_{[k]}-\mu_{[k]}}{\sigma/\sqrt{n}} < (h_k+\tfrac{1}{2})r_k + a_{n0\ldots 0k} \right.$$

(where $a_{ni_1\cdots i_k 1},\ldots, a_{ni_1\cdots i_k k}$ converge stochastically to zero when $\alpha_{i_1\cdots i_k}$ is the parameter of the density of $X(n)$ $(i_1,\ldots,i_k = 0,1))$ when $X(n) \varepsilon S_n(\mu, h_1,\ldots,h_k)$ and $M = K_n(X(n)|\alpha_{0\ldots 0})$. The $a_{ni_1\ldots i_k j}$ $(j = 1,\ldots,k)$ may depend on $X(n),\mu,h_1,\ldots,h_k$.

For, e.g., the first comparison of (5.3.57), $K_n(x|\alpha_{00\ldots 0}) \geq K_n(x|\alpha_{10\ldots 0})$, so by (5.3.55) and the fact that $x_1 \leq x_i$ $(i = 2,\ldots,k)$,

$$\sum_{\beta\varepsilon S_k} a'(\beta) \geq \sum_{\beta\varepsilon S_k} a'(\beta) e^{r_1\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)-(h_1+\frac{1}{2})r_1^2}$$

$$\geq \sum_{\beta\varepsilon S_k} a'(\beta) e^{r_1\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)-(h_1+\frac{1}{2})r_1^2};$$

$$1 \geq e^{r_1\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}-(h_1+\frac{1}{2})r_1^2}.$$

From here the proof is essentially that which follows (5.3.27).

<u>Rule for making comparisons</u>. For each of the $k!$ vectors $i_1,\ldots,i_k$, one must prove k relations similar to (5.3.57), with appropriate modifications of "<" to ">". For these, compare the given $\alpha_{i_1\ldots i_k}$ with the k others which have $i_1',\ldots,i_k'$'s which differ from the given $i_1,\ldots,i_k$ in only one place. (This rule, suggested by the $k = 2$ results, works when $k > 2$.)

To illustrate our method, we will now study, e.g., the second comparison of (5.3.57). Since $K_n(x|\alpha_{000\ldots0}) \geq K_n(x|\alpha_{010\ldots0})$,

$$\sum_{\beta\varepsilon S_k} a'(\beta) \geq \sum_{\beta\varepsilon S_k} a'(\beta) e^{r_2 \frac{x_{\beta(2)}-\mu_{[2]}}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2}$$

$$\geq \sum_{\substack{\beta\varepsilon S_k \\ \beta(2)=2}} a'(\beta) e^{r_2 \frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2} ;$$

$$1 + \frac{\sum_{\beta\varepsilon S_k} a'(\beta) - \sum_{\substack{\beta\varepsilon S_k \\ \beta(2)=2}} a'(\beta)}{\sum_{\substack{\beta\varepsilon S_k \\ \beta(2)=2}} a'(\beta)} \geq e^{r_2 \frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}} - (h_2 + \frac{1}{2}) r_2^2} .$$

Now the proof proceeds as at (5.3.30), and a relation like (5.3.31) holds because what is left in $\sum_{\beta\varepsilon S_k} a'(\beta)$ after $\sum_{\substack{\beta\varepsilon S_k \\ \beta(2)=2}} a'(\beta)$ is removed, makes the "wrong" associations and thus tends to zero, while the denominator does not.

<u>THEOREM</u>: For $\Theta(\eta^*)$ and any fixed $r = (r_1,\ldots,r_k) > 0$, $(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$ is a sequence of GMLE's for estimation of

(5.3.58) $(\mu_{[1]},\ldots,\mu_{[k]})$ based on $X(n) = (\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$. It thus possesses, for all $r = (r_1,\ldots,r_k) > 0$, the property of Theorem (5.3.13).

<u>Proof</u>: Theorems (5.3.17) and (5.3.52) establish conditions (A') and (B'), respectively, for all $r > 0$. We therefore have a sequence of GMLE's possessing the property of Theorem (5.3.13), or more properly its extension to $k \geq 2$, for all $r > 0$.

If T and U are estimators of θ, then U is said to be <u>more concentrated</u> (about θ) than T if

(5.3.59) $$P_\theta[-r \leq U-\theta \leq r] \geq P_\theta[-r \leq T-\theta \leq r]$$

for all $\theta \varepsilon \Theta$ and all $r > 0$. (This definition, which appears for perhaps the first time in print in Lawton (1968), is known to the present author to have been stated by Professor Lionel Weiss as early as March 1965 in lectures at Cornell University.) If T_n and U_n estimate θ, then U_n is said to be of higher <u>large sample concentration</u> (about θ) than T_n if

(5.3.60) $$\lim_{n\to\infty} P_\theta[-r \leq k(n)(U_n-\theta) \leq r] \geq \lim_{n\to\infty} P_\theta[-r \leq k(n)(T_n-\theta) \leq r],$$

where $k(n)$ is such that $k(n)(U_n-\theta)$ approaches a limiting distribution, for all $\theta \varepsilon \Theta$ and all $r > 0$. The GMLE $(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$ has, using a k-dimensional generalization of (5.3.60), desirable large sample concentration in comparison to the class of estimators of Theorem (5.3.13).

We will now show (for $k = 2$, the $k > 2$ extension being similar) that, by finding one GMLE, we find <u>a class of GMLE's</u>.

<u>LEMMA</u>: Suppose $\lim_{n\to\infty} P_{\theta_n}[Z_n < y] = L(y)$, with $L(\cdot)$ a continuous

(5.3.61) d.f.. Then, if $\lim_{n\to\infty} c_n = 0$,

$$\lim_{n\to\infty} P_{\theta_n}[Z_n < y+c_n] = L(y).$$

Proof: If all but a finite number of the c_n are positive, then $L(y) \leq \lim_{n\to\infty} P_{\theta_n}[Z_n < y+c_n]$ and (since eventually all c_n are less than c_m, m fixed)

$$(5.3.62) \qquad \lim_{n\to\infty} P_{\theta_n}[Z_n < y+c_n] \leq L(y+c_m).$$

Taking the limit on m in (5.3.62) and using the continuity of $L(\cdot)$ the desired result follows. (If all but a finite number of the c_n are negative, the proof is similar.)

If infinitely many c_n are positive and infinitely many c_n are negative, suppose $c_r < 0, c_s > 0$. Then

$$(5.3.63) \qquad L(y+c_r) \leq \lim_{n\to\infty} P_{\theta_n}[Z_n < y+c_n] \leq L(y+c_s)$$

since eventually $c_r \leq c_n \leq c_s$. Taking limits in (5.3.63) over $\{r: c_r < 0\}$ and $\{s: c_s > 0\}$ on the ℓ.h.s. and r.h.s. (respectively) the desired result follows. Note that this is a special case of, with an even simpler proof than, Cramér's Theorem (see, e.g., Fisz (1963), p.236).

THEOREM: If $\{Z_{n1}(X(n),r), Z_{n2}(X(n),r)\}$ is a sequence of GMLE's then so is

$$(5.3.65) \qquad \{Z_{n1} + o_1(1/k_1(n)),\ Z_{n2} + o_2(1/k_2(n))\},$$

(5.3.64) where $o_i(1/k_i(n))$ $(i = 1,2)$ is a quantity such that

$$\lim_{n\to\infty} \frac{o_i(1/k_i(n))}{1/k_i(n)} = \lim_{n\to\infty} k_i(n) o_i(1/k_i(n)) = 0.$$

Proof: We will show that, for the new sequence, conditions (A') and

(B') (see (5.3.5) and (5.3.6)) hold.

Since (A') holds for the original sequence $\{Z_{n1}, Z_{n2}\}$ with $L(\cdot|\theta)$ a <u>continuous</u> d.f., it will also hold for (5.3.65), by Lemma (5.3.61) (more properly, by its multi-dimensional analog, which is proven similarly).

Since (B') holds for the original sequence $\{Z_{n1}, Z_{n2}\}$ with constants $a_{nij} = (a_{nij1}, a_{nij2})$ $(i,j = 0,1)$, it will hold for sequence (5.3.65) with a'_{nij} given by

$$\begin{cases} a'_{n001} = a_{n001} - k_1(n)o_1(1/k_1(n)), \; a'_{n002} = a_{n002} - k_2(n)o_2(1/k_2(n)) \\ a'_{n011} = a_{n011} - k_1(n)o_1(1/k_1(n)), \; a'_{n012} = a_{n012} - k_2(n)o_2(1/k_2(n)) \\ a'_{n101} = a_{n101} - k_1(n)o_1(1/k_1(n)), \; a'_{n102} = a_{n102} - k_2(n)o_2(1/k_2(n)) \\ a'_{n111} = a_{n111} - k_1(n)o_1(1/k_1(n)), \; a'_{n112} = a_{n112} - k_2(n)o_2(1/k_2(n)). \end{cases}$$

(Whenever the a_{nij} converge in probability to zero the a'_{nij} do also.)

A typical $o_i(1/k_i(n))$ might be $1/\{k_i(n)n^{\delta_i}\}$ with $\delta_i > 0$ fixed $(i = 1,2)$. In comparing any two members of this class of GMLE's with each other, we find by Theorem (5.3.13) that they have the same asymptotic efficiency (in the sense of Theorem (5.3.13)).

After results (5.3.61) and (5.3.64) were obtained, the author's attention was called to the latter part of section 3 of a preliminary version of Weiss and Wolfowitz (1967b), where a generalization of Theorem (5.3.64) was stated without proof. Namely, if $\{Z_{n1}(X(n),r), Z_{n2}(X(n),r)\}$ is a sequence of GMLE's then so is $\{Z_{n1}+T'_{n1}, Z_{n2}+T'_{n2}\}$ where (T'_{n1}, T'_{n2}) is such that, uniformly in θ, (5.3.66) $$\lim_{n\to\infty} P_\theta[|k_1(n)T'_{n1}| < \delta,\ |k_2(n)T'_{n2}| < \delta] = 1$$

for any given $\delta > 0$. Our proof can be generalized to this case. (Note that in the published version of Weiss and Wolfowitz (1967b) condition (5.3.66) has apparently been weakened.) These results will now be used to compare the MLE and the GMLE with regard to asymptotic efficiency when $k = 2$.

LEMMA: For any $a > 0$, $P_\mu[\bar{X}_{[2]}-\bar{X}_{[1]} > a\,\sigma/\sqrt{n}]$ is minimized (over $\mu \in \Theta(\eta^*)$ i.e. over μ such that $\mu_{[2]} = \mu_{[1]} + \eta$ for some

(5.3.67) $\eta \geq \eta^* > 0$) at $\mu_{[2]} = \mu_{[1]} + \eta^*$. Also

$P_{\mu_{[2]}=\mu_{[1]}+\eta^*}[\bar{X}_{[2]}-\bar{X}_{[1]} > a\,\sigma/\sqrt{n}] \to 1$ as $n\to\infty$.

Proof: By Theorem (B.3.2),

$$P_\mu[\bar{X}_{[2]}-\bar{X}_{[1]} > a\sigma/\sqrt{n}] = \frac{\sqrt{n}}{2\sigma\sqrt{\pi}} \int_{a\sigma/\sqrt{n}}^{\infty} \left\{ e^{-\frac{1}{4}\left(\frac{y-\eta}{\sigma/\sqrt{n}}\right)^2} + e^{-\frac{1}{4}\left(\frac{y+\eta}{\sigma/\sqrt{n}}\right)^2} \right\} dy$$

(5.3.68) $$= \frac{1}{\sqrt{2\pi}} \int_{\frac{a\frac{\sigma}{\sqrt{n}}-\eta}{\sqrt{2}\sigma/\sqrt{n}}}^{\infty} e^{-\frac{1}{2}y^2}\,dy + \frac{1}{\sqrt{2\pi}} \int_{\frac{a\frac{\sigma}{\sqrt{n}}+\eta}{\sqrt{2}\sigma/\sqrt{n}}}^{\infty} e^{-\frac{1}{2}y^2}\,dy$$

$$= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\frac{-a\frac{\sigma}{\sqrt{n}}-\eta}{\sqrt{2}\sigma/\sqrt{n}}} + \int_{\frac{a\frac{\sigma}{\sqrt{n}}-\eta}{\sqrt{2}\sigma/\sqrt{n}}}^{\infty} \right) e^{-\frac{1}{2}y^2}\,dy.$$

By the formula for differentiation with respect to a parameter (e.g., Wadsworth and Bryan (1960), p. 2) or by the Chain Rule, since

$(a\frac{\sigma}{\sqrt{n}}+\eta)^2 > (a\frac{\sigma}{\sqrt{n}}-\eta)^2$,

$$\frac{d}{d\eta}P_\mu[\overline{X}_{[2]}-\overline{X}_{[1]} > a\sigma/\sqrt{n}] = \frac{1}{\sqrt{2\pi}}\left[e^{\frac{1}{2}\left(\frac{a\frac{\sigma}{\sqrt{n}}+\eta}{\sqrt{2}\sigma/\sqrt{n}}\right)^2}\frac{1}{\sqrt{2}\sigma/\sqrt{n}} - e^{\frac{1}{2}\left(\frac{a\frac{\sigma}{\sqrt{n}}-\eta}{\sqrt{2}\sigma/\sqrt{n}}\right)^2}\frac{1}{\sqrt{2}\sigma/\sqrt{n}}\right] > 0.$$

Hence $P_\mu[\overline{X}_{[2]}-\overline{X}_{[1]} > a\sigma/\sqrt{n}]$ is an increasing function of $\eta \geq \eta^* > 0$, and is therefore minimized when $\eta = \eta^* > 0$ (i.e. when $\mu_{[2]} = \mu_{[1]}+\eta^*$). That this minimal probability $\to 1$ as $n\to\infty$ follows from (5.3.68).

(5.3.69) <u>LEMMA</u>: $\left|\frac{d^2n}{\sigma^2}-\varepsilon_o\right| \leq 2$, where ε_o is the positive solution of (5.1.39).

<u>Proof</u>: From (5.1.39) and the fact that $\coth(x) > 1$ for $x > 0$,

$$\left|\frac{d^2n}{\sigma^2}-\varepsilon_o\right| = \left|\varepsilon_o-\varepsilon_o\coth(\varepsilon_o/2)\right| = \varepsilon_o(\coth(\varepsilon_o/2)-1).$$

Using an expression for $\coth(\varepsilon_o/2)$ which was found in the proof of Lemma (5.1.33), this becomes

$$\left|\frac{d^2n}{\sigma^2}-\varepsilon_o\right| = \varepsilon_o\left(\frac{e^{\varepsilon_o/2}+e^{-\varepsilon_o/2}}{e^{\varepsilon_o/2}-e^{-\varepsilon_o/2}}-1\right) = \varepsilon_o\frac{2e^{-\varepsilon_o/2}}{e^{\varepsilon_o/2}-e^{-\varepsilon_o/2}} = 2\frac{\varepsilon_o}{e^{\varepsilon_o}-1} \leq 2,$$

since (for $x \geq 0$) $x/(e^x-1) \leq 1$, or $x \leq e^x-1$, because $x+1 \leq e^x = 1+x+\frac{x^2}{2!}+\ldots.$

In the notation at (5.3.66), we wish to show that the MLE $\{\overline{X}_{[1]}+T'_{n1},\overline{X}_{[2]}+T'_{n2}\}$ is such that (5.3.66) holds, with $k_1(n) = k_2(n) = \sqrt{n}/\sigma$. By Theorem (5.1.37),

$$|T'_{n1}| = |\hat{\mu}_{[1]}-\overline{X}_{[1]}|$$

$$= \begin{cases} \left|\dfrac{\overline{X}_{[1]}+\overline{X}_{[2]}}{2} - \overline{X}_{[1]}\right| & \text{if } 0 \leq \overline{X}_{[2]}-\overline{X}_{[1]} \leq \sqrt{2}\sigma/\sqrt{n} \\ \left|\dfrac{\overline{X}_{[1]}+\overline{X}_{[2]}}{2} - \dfrac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2\coth(\varepsilon_o/2)}-\overline{X}_{[1]}\right| & \text{if } \quad \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n} \end{cases}$$

(5.3.70)

$$= \begin{cases} \dfrac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2} & \text{if } 0 \leq \overline{X}_{[2]}-\overline{X}_{[1]} \leq \sqrt{2}\sigma/\sqrt{n} \\ \dfrac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2}\left|1- \dfrac{1}{\coth(\varepsilon_o/2)}\right| & \text{if } \quad \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n} \end{cases}$$

and $|T'_{n2}| = |\hat{\mu}_{[2]}-\overline{X}_{[2]}|$ turns out to be the same. Thus, using the definition $d = \overline{X}_{[2]}-\overline{X}_{[1]}$ and the fact that $\varepsilon_o\coth(\varepsilon_o/2) = d^2n/\sigma^2$, for any $\delta > 0$

$$P_\theta[|k_1(n)T'_{n1}| < \delta, |k_2(n)T'_{n2}| < \delta] = P_\mu[|T'_{n1}| < \delta\sigma/\sqrt{n}]$$

$$= P_\mu[\overline{X}_{[2]}-\overline{X}_{[1]} < 2\delta\sigma/\sqrt{n}, 0 \leq \overline{X}_{[2]}-\overline{X}_{[1]} \leq \sqrt{2}\sigma/\sqrt{n}]$$

$$+ P_\mu\left[\frac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2}\left|1- \frac{1}{\coth(\varepsilon_o/2)}\right| < \delta\sigma/\sqrt{n}, \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right]$$

$$\geq P_\mu\left[\frac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2}\left|1- \frac{1}{\coth(\varepsilon_o/2)}\right| < \delta\sigma/\sqrt{n}, \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right]$$

(5.3.71)

$$= P_\mu\left[\frac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2}\left|1-\frac{\varepsilon_o\sigma^2}{d^2n}\right| < \delta\frac{\sigma}{\sqrt{n}},\ \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right]$$

$$= P_\mu\left[\frac{\overline{X}_{[2]}-\overline{X}_{[1]}}{2}\ \frac{|d^2n-\varepsilon_o\sigma^2|}{d^2n} < \delta\sigma/\sqrt{n},\ \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right]$$

$$= P_\mu\left[\frac{1}{2}\ \frac{|d^2n-\varepsilon_o\sigma^2|}{n} < (\overline{X}_{[2]}-\overline{X}_{[1]})\delta\sigma/\sqrt{n},\ \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right]$$

$$= P_\mu\left[\overline{X}_{[2]}-\overline{X}_{[1]} > \frac{1}{2\delta}\ \frac{\sigma}{\sqrt{n}}\left|\frac{d^2n}{\sigma^2}-\varepsilon_o\right|,\ \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}\right].$$

<u>THEOREM</u>: For the MLE when $k = 2$, uniformly in μ, for any given $\delta > 0$,

(5.3.72)
$$\lim_{n\to\infty} P_\mu\,[\,|k_1(n)T'_{n1}| < \delta, |k_2(n)T'_{n2}| < \delta] = 1.$$

<u>Proof</u>: By Lemma (5.3.69) and equation (5.3.71),

$$P_\mu\,[\,|k_1(n)T'_{n1}| < \delta,\ |k_2(n)T'_{n2}| < \delta]$$

(5.3.73)
$$\geq P_\mu\,[\overline{X}_{[2]}-\overline{X}_{[1]} > \frac{1}{2\delta}\ \frac{\sigma}{\sqrt{n}}2,\ \overline{X}_{[2]}-\overline{X}_{[1]} > \sqrt{2}\sigma/\sqrt{n}]$$

$$= P_\mu\,[\overline{X}_{[2]}-\overline{X}_{[1]} > \max(\sqrt{2},\tfrac{1}{\delta})\sigma/\sqrt{n}].$$

By Lemma (5.3.67), the last member of (5.3.73) can be bounded below, for $\mu \in \Theta(n^*)$, in such a way that the bound $\to 1$ as $n\to\infty$.

By Theorem (5.3.72) it follows, as noted above (5.3.66), that the MLE and the GMLE have (for $k = 2$) the same asymptotic efficiency, and that the MLE is a GMLE. This proves asymptotic efficiency properties for the MLE which do not follow directly from the standard theory, which assumes i.i.d. observations.

5.4. MAXIMUM PROBABILITY ESTIMATORS (MPE's)

Maximum probability estimators were introduced by Weiss and Wolfowitz (1967b) for much the same reason as GMLE's were introduced by Weiss and Wolfowitz (1966), as discussed in Section 5.3 above. Weiss and Wolfowitz (1967b), pp. 202-203, proved that, for the case of $m = 1$ parameter, every GMLE is an MPE; thus MPE's extend the notion of GMLE's (and by finding a GMLE we find a fortiori an MPE). We now study the extension of this result to $m \geq 1$ parameters, first summarizing Weiss and Wolfowitz's results.

Let Θ and $\overline{\Theta}$ be as in (the m-dimensional analog of) (5.3.1), let $X(n)$ be as in (5.3.2), and let $K_n(x|\theta)$ and μ_n be as in (5.3.3).

<u>DEFINITION:</u> Let R be a fixed region of $\mathcal{R}^m$, let $k(n) = (k_1(n),\ldots,k_m(n))$ be such that $k(n)\to\infty$, let $d = (d_1,\ldots,d_m)$, and define

(5.4.1)
$$d - R/k(n) = \{(z_1,\ldots,z_m) \in \overline{\Theta} : d_i - y_i/k_i(n) = z_i,\ i = 1,\ldots,m,\ (y_1,\ldots,y_m) \in R\}.$$

<u>DEFINITION:</u> Z_n is a <u>maximum probability estimator</u> with respect to R and $k(n)$ if (for a.e. (μ_n) value x of $X(n)$) $Z_n(x)$ equals a $d \in \overline{\Theta}$ such that

(5.4.2)
$$\int_{d-[k(n)]^{-1}R} \cdots \int K_n(x|\theta)d\theta_1\ldots d\theta_m = \sup_{t\in\overline{\Theta}} \int_{t-[k(n)]^{-1}R} \cdots \int K_n(x|\theta)d\theta_1\ldots d\theta_m.$$

CONDITION: For each $h > 0$ and $\theta_o \in \Theta$

(5.4.3) $$\lim_{n\to\infty} P_\theta[k(n)(Z_n-\theta) \in R] = \beta$$

uniformly for all $\theta \in H = \{\theta: |k(n)(\theta-\theta_o)| \leq h\}$.

CONDITION: For each $\theta_o \in \Theta$

(5.4.4) $$\lim_{\substack{n\to\infty \\ M\to\infty}} P_\theta[|k(n)(Z_n-\theta)| < M] = 1$$

uniformly for all θ in some neighborhood of θ_o.

CONDITION: For each $\theta_o \in \Theta$ and $h > 0$

(5.4.5) $$\lim_{n\to\infty} \{P_\theta[k(n)(T_n-\theta) \in R] - P_{\theta_o}[k(n)(T_n-\theta_o) \in R]\} = 0$$

uniformly for all $\theta \in H = \{\theta: |k(n)(\theta-\theta_o)| \leq h\}$.

(5.4.6) THEOREM: Let $\{Z_n\}$ be an MPE with respect to R and $k(n)$. Suppose $\{Z_n\}$ satisfies (5.4.3) and (5.4.4). Let $\{T_n\}$ be any estimator which satisfies (5.4.5). Then (for each $\theta_o \in \Theta$)

$$\beta \geq \overline{\lim_{n\to\infty}}\, P_{\theta_o}[k(n)(T_n-\theta_o) \in R].$$

(5.4.7) THEOREM: Let W_n be a GMLE (with respect to $r = (r_1,\ldots,r_m) > 0$) for the estimation of $\theta = (\theta_1,\ldots,\theta_m) \in \Theta$ ($m \geq 1$). Choose $R = \{(y_1,\ldots,y_m): -r_i/2 < y_i \leq r_i/2,\ i = 1,\ldots,m\}$ and $k(n)$ as for the GMLE. If the MPE (w.r.t. this R and $k(n)$) satisfies (5.4.3) and (5.4.4), and if the GMLE satisfies (5.4.5), then the GMLE is (in the equivalence class of) such an MPE.

Proof: Let Z_n be the MPE w.r.t. this R and k(n). It then satisfies the condition of Theorem (5.3.13). Thus (for each $\theta_o \in \Theta$)

$$(5.4.8) \qquad \varliminf_{n\to\infty} P_{\theta_o}[k(n)(W_n-\theta_o) \in R] \geq \varlimsup_{n\to\infty} P_{\theta_o}[k(n)(Z_n-\theta_o) \in R].$$

The GMLE W_n satisfies (5.4.5) and thus the conclusion of Theorem (5.4.6) holds: for each $\theta_o \in \Theta$

$$(5.4.9) \qquad \varliminf_{n\to\infty} P_{\theta_o}[k(n)(Z_n-\theta_o) \in R] \geq \varlimsup_{n\to\infty} P_{\theta_o}[k(n)(W_n-\theta_o) \in R].$$

Then (see Weiss and Wolfowitz (1967b), p. 198) the GMLE is (in the equivalence class of such) an MPE.

The result of Weiss and Wolfowitz (1967b) for the case m = 1 is somewhat stronger than our Theorem (5.4.7) for the case $m \geq 1$: they show that the MPE satisfies (5.4.3) and (5.4.4). (They assume, as we do, that the GMLE satisfies (5.4.5), which is stronger than (A') of (5.3.5).) Our result (more precisely, a slight extension of our result) says that if the MPE for a problem is "good" (i.e., satisfies (5.4.3) and (5.4.4)), then the GMLE (if it meets (5.4.5)) is equivalent to it. Note that the analog for m > 1 of Weiss and Wolfowitz's result for m = 1 is false. E.g., Weiss and Wolfowitz (1967b), p. 198, last paragraph, note an example (with m = 2) where the MPE is not "good" although the GMLE is. (Weiss and Wolfowitz give a method for attacking the problem, in such cases, by modifying it slightly and thereby obtaining (often "good") MPE's.)

We will now study in detail the MPE of the ranked means. Although we have seen that, in general, for m > 1 parameters even if a GMLE and an MPE both exist the MPE may not be good, in our case the MPE is shown (for the case $m = 2$) to have all the good properties of the GMLE. Thus, let $\Theta = \{\mu : \mu \in \Omega_o, \mu_1 = \mu_{[1]}, \ldots, \mu_k = \mu_{[k]}\}$ and $\overline{\Theta} = R^k$, and let $X(n)$, $K_n(x|\mu)$, μ_n be as specified in (5.3.14). Fix $r = (r_1, \ldots, r_k) > 0$, and choose $k_1(n) = \ldots = k_k(n) = \sqrt{n}/\sigma$, $R = \{(y_1, \ldots, y_k): -r_i/2 < y_i \le r_i/2, i = 1, \ldots, k\}$. Then

$$\begin{aligned} d - [k(n)]^{-1}R &= \{(z_1, \ldots, z_k) \in \overline{\Theta}: \\ &\qquad d_i - y_i/k_i(n) = z_i,\ i = 1, \ldots, k,\ (y_1, \ldots, y_k) \in R\} \\ &= \{(z_1, \ldots, z_k): d_i - \frac{r_i}{2k_i(n)} \le z_i < d_i + \frac{r_i}{2k_i(n)},\ i = 1, \ldots, k\}, \end{aligned} \tag{5.4.10}$$

and

$$\begin{aligned} &\sup_{t \in \overline{\Theta}} \int \cdots \int_{t-[k(n)]^{-1}R} K_n(x|\mu)\, d\mu_{[1]} \cdots d\mu_{[k]} \\ &= \sup_{t_1, \ldots, t_k} \int_{t_k - \frac{r_k}{2}\sigma/\sqrt{n}}^{t_k + \frac{r_k}{2}\sigma/\sqrt{n}} \cdots \int_{t_1 - \frac{r_1}{2}\sigma/\sqrt{n}}^{t_1 + \frac{r_1}{2}\sigma/\sqrt{n}} K_n(x|\mu)\, d\mu_{[1]} \cdots d\mu_{[k]}. \end{aligned} \tag{5.4.11}$$

For the case $k = 2$, (5.4.11) becomes (when $\overline{X}_{[1]} = x_1$ and $\overline{X}_{[2]} = x_2$)

$$\sup_{t_1, t_2} \frac{n}{2\pi\sigma^2} \int_{t_2 - \frac{r_2}{2}\sigma/\sqrt{n}}^{t_2 + \frac{r_2}{2}\sigma/\sqrt{n}} \int_{t_1 - \frac{r_1}{2}\sigma/\sqrt{n}}^{t_1 + \frac{r_1}{2}\sigma/\sqrt{n}} \left\{ e^{-\frac{1}{2}\left(\frac{x_1 - \mu_{[1]}}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_2 - \mu_{[2]}}{\sigma/\sqrt{n}}\right)^2} \right.$$

$$\left. + e^{-\frac{1}{2}\left(\frac{x_2 - \mu_{[1]}}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{x_1 - \mu_{[2]}}{\sigma/\sqrt{n}}\right)^2} \right\} d\mu_{[1]}\, d\mu_{[2]}$$

$$= \sup_{t_1,t_2}\left[\int_{\frac{t_2-x_2}{\sigma/\sqrt{n}}-\frac{r_2}{2}}^{\frac{t_2-x_2}{\sigma/\sqrt{n}}+\frac{r_2}{2}}\int_{\frac{t_1-x_1}{\sigma/\sqrt{n}}-\frac{r_1}{2}}^{\frac{t_1-x_1}{\sigma/\sqrt{n}}+\frac{r_1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu_2^2}d\nu_1\,d\nu_2\right.$$

$$\left.+\int_{\frac{t_2-x_1}{\sigma/\sqrt{n}}-\frac{r_2}{2}}^{\frac{t_2-x_1}{\sigma/\sqrt{n}}+\frac{r_2}{2}}\int_{\frac{t_1-x_2}{\sigma/\sqrt{n}}-\frac{r_1}{2}}^{\frac{t_1-x_2}{\sigma/\sqrt{n}}+\frac{r_1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu_2^2}d\nu_1\,d\nu_2\right] \qquad (5.4.12)$$

$$= \sup_{t_1,t_2}\left[\left\{\Phi\left(\frac{t_1-x_1}{\sigma/\sqrt{n}}+\frac{r_1}{2}\right)-\Phi\left(\frac{t_1-x_1}{\sigma/\sqrt{n}}-\frac{r_1}{2}\right)\right\}\cdot\left\{\Phi\left(\frac{t_2-x_2}{\sigma/\sqrt{n}}+\frac{r_2}{2}\right)-\Phi\left(\frac{t_2-x_2}{\sigma/\sqrt{n}}-\frac{r_2}{2}\right)\right\}\right.$$

$$\left.+\left\{\Phi\left(\frac{t_1-x_2}{\sigma/\sqrt{n}}+\frac{r_1}{2}\right)-\Phi\left(\frac{t_1-x_2}{\sigma/\sqrt{n}}-\frac{r_1}{2}\right)\right\}\cdot\left\{\Phi\left(\frac{t_2-x_1}{\sigma/\sqrt{n}}+\frac{r_2}{2}\right)-\Phi\left(\frac{t_2-x_1}{\sigma/\sqrt{n}}-\frac{r_2}{2}\right)\right\}\right].$$

<u>LEMMA</u>: Let $d \equiv \frac{\sqrt{n}}{\sigma}(x_2-x_1)$, $t_1 = x_1 + a_1\sigma/\sqrt{n}$, $t_2 = x_2 - a_2\sigma/\sqrt{n}$.

Then an MPE is (t_1,t_2) with a_1,a_2 which achieve

$$\sup_{a_1,a_2}[\{\Phi(a_1+r_1/2)-\Phi(a_1-r_1/2)\}\{\Phi(a_2+r_2/2)-\Phi(a_2-r_2/2)\}$$
$$+\{\Phi(a_1-d+r_1/2)-\Phi(a_1-d-r_1/2)\}\{\Phi(a_2-d+r_2/2)-\Phi(a_2-d-r_2/2)\}]. \qquad (5.4.13)$$

<u>Proof</u>: By definition (5.4.2), for our case as specified above (5.4.10),

the MPE is (t_1,t_2) which achieves the supremum in (5.4.12). If we use $d = \frac{\sqrt{n}}{\sigma}(x_2-x_1)$ and transform via $t_1 = x_1 + a_1\sigma/\sqrt{n}$, $t_2 = x_2 - a_2\sigma/\sqrt{n}$, this (t_1,t_2) will be specified by the (a_1,a_2) which achieves the

$$\sup_{a_1,a_2} [\{\Phi(a_1+r_1/2) - \Phi(a_1-r_1/2)\}\{\Phi(-a_2+r_2/2) - \Phi(-a_2-r_2/2)\}$$

$$+ \{\Phi(a_1-d+r_1/2) - \Phi(a_1-d-r_1/2)\}\{\Phi(-a_2+d+r_2/2) - \Phi(-a_2+d-r_2/2)\}].$$

Using the relation $\Phi(x) = 1 - \Phi(-x)$ $(x \in \mathbb{R})$, this becomes as specified in the statement of the lemma.

(5.4.14) LEMMA: The supremum of (5.4.13) occurs only at (a_1,a_2) with $0 < a_1 < d$, $0 < a_2 < d$.

Proof: By reasoning as at (5.1.5), the supremum must occur at a critical point. However, if we set the partial derivative with respect to a_1 equal to zero we obtain

$$\frac{\phi(a_1+r_1/2) - \phi(a_1-r_1/2)}{\phi(a_1-d+r_1/2) - \phi(a_1-d-r_1/2)} = -\frac{\Phi(a_2-d+r_2/2) - \Phi(a_2-d-r_2/2)}{\Phi(a_2+r_2/2) - \Phi(a_2-r_2/2)} .$$

Since the r.h.s. is always < 0, the l.h.s. must always be < 0. Now, the denominator of the l.h.s. is positive (negative) iff $a_1 < d$ $(a_1 > d)$. Thus, we must have

$$\phi(a_1+\frac{r_1}{2}) - \phi(a_1-\frac{r_1}{2}) \begin{matrix} < 0 & \text{if} & a_1 < d \\ > 0 & \text{if} & a_1 > d \end{matrix}$$

i.e.

$$\begin{matrix} a_1 > 0 & \text{if} & a_1 < d \\ a_1 < 0 & \text{if} & a_1 > d. \end{matrix}$$

This proves the result for a_1; the result for a_2 follows similarly.

(5.4.15) LEMMA: By imposing a consistency criterion for estimators similar to (5.1.4), we may restrict ourselves to (a_1,a_2) with $a_1 + a_2 \leq d$.

Proof: In order that we have $t_1 \leq t_2$, we must have $x_1 + a_1\sigma/\sqrt{n} \leq x_2 - a_2\sigma/\sqrt{n}$, i.e., $a_1 + a_2 \leq \frac{\sqrt{n}}{\sigma}(x_2-x_1) = d$.

Note that, in the region of (a_1,a_2)-space in which Lemma (5.4.14) tells us the supremum of (5.4.13) must lie, we have symmetry (of values of (5.4.13)) about the line $a_1 + a_2 = d$; see Figure (5.4.16). Thus, our consistency criterion only eliminates an illogical duplicate maximizing point.

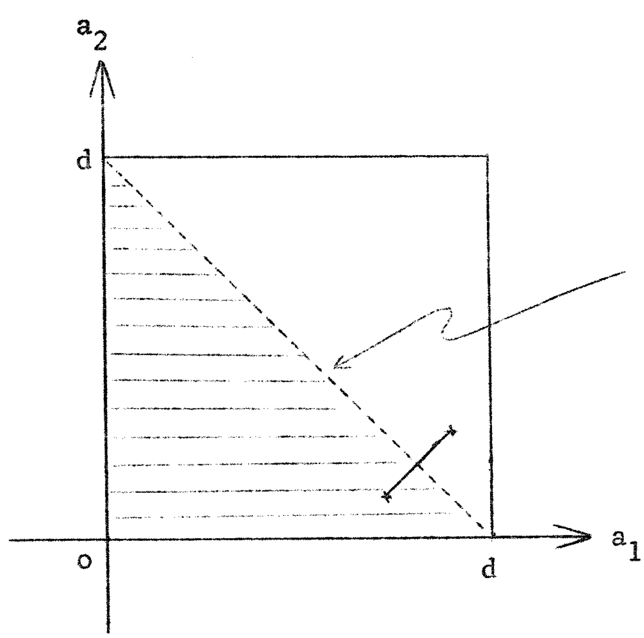


Figure (5.4.16)

(5.4.17) LEMMA: For any fixed $\delta > 0$, there is a $K(r_1,r_2,\delta)$ such that if $d \geq K(r_1,r_2,\delta)$ then (5.4.13) is maximized (in the shaded region I: $a_1 > 0$, $a_2 > 0$, $a_1 + a_2 \leq d$ of Figure (5.4.16)) inside the disk D: $a_1^2 + a_2^2 \leq \delta$.

Proof: Let $f_1 = \{\Phi(a_1+r_1/2) - \Phi(a_1-r_1/2)\}\{\Phi(a_2+r_2/2) - \Phi(a_2-r_2/2)\}$, $f_2 = \{\Phi(a_1-d+r_1/2) - \Phi(a_1-d-r_1/2)\}\{\Phi(a_2-d+r_2/2) - \Phi(a_2-d-r_2/2)\}$; then (5.4.13) is $\sup\limits_{(a_1,a_2) \text{ in } I} (f_1+f_2)$.

Now over $(a_1,a_2) \in I$, f_1 is maximized at $(a_1,a_2) = (0,0)$ and decreases as a_1 and a_2 increase. Thus, if we move (a_1,a_2) outside D, the loss in f_1 is at least $f_1((0,0))$ minus the largest value of $f_1((a_1,a_2))$ on the boundary of D inside I; there $a_1^2 + a_2^2 = \delta$, so

$$\sup_{\substack{a_1^2+a_2^2=\delta \\ (a_1,a_2) \text{ in } I}} f_1((a_1,a_2)) = \sup_{0\leq a_1\leq\delta} \{\Phi(a_1+r_1/2) - \Phi(a_1-r_1/2)\}\cdot$$

$$\cdot\{\Phi(\sqrt{\delta-a_1^2}+r_2/2) - \Phi(\sqrt{\delta-a_1^2}-r_2/2)\}$$

$$\leq \{\Phi(c_1\delta+r_1/2) - \Phi(c_1\delta-r_1/2)\}\{\Phi(r_2/2) - \Phi(-r_2/2)\},$$

where we may suppose without loss that $c_1 = c_1(r_1,r_2,\delta) > 0$. (This can only fail if the supremum occurs at $(a_1,a_2) = (0,\delta)$, in which case we may reverse the roles played by a_1 and a_2 in our inequality and the argument below will go through similarly.) Thus, the loss in f_1 via going outside D is at least

$$\{\Phi(r_1/2) - \Phi(-r_1/2)\}\{\Phi(r_2/2) - \Phi(-r_2/2)\}$$

$$-\{\Phi(c_1\delta+r_1/2) - \Phi(c_1\delta-r_1/2)\}\{\Phi(r_2/2) - \Phi(-r_2/2)\}$$

$$= \{\Phi(r_2/2) - \Phi(-r_2/2)\}[\{\Phi(r_1/2) - \Phi(-r_1/2)\}-\{\Phi(c_1\delta+r_1/2) - \Phi(c_1\delta-r_1/2)\}]$$

$$= c_2(r_2)c_3(r_1,r_2,\delta) \qquad \text{(say)}.$$

The gain in f_2 (which increases as a_1 and a_2 increase in I) is less than

$$\sup_{(a_1,a_2) \text{ in } I} \Phi(a_1-d+r_1/2)\Phi(a_2-d+r_2/2)$$

$$\leq \sup_{(a_1,a_2) \text{ in } I} \Phi(a_1-d+\max(r_1,r_2))\Phi(a_2-d+\max(r_1,r_2))$$

$$= \sup_{\substack{a_1+a_2=d \\ a_1,a_2\geq 0}} \Phi(a_1-d+\max(r_1,r_2))\Phi(a_2-d+\max(r_1,r_2))$$

$$= \sup_{0\leq a_1\leq d} \Phi(a_1-d+\max(r_1,r_2))\Phi(-a_1+\max(r_1,r_2)).$$

We will show that

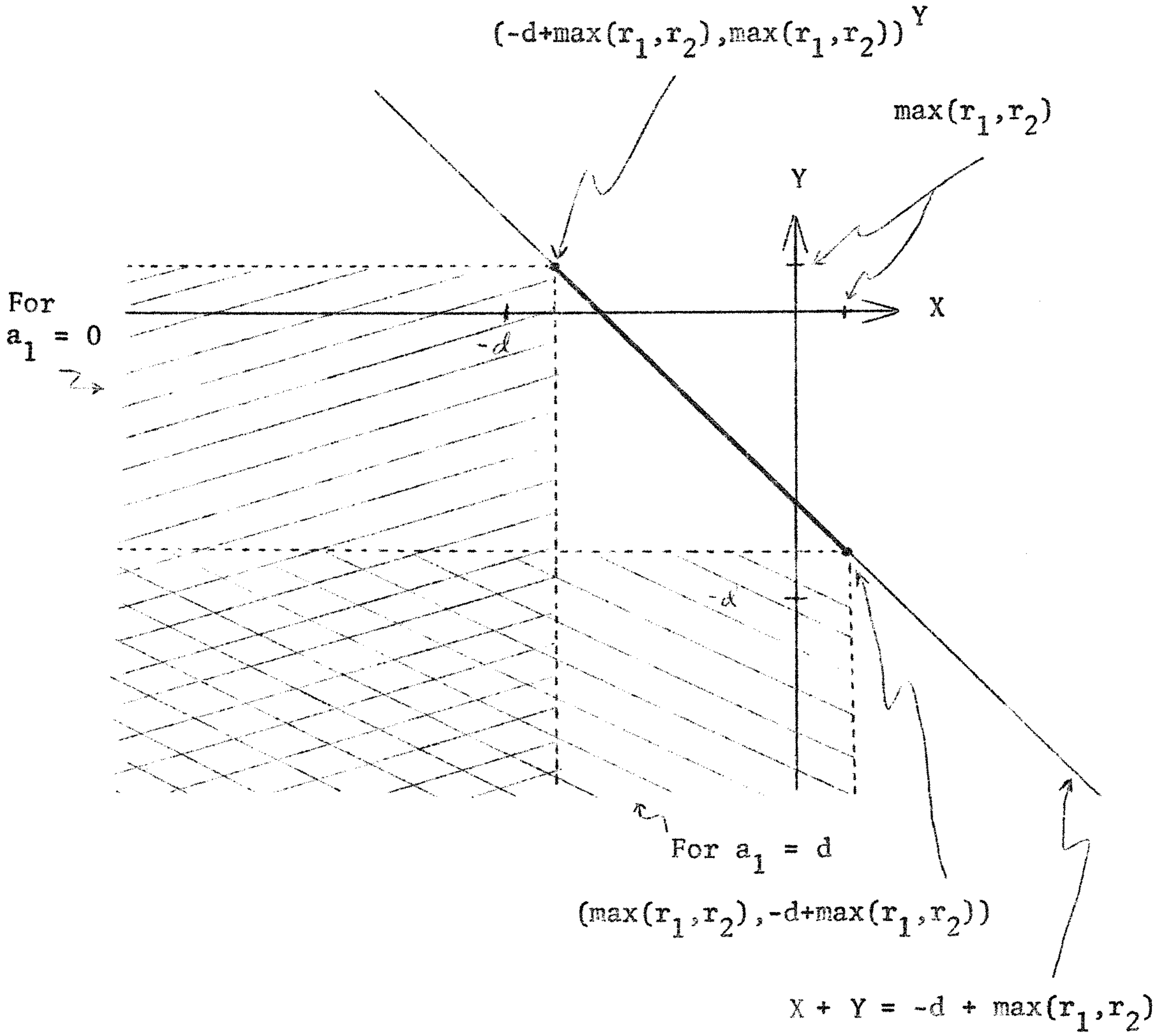
$$\lim_{d\to\infty} \sup_{0\leq a_1\leq d} \Phi(a_1-d+\max(r_1,r_2))\Phi(-a_1+\max(r_1,r_2)) = 0. \tag{5.4.18}$$

Thus, there will exist a $K(r_1,r_2,\delta)$ such that $d \geq K(r_1,r_2,\delta)$ implies the gain is less than $c_2(r_2)\ c_3(r_1,r_2,\delta)$, which will prove the lemma.

Let X and Y be i.i.d. N(0,1) r.v.'s. Then (5.4.18) is equal to

$$\lim_{d\to\infty} \sup_{0\leq a_1\leq d} P[X \leq a_1-d+\max(r_1,r_2),\ Y \leq -a_1+\max(r_1,r_2)], \tag{5.4.19}$$

which involves the probability in a certain rectangle in R^2, as illustrated in Figure (5.4.20).



$(-d+\max(r_1,r_2),\max(r_1,r_2))$
Y
$\max(r_1,r_2)$
Y
X
For
$a_1 = 0$
$-d$
$-d$
For $a_1 = d$
$(\max(r_1,r_2),-d+\max(r_1,r_2))$
$X + Y = -d + \max(r_1,r_2)$

Figure (5.4.20)

Thus, (5.4.19) is less than or equal to the limit of the supremum of the probability to the left of the line $X + Y = -d + \max(r_1,r_2)$,

$$\lim_{d\to\infty} \sup_{0\le a_1\le d} P[X+Y \le -d+\max(r_1,r_2)] = \lim_{d\to\infty} P[X+Y \le -d+\max(r_1,r_2)] = 0.$$

(5.4.21) THEOREM: For $\mu \in \Theta(\eta^*)$ (see (5.3.15)), the MPE (t_1,t_2) is equivalent to the GMLE $(\overline{X}_{[1]},\overline{X}_{[2]})$ found in Section 5.3, and thus has the same optimum property as that GMLE.

Proof: We wish to show that, for each $\mu \in \Theta(\eta^*)$ and for each fixed $\delta > 0$,

$$\begin{aligned} 1 &= \lim_{n\to\infty} P_\mu[k(n)\max(|t_1-\overline{X}_{[1]}|,|t_2-\overline{X}_{[2]}|) < \delta] \\ &= \lim_{n\to\infty} P_\mu[\tfrac{\sqrt{n}}{\sigma}\max(|a_1|\sigma/\sqrt{n},|a_2|\sigma/\sqrt{n}) < \delta] \\ &= \lim_{n\to\infty} P_\mu[\max(a_1,a_2) < \delta], \end{aligned}$$

where the last equality uses Lemma (5.4.14). Now by Theorem (B.3.2), the density of $d = \frac{\sqrt{n}}{\sigma}(\overline{X}_{[2]}-\overline{X}_{[1]})$ for $y \ge 0$ is

$$\frac{1}{2\sqrt{\pi}}\left\{e^{-\frac{1}{4}\left(y-\frac{\eta}{\sigma/\sqrt{n}}\right)^2} + e^{-\frac{1}{4}\left(y+\frac{\eta}{\sigma/\sqrt{n}}\right)^2}\right\}$$

where $\eta = \mu_{[2]} - \mu_{[1]}$. Thus $\lim_{n\to\infty} P_\mu[d \ge K(r_1,r_2,\delta)] = 1$, so using Lemma (B.2.1),

$$\lim_{n\to\infty} P_\mu[\max(a_1,a_2) < \delta] = \lim_{n\to\infty} P_\mu[\max(a_1,a_2) < \delta \mid d \ge K(r_1,r_2,\delta)] = 1,$$

where the last step uses Lemma (5.4.17).

CHAPTER 6. INTERVAL ESTIMATION

6.1. GENERAL FORMULATION

Consider joint confidence interval estimation of $\mu_{[1]},\ldots,\mu_{[k]}$. Our observed statistics under Rule (1.3.2) are X_{ij} ($i = 1,\ldots,k$; $j = 1,\ldots,n$); we take $\overline{X}_1,\ldots,\overline{X}_k$ to be fundamental as at (5.1.1) (note, as has been pointed out by Bechhofer, Kiefer, and Sobel (1968), Part I, Remark 4.1.2, that $\overline{X}_1,\ldots,\overline{X}_k$ are sufficient and transitive for $\mu_1,\ldots,\mu_k$ after n stages; see p. 426 and Theorem 10.1 of Bahadur (1954), as well as pp. 334ff of Ferguson (1967) for details of these notions), choose our interval to be of the form

$$\begin{aligned} I &= I(\overline{X}_1,\ldots,\overline{X}_k) \\ &= \{\mu_{[1]},\ldots,\mu_{[k]}: g_1 \le \mu_{[1]} \le h_1, \ldots, g_k \le \mu_{[k]} \le h_k\}, \end{aligned} \tag{6.1.1}$$

where $g_1,h_1;\ldots;g_k,h_k$ are functions of $\overline{X}_1,\ldots,\overline{X}_k$, and ask two invariance conditions (involving relabeling of populations and shifts of location).

(6.1.2) <u>SYMMETRY INVARIANCE</u>: For all $\beta \in S_k$,

$$I(\overline{X}_1,\ldots,\overline{X}_k) = I(\overline{X}_{\beta(1)},\ldots,\overline{X}_{\beta(k)}).$$

(6.1.3) <u>LOCATION INVARIANCE</u>: For all $c \in R$,

$$I(\overline{X}_1+c,\ldots,\overline{X}_k+c) = I(\overline{X}_1,\ldots,\overline{X}_k) + c.$$

Weiss (1963) pointed out, in another context, that (6.1.2) and (6.1.3) are not necessarily the only or the best ways to compensate for permutations and shifts of location, respectively: there may be other ways to compensate which yield the same interval.

(6.1.4) LEMMA: Under condition (6.1.2), $I(\overline{X}_1,\ldots,\overline{X}_k)$ must be of the form $I(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$.

Proof: Condition (6.1.2) implies that I depends only on the ordered $\overline{X}_i$ $(i = 1,\ldots,k)$.

(6.1.5) DEFINITION: Let $a_1,\ldots,a_k$ $(a_1 \geq 0, \ldots, a_k \geq 0; a_1 + \ldots + a_k = 1)$, b^* $(0 < b^* < \infty)$, and (G,H) $(-\infty \leq G \leq H \leq +\infty)$ be constants pre-set by the experimenter.

We now take our loss function to be a weighted sum of the probability that $I(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})$ doesn't cover $\mu_{[i]}$ plus a multiple of a quantity related to the length of the interval on $\mu_{[i]}$ $(i = 1,\ldots,k)$:

(6.1.6) LOSS FUNCTION: $W(\mu;I(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})) =$

$$= \sum_{i=1}^{k} a_i\{P_\mu[\mu_{[i]} \notin (g_i,h_i)] + b^* \min(h_i-g_i,H-G)\}.$$

Note that the length h_i-g_i is the special case of $\min(h_i-g_i,H-G)$ where the experimenter chooses (G,H) with $H-G = +\infty$.

(6.1.7) RISK FUNCTION:

$$r(\mu;I(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})) = E_\mu W(\mu;I(\overline{X}_{[1]},\ldots,\overline{X}_{[k]})).$$

Thus,

$$r(\mu;I) = E_\mu W(\mu;I)$$

(6.1.8)
$$= \sum_{i=1}^{k} a_i E_\mu\{P_\mu[\mu_{[i]} \notin (g_i,h_i)] + b^* \min(h_i-g_i,H-G)\}$$

$$= \sum_{i=1}^{k} a_i P_\mu [\mu_{[i]} \notin (g_i, h_i)] + b^* \sum_{i=1}^{k} a_i E_\mu \min(h_i - g_i, H-G)$$

$$= 1 - \sum_{i=1}^{k} a_i P_\mu [\mu_{[i]} \varepsilon (g_i, h_i)] + b^* \sum_{i=1}^{k} a_i E_\mu \min(h_i - g_i, H-G).$$

Our aim now is to find functions $g_1, h_1; \ldots; g_k, h_k$ which are in some sense optimal with respect to (6.1.7), e.g. which achieve the minimum

(6.1.9) $$\inf_{g_1, h_1; \ldots; g_k, h_k} \ \sup_{\mu \varepsilon \Omega_o(\mu_{[k]})} r(\mu; I(\overline{X}_{[1]}, \ldots, \overline{X}_{[k]}))$$

and provide a minimax invariant confidence interval. (The μ in (6.1.9) will be non-randomized, since μ is a fixed unknown and not a random variable; $I(\overline{X}_{[1]}, \ldots, \overline{X}_{[k]})$ will be considered non-randomized also.) Although we have been unable to carry out (6.1.9) or other optimization in the general case, results for special cases are obtained below. Note, for use below, that by Lemma (6.1.4) and (6.1.3) with $c = -\overline{X}_{[i]}$ we have the

<u>THEOREM</u>: Under conditions (6.1.2) and (6.1.3), $I(\overline{X}_1, \ldots, \overline{X}_k)$ must be of the form $I(\overline{X}_{[1]}, \ldots, \overline{X}_{[k]})$ with (for $i = 1, \ldots, k$)

(6.1.10) $$g_i = \overline{X}_{[i]} - g_i^*(\overline{X}_{[i]} - \overline{X}_{[1]}, \ldots, \overline{X}_{[i]} - \overline{X}_{[k]})$$

$$h_i = \overline{X}_{[i]} + h_i^*(\overline{X}_{[i]} - \overline{X}_{[1]}, \ldots, \overline{X}_{[i]} - \overline{X}_{[k]}).$$

6.2. INTERVALS OF FIXED WIDTH WITHIN A CERTAIN SUBCLASS

In Theorem (6.1.10) we looked at the form of intervals of type (6.1.1) under two invariance conditions. We now study the subclass of joint intervals

$$I_N(\overline{X}_{[1]},\ldots,\overline{X}_{[k]}) = \{\mu_{[1]},\ldots,\mu_{[k]}: \overline{X}_{[1]}-g_1^* \le \mu_{[1]} \le \overline{X}_{[1]}+h_1^*,\ldots,\overline{X}_{[k]}-g_k^* \le \mu_{[k]} \le \overline{X}_{[k]}+h_k^*\}, \tag{6.2.1}$$

which utilize the "natural" estimators $\overline{X}_{[i]}$ of $\mu_{[i]}$ $(i = 1,\ldots,k)$ strongly by taking $g_1^*,h_1^*;\ldots;g_k^*,h_k^*$ to be constants. Further, we will suppose the experimenter has specified positive constants $d_1,\ldots,d_k$, and wishes the interval about $\mu_{[i]}$ to be of length d_i $(i = 1,\ldots,k)$. We then study intervals of fixed width within subclass (6.2.1), i.e. the subclass of joint intervals

$$\begin{aligned} &I_{F,N}(\overline{X}_{[1]},\ldots,\overline{X}_{[k]}) \\ &= \{\mu_{[1]},\ldots,\mu_{[k]}: \overline{X}_{[1]}+(h_1^*-d_1) \le \mu_{[1]} \le \overline{X}_{[1]}+h_1^*,\ldots, \\ &\qquad \overline{X}_{[k]}+(h_k^*-d_k) \le \mu_{[k]} \le \overline{X}_{[k]}+h_k^*\}. \end{aligned} \tag{6.2.2}$$

Then (here it is logical to choose $(G,H) = (-\infty,+\infty)$)

$$\begin{aligned} &r(\mu;I_{F,N}) \\ &= 1 - \sum_{i=1}^{k} a_i P_\mu[\mu_{[i]} \in (\overline{X}_{[i]}+(h_i^*-d_i),\overline{X}_{[i]}+h_i^*)] + b^*\sum_{i=1}^{k} a_i d_i \\ &= 1 + b^*\sum_{i=1}^{k} a_i d_i - \sum_{i=1}^{k} a_i P_\mu[\mu_{[i]}-h_i^* \le \overline{X}_{[i]} \le \mu_{[i]}-h_i^*+d_i], \end{aligned} \tag{6.2.3}$$

which is of the form constant (specified by the experimenter) minus a weighted sum of probabilities of coverage of $\mu_{[1]},\ldots,\mu_{[k]}$. To find the $h_1^*,\ldots,h_k^*$ which are optimal in the sense of (6.1.9) (minimax) within subclass $I_{F,N}$ of (6.2.2), we must find the h_i^*'s which achieve

$$\text{(6.2.4)} \qquad \sup_{h_1^*,\ldots,h_k^*} \; \inf_{\mu\varepsilon\Omega_o(\mu_{[k]})} \sum_{i=1}^{k} a_i P_\mu[\mu_{[i]}-h_i^* \le \overline{X}_{[i]} \le \mu_{[i]}-h_i^*+d_i].$$

For the case $a_1 = \ldots = a_{k-1} = 0$, $a_k = 1$, suppose we set $h_k^* = d_k/2$. Then Lal Saxena and Tong (1968) claim in an abstract that

$$\inf_{\mu\varepsilon\Omega_o(\mu_{[k]})} P_\mu[\mu_{[k]}-d_k/2 \le \overline{X}_{[k]} \le \mu_{[k]}+d_k/2]$$

occurs at $\mu_{[1]} = \ldots = \mu_{[k]}$, and therefore equals $\left[\Phi\left(\frac{d_k}{2}\frac{\sqrt{n}}{\sigma}\right)\right]^k$ $- \left[\Phi\left(-\frac{d_k}{2}\frac{\sqrt{n}}{\sigma}\right)\right]^k$; i.e., if one uses the interval $(\overline{X}_{[k]}-d_k/2,\ \overline{X}_{[k]}+d_k/2)$ for $\mu_{[k]}$ then the probability of converge is a minimum when $\mu_{[1]} = \ldots = \mu_{[k]}$.

6.3. UPPER AND LOWER INTERVALS WITHIN A CERTAIN SUBCLASS

The subclass of joint intervals I_N of (6.2.1) utilizes $\overline{X}_{[i]}$ as an estimator of $\mu_{[i]}$ strongly $(i = 1,\ldots,k)$. For problems in which we wish an upper (lower) joint confidence interval on $\mu_{[1]},\ldots,\mu_{[k]}$ we will set $g_1^* = \ldots = g_k^* = +\infty$ $(h_1^* = \ldots = h_k^* = +\infty)$ in (6.2.1). Then our interval is in one of the classes

$$I_{N,U} = \{\mu_{[1]},\ldots,\mu_{[k]}: \mu_{[1]} \leq \overline{X}_{[1]}+h_1^*,\ldots,\mu_{[k]} \leq \overline{X}_{[k]}+h_k^*\} \tag{6.3.1}$$

$$I_{N,L} = \{\mu_{[1]},\ldots,\mu_{[k]}: \overline{X}_{[1]}-g_1^* \leq \mu_{[1]},\ldots,\overline{X}_{[k]}-g_k^* \leq \mu_{[k]}\} \tag{6.3.2}$$

and

$$r(\mu;I_{N,U}) = 1 - \sum_{i=1}^{k} a_i P_\mu [\overline{X}_{[i]} \geq \mu_{[i]}-h_i^*] + b^*(H-G) \tag{6.3.3}$$

$$r(\mu;I_{N,L}) = 1 - \sum_{i=1}^{k} a_i P_\mu [\overline{X}_{[i]} \leq \mu_{[i]}+g_i^*] + b^*(H-G). \tag{6.3.4}$$

For the case of upper intervals we may choose $H-G = 0$ without loss. Then

$$r(\mu;I_{N,U}) = \sum_{i=1}^{k} a_i P_\mu [\overline{X}_{[i]} \leq \mu_{[i]}-h_i^*]. \tag{6.3.5}$$

Similarly, for the case of lower intervals we may choose $H-G = 0$ without loss. Then

$$r(\mu;I_{N,L}) = \sum_{i=1}^{k} a_i P_\mu [\overline{X}_{[i]} \geq \mu_{[i]}+g_i^*]. \tag{6.3.6}$$

<u>THEOREM</u>: For any i ($1 \leq i \leq k$), if $a_i = 1$ (thus $a_j = 0$ for $j \neq i$) then the risk (6.3.5) ((6.3.6)) is the probability that our upper (lower) interval doesn't cover $\mu_{[i]}$, and is maximized over $\mu \in \Omega_o(\mu_{[i]})$ at $\mu = (\overbrace{-\infty,\ldots,-\infty}^{i-1 \text{ terms}}, \overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{k-i+1 \text{ terms}})$

(6.3.7)

$(\mu = (\overbrace{\mu_{[i]},\ldots,\mu_{[i]}}^{i \text{ terms}}, \overbrace{+\infty,\ldots,+\infty}^{k-i \text{ terms}}))$.

Thus, for any γ ($0 < \gamma < 1$) an upper (lower) confidence interval of minimal probability of coverage γ is $(-\infty, \overline{X}_{[i]}+h_i^*)$ $((\overline{X}_{[i]}-g_i^*, +\infty))$ with $h_i^* = (\sigma/\sqrt{n})\Phi^{-1}(\gamma^{\frac{1}{k-i+1}})$ $(g_i^* = (\sigma/\sqrt{n})\Phi^{-1}(\gamma^{\frac{1}{i}}))$.

<u>Proof</u>: <u>Upper Interval</u>. For any i ($1 \leq i \leq k$), if $a_i = 1$, $a_j = 0$ ($j \neq i$) then

$$\sup_{\mu \in \Omega_o(\mu_{[i]})} r(\mu; I_{N,U})$$

$$= \sup_{\mu \in \Omega_o(\mu_{[i]})} P_\mu[\overline{X}_{[i]} \leq \mu_{[i]} - h_i^*] = \sup_{\mu \in \Omega_o(\mu_{[i]})} F_{\overline{X}_{[i]}}(\mu_{[i]} - h_i^*)$$

$$= \lim_{M \to +\infty} P_{\mu_{[1]}=\cdots=\mu_{[i-1]}=-M,\mu_{[i]}=\cdots=\mu_{[k]}}[\overline{X}_{[i]} \leq \mu_{[i]} - h_i^*]$$

since, for $i = 1,\ldots,k$ and $x \in \mathcal{R}$, $F_{\overline{X}_{[i]}}(x)\downarrow$ as $\mu_\ell \uparrow$ ($\ell = 1,\ldots,k$) by Theorem (2.1.11). It follows by a modification of the proof of Case 1 of Theorem (2.2.4) (using 1 for x) that

$$\sup_{\mu \in \Omega_o(\mu_{[i]})} r(\mu; I_{N,U}) = P_{\mu_{[1]}=\cdots=\mu_{[i-1]}=-\infty,\mu_{[i]}=\cdots=\mu_{[k]}}[\overline{X}_{[i]} \leq \mu_{[i]} - h_i^*]$$

$$= P\left[\min(Y_1,\ldots,Y_{k-i+1}) \leq \frac{-h_i^*}{\sigma/\sqrt{n}}\right]$$

where $Y_1,\ldots,Y_{k-i+1}$ are k-i+1 independent N(0,1) r.v.'s. To make the minimal probability of coverage γ $(0 < \gamma < 1)$ we set h* so that

$$1-\gamma = P\left[\min(Y_1,\ldots,Y_{k-i+1}) \le \frac{-h_i^*}{\sigma/\sqrt{n}}\right]$$

$$= 1 - P\left[\min(Y_1,\ldots,Y_{k-i+1}) \ge \frac{-h_i^*}{\sigma/\sqrt{n}}\right]$$

$$= 1 - \left[1-\Phi\left(\frac{-h_i^*}{\sigma/\sqrt{n}}\right)\right]^{k-i+1} = 1 - \left[\Phi\left(\frac{h_i^*}{\sigma/\sqrt{n}}\right)\right]^{k-i+1};$$

thus $\gamma = \left[\Phi\left(\frac{h_i^*}{\sigma/\sqrt{n}}\right)\right]^{k-i+1}$, $\gamma^{\frac{1}{k-i+1}} = \Phi\left(\frac{h_i^*}{\sigma/\sqrt{n}}\right)$, and $h_i^* = (\sigma/\sqrt{n})\Phi^{-1}(\gamma^{\frac{1}{k-i+1}})$.

<u>Lower Interval</u>. For any i $(1 \le i \le k)$, if $a_i = 1$, $a_j = 0$ $(j \ne i)$ then by Theorem (2.1.11)

$$\sup_{\mu\in\Omega_o(\mu_{[i]})} r(\mu;I_{N,L}) = \sup_{\mu\in\Omega_o(\mu_{[i]})} P_\mu[\overline{X}_{[i]} \ge \mu_{[i]}+g_i^*]$$

$$= \lim_{M\to+\infty} P_{\mu_{[1]}=\cdots=\mu_{[i]},\mu_{[i+1]}=\cdots=\mu_{[k]}=M}[\overline{X}_{[i]} \ge \mu_{[i]}+g_i^*].$$

By a modification of the proof of Case 2 of Theorem (2.2.4),

$$\sup_{\mu\in\Omega_o(\mu_{[i]})} r(\mu;I_{N,L}) = P_{\mu_{[1]}=\cdots=\mu_{[i]},\mu_{[i+1]}=\cdots=\mu_{[k]}=+\infty}[\overline{X}_{[i]} \ge \mu_{[i]}+g_i^*]$$

$$= P\left[\max(Y_1,\ldots,Y_i) \ge \frac{g_i^*}{\sigma/\sqrt{n}}\right]$$

where $Y_1,\ldots,Y_i$ are i independent N(0,1) r.v.'s. To make the minimal probability of coverage γ $(0 < \gamma < 1)$ we set g* so that

$$1 - \gamma = P\left[\max(Y_1,\ldots,Y_i) \geq \frac{g_i^*}{\sigma/\sqrt{n}}\right]$$

$$= 1 - P\left[\max(Y_1,\ldots,Y_i) \leq \frac{g_i^*}{\sigma/\sqrt{n}}\right] = 1 - \left[\Phi\left(\frac{g_i^*}{\sigma/\sqrt{n}}\right)\right]^i;$$

thus $\gamma = \left[\Phi\left(\frac{g_i^*}{\sigma/\sqrt{n}}\right)\right]^i$, $\gamma^{\frac{1}{i}} = \Phi\left(\frac{g_i^*}{\sigma/\sqrt{n}}\right)$, and $g_i^* = (\sigma/\sqrt{n})\Phi^{-1}(\gamma^{\frac{1}{i}})$.

(6.3.8) THEOREM: The upper confidence interval of (6.3.7) on $\mu_{[i]}$ which has minimal probability of coverage γ has maximal probability of coverage $1-\left[1-\gamma^{\frac{1}{k-i+1}}\right]^i$ $(i = 1,\ldots,k;\ 0 < \gamma < 1)$.

The lower confidence interval of (6.3.7) on $\mu_{[i]}$ which has minimal probability of coverage γ has maximal probability of coverage $1-\left[1-\gamma^{\frac{1}{i}}\right]^{k-i+1}$ $(i = 1,\ldots,k;\ 0 < \gamma < 1)$.

The proof of Theorem (6.3.8) is similar to that of Theorem (6.3.7) and will be omitted. Note that (6.3.7) and (6.3.8) also hold when $k = 1$, in which case the upper and lower intervals on $\mu_{[1]}$ are exact. The following table illustrates the maximal degree of overprotection.

Table 6.3.9. $1-\left[1-\gamma^{\frac{1}{k-i+1}}\right]^i$

k	1	2		3		
γ	i = 1	i = 1	i = 2	i = 1	i = 2	i = 3
.99	.99	.995	1.000	.997	1.000	1.000
.95	.95	.975	.998	.983	.999	1.000
.90	.90	.949	.99	.965	.997	.999
.80	.80	.894	.96	.923	.989	.992
.70	.70	.837	.91	.888	.973	.973
.60	.60	.775	.84	.843	.949	.936
.50	.50	.707	.75	.794	.914	.875

For the special case i = k, Fraser (1952), p. 579, gave the upper interval on $\mu_{[k]}$ of Theorem (6.3.7) as one with probability of coverage at least γ. Fraser proves that under mild conditions an upper confidence interval for $\mu_{[k]}$ $(k \geq 2)$, with probability of coverage γ $(0 < \gamma < 1)$ for all $\mu \in \Omega_o$, does not exist.

Our results above extend to certain location parameter families if, instead of set-up (1.3.1) (normal distributions), we take set-up (2.1.1) with assumption (2.1.2) (a location parameter family with finite mean).

THEOREM: Suppose we have location parameter populations as in (2.1.1) and assumption (2.1.2) holds. For any i $(1 \leq i \leq k)$, if $a_i = 1$ (thus $a_j = 0$ for $j \neq i$) then the risk (6.3.5) ((6.3.6)) is the probability that our upper (lower) interval does not cover $\mu_{[i]}$ and is maximized over $\mu \in \Omega_o(\mu_{[i]})$ at

(6.3.10) $$\mu = (\underbrace{-\infty,\ldots,-\infty}_{i-1\text{ terms}}, \underbrace{\mu_{[i]},\ldots,\mu_{[i]}}_{k-i+1\text{ terms}}) \quad (\mu = (\underbrace{\mu_{[i]},\ldots,\mu_{[i]}}_{i\text{ terms}}, \underbrace{+\infty,\ldots,+\infty}_{k-i\text{ terms}})).$$

Thus, for any γ $(0 < \gamma < 1)$ an upper (lower) confidence interval of minimal probability of coverage γ is $(-\infty, \overline{X}_{[i]}+h_i^*)$ $((\overline{X}_{[i]}-g_i^*,+\infty))$ with $h_i^* = -G_n^{-1}(1-\gamma^{\frac{1}{k-i+1}}|f) + E_f$ $(g_i^* = G_n^{-1}(\gamma^{\frac{1}{i}}|f) - E_f)$. If $g_n(x|f)$ is symmetric about $x = 0$ this becomes $h_i^* = G_n^{-1}(\gamma^{\frac{1}{k-i+1}}|f) + E_f$ $(g_i^* = G_n^{-1}(\gamma^{\frac{1}{i}}|f) - E_f)$.

<u>Proof</u>: <u>Upper Interval</u>. For any i $(1 \le i \le k)$, if $a_i = 1$, $a_j = 0$ $(j \ne i)$ then by Theorem (2.1.11)

$$\sup_{\mu\in\Omega_o(\mu_{[i]})} r(\mu;I_{N,U}) = \sup_{\mu\in\Omega_o(\mu_{[i]})} P_\mu[\overline{X}_{[i]} \le \mu_{[i]}-h_i^*]$$

$$= \lim_{M\to+\infty} P_{\mu_{[1]}=\ldots=\mu_{[i-1]}=-M,\mu_{[i]}=\ldots=\mu_{[k]}}[\overline{X}_{[i]} \le \mu_{[i]}-h_i^*]$$

$$= \lim_{M\to+\infty} H_M(\mu_{[i]}-h_i^*),$$

where $H_M(x) = P_\mu[\overline{X}_{[i]}\le x]$ with $\mu = (-M,\ldots,-M,\mu_{[i]},\ldots,\mu_{[i]})$. Now $H_M(x)\to H_\infty(x)$ for all x by the expression for $F_{\overline{X}_{[i]}}(x)$ given in the proof of Lemma (2.2.5). Thus

$$\sup_{\mu\in\Omega_o(\mu_{[i]})} r(\mu;I_{N,U}) = P_{\mu_{[1]}=\ldots=\mu_{[i-1]}=-\infty,\mu_{[i]}=\ldots=\mu_{[k]}}[\overline{X}_{[i]} \le \mu_{[i]}-h_i^*]$$

$$= P[\min(Y_1,\ldots,Y_{k-i+1}) \le -h_i^*+E_f]$$

where $Y_1,\ldots,Y_{k-i+1}$ are (see (2.1.7)) $k-i+1$ independent r.v.'s each with d.f. $G_n(y|f)$. It follows that to make the minimal probability of coverage γ $(0 < \gamma < 1)$ we set h_i^* so that $h_i^* = -G_n^{-1}(1-\gamma^{\frac{1}{k-i+1}}|f) + E_f$.

Lower Interval. This case follows in a similar manner.

THEOREM: The upper confidence interval of (6.3.10) on $\mu_{[i]}$ which has minimal probability of coverage γ has maximal probability of coverage $1-\left[1-\gamma^{\frac{1}{k-i+1}}\right]^i$ $(i = 1,\ldots,k;\ 0 < \gamma < 1)$. (6.3.11)

The lower confidence interval of (6.3.10) on $\mu_{[i]}$ which has minimal probability of coverage γ has maximal probability of coverage $1-\left[1-\gamma^{\frac{1}{i}}\right]^{k-i+1}$ $(i = 1,\ldots,k;\ 0 < \gamma < 1)$.

The proof of Theorem (6.3.11) will be omitted. Note that this result implies that Table 6.3.9 provides an analysis of maximal overprotection for our location parameter case as well as for the normal case. For the special case $i = k$, Fraser (1952), p. 576, gave the upper interval on $\mu_{[k]}$ of Theorem (6.3.10) as one with probability of coverage at least γ. Fraser proves that under mild conditions an upper confidence interval for $\mu_{[k]}$ $(k \geq 2)$, with probability of coverage γ $(0 < \gamma < 1)$ for all $\mu \in \Omega_0$, does not exist if $f(x-\mu)$ satisfies a condition of bounded completeness. We will now extend this result to $\mu_{[i]}$ $(1 \leq i \leq k;\ k \geq 2)$; our mild conditions are slightly stronger than Fraser's.

DEFINITION: For $1 \leq i \leq k$, let $g_i(x_1,\ldots,x_k)$ be a real-valued function such that for any j $(1 \leq j \leq k)$

(6.3.12) $g_i(x_1,\ldots,x_k) \le g_i(x_1,\ldots,x_{j-1},x_j+\delta,x_{j+1},\ldots,x_k)$

for all $x_1,\ldots,x_k \in \mathcal{R}$ and $\delta > 0$.

DEFINITION: For $1 \le i \le k$, let

(6.3.13) $$\phi_{\theta,i}(x_1,\ldots,x_k) = \begin{cases} 1 \text{ if } g_i(x_1,\ldots,x_k) \ge \theta \\ 0 \text{ if } g_i(x_1,\ldots,x_k) < \theta. \end{cases}$$

DEFINITION: For any i $(1 \le i \le k)$ for $\ell = 1,2,\ldots$ let

(6.3.14) $$R_i(y_1,\ldots,y_\ell) = \begin{cases} \text{the } \underline{ith} \text{ smallest of } y_1,\ldots,y_\ell & \text{if } \ell \ge i \\ +\infty & \text{if } \ell < i. \end{cases}$$

Let $R_o(y_1,\ldots,y_\ell) = -\infty$ if $\ell \ge 1$.

DEFINITION: For $1 \le \ell \le k$, let

(6.3.15) $$S_\ell = \{(x_1,\ldots,x_k)\colon R_i(x_j,j\neq\ell) > x_\ell > R_{i-1}(x_j,j\neq\ell)\}.$$

Note that $\phi_{\theta,i}(x_1,\ldots,x_k)$ is a monotone non-decreasing function of $x_1,\ldots,x_k$ and that $S_1,\ldots,S_k$ are disjoint sets whose union is $\mathcal{R}^k$.

ASSUMPTION: $G_n(y-\theta|f)$ is boundedly complete (each-sided),

(6.3.16) i.e. $Eg(X) = \int_{-\infty}^{\infty} g(x)dG_n(x-\theta|f) = 0$ for a dense set of $\theta(<0$ or $>0)$ and $|g(x)| < M$ imply $g(x) = 0$ (a.e.).

THEOREM: Suppose we have location parameter populations as in (2.1.1) and assumptions (2.1.2) and (6.3.16) hold. Fix

(6.3.17) i $(1 \le i \le k;\ k \ge 2)$. Then an upper confidence interval for $\mu_{[i]}$, with probability of covering γ $(0 < \gamma < 1)$ for all $\mu \in \Omega_o$, and satisfying (6.3.12), does not exist.

Proof: Assume that $g_i(x_1,\ldots,x_k)$ satisfies (6.3.12) and yields an upper confidence interval for $\mu_{[i]}$ with probability of covering γ $(0 < \gamma < 1)$ for all $\mu \in \Omega_o$. We have

$$\gamma \overset{\mu}{\equiv} P_\mu[\mu_{[i]} \leq g_i(\overline{X}_1,\ldots,\overline{X}_k)] = E\phi_{\mu_{[i]},i}(\overline{X}_1,\ldots,\overline{X}_k)$$

$$= E\int_{-\infty}^{\infty}\phi_{\mu_{[i]},i}(\overline{X}_1,\ldots,\overline{X}_{\ell-1},x_\ell,\overline{X}_{\ell+1},\ldots,\overline{X}_k)dG_n(x_\ell-\mu_{[i]}+E_f|f) \text{ if } \mu_\ell = \mu_{[i]}$$

$$= E[\beta^{(\ell)}_{\mu_{[i]},i}(\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k)] \quad \text{if } \mu_\ell = \mu_{[i]}$$

where

$$\beta^{(\ell)}_{\mu_{[i]},i}(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) = \int_{-\infty}^{\infty}\phi_{\mu_{[i]},i}(x_1,\ldots,x_k)dG_n(x_\ell-\mu_{[i]}+E_f|f).$$

We now derive conditions on the function $\beta^{(\ell)}_{o,i}$. From the expression above it is seen that (if $\mu_\ell = \mu_{[i]}$)

$$E[\beta^{(\ell)}_{o,i}(\overline{X}_1,\ldots,\overline{X}_{\ell-1},\overline{X}_{\ell+1},\ldots,\overline{X}_k) - \gamma] = 0.$$

Hence, as in Fraser (1952), p. 580,

$$\beta^{(\ell)}_{o,i}(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) = \gamma \quad \text{(a.e.)}.$$

Using the above condition on $\beta^{(\ell)}_{o,i}$, we obtain conditions on the function $\phi_{o,i}(x_1,\ldots,x_k)$.

$$\gamma = \beta^{(\ell)}_{o,i}(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) \quad \text{(a.e.)}$$

$$= \int_{-\infty}^{\infty}\phi_{o,i}(x_1,\ldots,x_k)dG_n(x_\ell+E_f|f).$$

Consider fixed $x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k$. Now $\phi_{o,i}(x_1,\ldots,x_k)$ is a monotone function of x_ℓ, and since it is a characteristic function it will have the following form:

$$u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) = \max\left(\begin{array}{l}\text{value of } x_\ell \text{ at which} \\ \phi_{o,i}(x_1,\ldots,x_k) \text{ jumps} \\ \text{from 0 to 1}\end{array}, R_{i-1}(x_j,j\neq\ell)\right)$$

$$\phi_{o,i}(x_1,\ldots,x_k) = \begin{cases} 0 \text{ if } R_{i-1}(x_j,j\neq\ell) < x_\ell < u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) \\ 1 \text{ if } u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) < x_\ell < \infty. \end{cases}$$

Using the function $u(x_1,\ldots,x_k)$ we obtain

$$\gamma = \int_{-\infty}^{R_{i-1}(x_j,j\neq\ell)} \phi_{o,i}(x_1,\ldots,x_k)dG_n(x_\ell+E_f|f) + \int_{u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)}^{\infty} dG_n(x_\ell+E_f|f). \tag{6.3.18}$$

However, since

$$0 \leq \int_{-\infty}^{R_{i-1}(x_j,j\neq\ell)} \phi_{o,i}(x_1,\ldots,x_k)dG_n(x_\ell+E_f|f)$$

$$\leq \int_{-\infty}^{R_{i-1}(x_j,j\neq\ell)} dG_n(x_\ell+E_f|f) = P[\overline{X}_\ell \leq R_{i-1}(x_j,j\neq\ell) + E_f]$$

then

$$G_n^{-1}(1-\gamma|f) - E_f \leq u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)$$

$$\leq G_n^{-1}(1-\gamma+P[\overline{X}_\ell \leq R_{i-1}(x_j,j\neq\ell)+E_f]) - E_f.$$

The inequality on $u(x_1,\ldots,x_k)$ implies that $\phi_{o,i}(x_1,\ldots,x_k) = 0$ for $(x_1,\ldots,x_k) \in S_\ell$ with $x_\ell < G_n^{-1}(1-\gamma|f) - E_f$. This is true for $\ell = 1,\ldots,k$; hence $\phi_{o,i}(x_1,\ldots,x_k) = 0$ if $R_i(x_j) < G_n^{-1}(1-\gamma|f) - E_f$. Consider now

$(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)$ having $R_{i-1}(x_j, j\neq\ell) < G_n^{-1}(1-\gamma|f)$; in (6.3.18), the first integral vanishes leaving

$$\gamma = \int_{u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)}^{\infty} dG_n(x_\ell + E_f|f).$$

Therefore $u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k) = G_n^{-1}(1-\gamma|f) - E_f$, if $R_{i-1}(x_j, j\neq\ell) < G_n^{-1}(1-\gamma|f)$. From this equality on $u(x_1,\ldots,x_{\ell-1},x_{\ell+1},\ldots,x_k)$, we obtain the following conditions on $\phi_{o,i}(x_1,\ldots,x_k)$:

$$\phi_{o,i}(x_1,\ldots,x_k) = \begin{cases} 0 & \text{if } \quad R_i(x_j) < G_n^{-1}(1-\gamma|f) - E_f \\ 1 & \text{if } \begin{cases} R_i(x_j) > G_n^{-1}(1-\gamma|f) - E_f \\ R_{i-1}(x_j) < G_n^{-1}(1-\gamma|f). \end{cases} \end{cases}$$

But since $\phi_{o,i}(x_1+\delta,\ldots,x_k+\delta)$ is monotone in δ, we have

$$\phi_{o,i}(x_1,\ldots,x_k) = \begin{cases} 0 & \text{if } R_i(x_j) < G_n^{-1}(1-\gamma|f) - E_f \\ 1 & \text{if } R_i(x_j) > G_n^{-1}(1-\gamma|f) - E_f. \end{cases}$$

Therefore

$$g_i(x_1,\ldots,x_k) \begin{matrix} < 0 & \text{if} & R_i(x_j) < G_n^{-1}(1-\gamma|f) - E_f \\ \geq 0 & \text{if} & R_i(x_j) > G_n^{-1}(1-\gamma|f) - E_f. \end{matrix}$$

Similarly

$$g_i(x_1,\ldots,x_k) \begin{matrix} < \mu_{[i]} & \text{if} & R_i(x_j) < G_n^{-1}(1-\gamma|f) + \mu_{[i]} - E_f \\ \geq \mu_{[i]} & \text{if} & R_i(x_j) > G_n^{-1}(1-\gamma|f) + \mu_{[i]} - E_f. \end{matrix}$$

This completely determines $g_i(x_1,\ldots,x_k)$: $g_i(x_1,\ldots,x_k) = R_i(x_j) - G_n^{-1}(1-\gamma|f) + E_f$. But we know that a constant added to this yields

$\overline{X}_{[i]} - G_n^{-1}(1-\gamma^{\frac{1}{k-i+1}}|f) + E_f$, which doesn't always yield γ; therefore this can't.

Note that the argument of Fraser (1952), p. 580 (top) showing that the interval for $\mu_{[k]}$ generated by his proof has coverage <u>at least</u> γ doesn't extend to our case, since although

$$\{R_{k-1}(x_j, j\neq \ell) \leq A\} \Rightarrow \{R_k(x_j) > A \quad \text{iff} \quad x_\ell > A\},$$

$$\{R_{i-1}(x_j, j\neq \ell) \leq A\} \not\Rightarrow \{R_i(x_j) > A \quad \text{iff} \quad x_\ell > A\}.$$

Note that (if we wish to consider location parameters and not means) restriction (2.1.2) can be dropped throughout this section and the results stated in terms of $\theta_{[1]},\ldots,\theta_{[k]}$.

APPENDIX A. MAXIMA AND MINIMA OF REAL-VALUED FUNCTIONS OF n REAL VARIABLES

A-1. $n = 2$

Although the case $n = 2$ is included in the case $n \geq 2$ of Section A-2, it will be convenient to have stated separately the results and notations of this special case. (Note that some authors, e.g. Kaplan (1952), p. 126, state these results in a somewhat more cumbersome manner.)

<u>THEOREM</u>: Let f have continuous second-order partial derivatives on an open set S in R^2. Let $(x_1^o, x_2^o) \in S$ be such that

$$\left.\frac{\partial f(x_1,x_2)}{\partial x_1}\right|_{(x_1^o,x_2^o)} = \left.\frac{\partial f(x_1,x_2)}{\partial x_2}\right|_{(x_1^o,x_2^o)} = 0,$$

and let

$$A = \left.\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2}\right|_{(x_1^o,x_2^o)}$$

(A.1.1) $$B = \left.\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2}\right|_{(x_1^o,x_2^o)}$$

$$C = \left.\frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}\right|_{(x_1^o,x_2^o)}.$$

Then (x_1^o, x_2^o) is

(i) a relative minimum if $B^2-AC < 0$, $A > 0$;

(ii) a relative maximum if $B^2-AC < 0$, $A < 0$;

(iii) of undecided nature if $B^2-AC = 0$; and

(iv) a saddle point if $B^2-AC > 0$.

APPENDIX A. MAXIMA AND MINIMA OF REAL-VALUED FUNCTIONS OF n REAL VARIABLES

A-2. $n \geq 2$

Even in Hancock (1960) and Apostol (1957) the presentation of the theory of maxima and minima is not as complete as we need (e.g., in order to show in total the asymptotic nature of $(\overline{X},\ldots,\overline{X})$ in Section 5.1). We therefore present a summary gathered from several sources.

<u>THEOREM</u>: Let f have continuous second-order partial derivatives on an open set S in R^n. Let $(x_1^o,\ldots,x_n^o) \in S$ be such that

$$\left.\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i}\right|_{(x_1^o,\ldots,x_n^o)} = 0 \qquad (i = 1,\ldots,n),$$

and let $Q = (d_{ij})$ where

$$d_{ij} = \left.\frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_i \partial x_j}\right|_{(x_1^o,\ldots,x_n^o)} \qquad (i,j = 1,\ldots,n).$$

(A.2.1) Then the real symmetric matrix Q is either

(i) positive definite, in which case $(x_1^o,\ldots,x_n^o)$ is a relative minimum;

(ii) negative definite, in which case $(x_1^o,\ldots,x_n^o)$ is a relative maximum;

(iii) semi-definite, in which case the nature of $(x_1^o,\ldots,x_n^o)$ is undecided; or

(iv) indefinite, in which case $(x_1^o,\ldots,x_n^o)$ is a saddle point.

Proof: In addition to previously-cited references, see Courant (1966), pp. 204-208.

THEOREM: A real symmetric matrix Q, having eigenvalues $\lambda_1,\ldots,\lambda_n$ (say) is

(A.2.2)

(i) positive definite iff $\lambda_i > 0$ $(i = 1,\ldots,n)$;

(ii) negative definite iff $\lambda_i < 0$ $(i = 1,\ldots,n)$;

(iii)(a) positive semi-definite iff $\lambda_i \geq 0$ $(i = 1,\ldots,n)$ and at least one $\lambda_j = 0$;

(b) negative semi-definite iff $\lambda_i \leq 0$ $(i = 1,\ldots,n)$ and at least one $\lambda_j = 0$; and

(iv) indefinite iff at least one λ_i is positive and at least one λ_j is negative.

Proof: Recall that the eigenvalues of a matrix Q are the n roots of the equation $|Q-\lambda I| = 0$, and see Wedderburn (1964), p. 92.

THEOREM: For the real symmetric matrix Q, let $\Delta = \det(Q)$ and $\Delta_o = 1$. Let Δ_{n-t} be the determinant of Q with its last t rows and columns deleted. (Note that $\Delta_n = \Delta$.) Then Q is

(A.2.3)

(i) positive definite iff $\Delta_o,\Delta_1,\ldots,\Delta_n$ are positive;

(ii) negative definite iff $\Delta_o,\Delta_1,\ldots,\Delta_n$ are alternately positive and negative;

(iii)(a) positive semi-definite iff all principal minors of Q are ≥ 0 and $\Delta = 0$;

(b) negative semi-definite iff all principal minors of Q are $\geq 0 (\leq 0)$ if their order is even (odd), and $\Delta = 0$; or

(iv) indefinite, otherwise.

Proof: For (i) and (ii), see (e.g.) Narayan (1962), pp. 165, 167. (Note that the reference cited by Apostol (1957) is inadequate; it proves a weaker theorem which utilizes more than the leading principal minors of Q.)

For (iii)(a), from Browne (1958), we know Q is positive semi-definite iff all principal minors of Q are ≥ 0 (see pp. 120-121, Theorem 46.5). If Q is to be positive semi-definite but not definite, then the condition should also specify $\Delta = 0$. (This modification holds for the $\Rightarrow$ implication by the well-known result $\Delta = \lambda_1 \ldots \lambda_n$, e.g. Faddeeva (1959), p. 14. The $\Leftarrow$ implication is clear.) We use, of course, Theorem (A.2.2).

For (iii)(b), note that for any matrix A of order i, $\det(-A) = (-1)^i \det(A)$, and that Q is negative semi-definite iff $-Q$ is positive semi-definite.

Note. A condition such as "$\Delta_o, \Delta_1, \ldots, \Delta_n \geq 0$ and $\Delta = 0$" will not suffice for (iii)(a) of Theorem (A.2.3). For example, consider

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note. If $n = 2$, $Q = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$, $\Delta = AC - B^2$, $\Delta_1 = A$, $\Delta_2 = \Delta$ and Q is

(i) positive definite iff $B^2\text{-}AC < 0$, $A > 0$;

(ii) negative definite iff $B^2\text{-}AC < 0$, $A < 0$;

(iii)(a) positive semi-definite iff $B^2\text{-}AC = 0$, $A \geq 0$, $C \geq 0$;

(b) negative semi-definite iff $B^2\text{-}AC = 0$, $A \leq 0$, $C \leq 0$; and

(iv) indefinite iff $\{B^2\text{-}AC = 0,\ A > 0,\ C < 0\}$ or $\{B^2\text{-}AC = 0,\ A < 0,\ C > 0\}$ or $\{B^2\text{-}AC > 0\}$.

Here, we have reduced the number of undecided cases ((iii) cases) "beyond" those, namely $B^2\text{-}AC = 0$, named in virtually all texts. (The cases separated out belong to (iv) and are therefore saddle points.) However, by a consideration of signs it is easy to see that the sets $\{B^2\text{-}AC = 0,\ A > 0,\ C < 0\}$ and $\{B^2\text{-}AC = 0,\ A < 0,\ C > 0\}$ are empty. (The reason for this is the need to have at least one positive and one negative eigenvalue, thus exhausting the supply of eigenvalues when $n = 2$.)

APPENDIX B. DISTRIBUTIONS OF VARIOUS FUNCTIONS OF CERTAIN RANDOM VARIABLES

B-1. JOINT DISTRIBUTION OF $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$

The joint density of $\overline{X}_1,\ldots,\overline{X}_k$ is

$$f_{\overline{X}_1,\ldots,\overline{X}_k}(y_1,\ldots,y_k)=f_{\overline{X}_1}(y_1)\ldots f_{\overline{X}_k}(y_k) \qquad (y_i \in R;\ i = 1,\ldots,k)$$

where $f_{\overline{X}_i}(\circ)$ is the $N(\mu_i,\sigma^2/n)$ density function $(i = 1,\ldots,k)$ (see (5.1.1)). It is well-known that then the joint density of the ordered $\overline{X}_i$ $(i = 1,\ldots,k)$, i.e. of $\overline{X}_{[1]}\leq\ldots\leq\overline{X}_{[k]}$, is

$$f_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)$$

$$= \begin{cases} \sum\limits_{\beta\varepsilon S_k} f_{\overline{X}_1,\ldots,\overline{X}_k}(x_{\beta(1)},\ldots,x_{\beta(k)}), & x_1\leq\ldots\leq x_k \\ 0 & \text{, otherwise} \end{cases}$$

(B.1.1) $$= \begin{cases} \sum\limits_{\beta\varepsilon S_k} (\sqrt{n}/\sigma)^k \phi\left(\dfrac{x_{\beta(1)}-\mu_1}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\dfrac{x_{\beta(k)}-\mu_k}{\sigma/\sqrt{n}}\right), & x_1\leq\ldots\leq x_k \\ 0 & \text{, otherwise} \end{cases}$$

$$= \begin{cases} \sum\limits_{\beta\varepsilon S_k} (\sqrt{n}/\sigma)^k \phi\left(\dfrac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\ldots\phi\left(\dfrac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right), & x_1\leq\ldots\leq x_k \\ 0 & \text{, otherwise.} \end{cases}$$

APPENDIX B. DISTRIBUTIONS OF VARIOUS FUNCTIONS OF CERTAIN RANDOM VARIABLES

B-2. LIMIT DISTRIBUTION OF $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$

The limiting distribution of $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$ (under certain parameter configurations) is of interest to us. Let $\{A_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ be sequences of events on some probability space (which may depend on n). Let $a = (a_1,\ldots,a_k) \in R^k$ be fixed, and denote the vector $(\mu_1+a_1\sigma/\sqrt{n},\ldots,\mu_k+a_k\sigma/\sqrt{n})$ by $\mu + a\sigma/\sqrt{n}$.

(B.2.1) <u>LEMMA</u>: If $\lim_{n\to\infty} P_n(B_n) = 1$, then (if either of the following limits exists) $\lim_{n\to\infty} P_n(A_nB_n) = \lim_{n\to\infty} P_n(A_n)$.

<u>Proof</u>: Suppose $\lim_{n\to\infty} P_n(B_n) = 1$. Then by taking limits in $P_n(B_n) \leq P_n(A_n \cup B_n) \leq 1$ we find $\lim_{n\to\infty} P_n(A_n \cup B_n) = 1$, and hence $\lim_{n\to\infty} \{P_n(B_n) - P_n(A_n \cup B_n)\} = 0$. Taking limits in $P_n(A_nB_n) = P_n(A_n) + \{P_n(B_n) - P_n(A_n \cup B_n)\}$ yields our result.

(B.2.2) <u>DEFINITION</u>: For $\mu \in \Omega_o$, let $p(n|\mu) = P_\mu[\overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$, where $\overline{X}_{(1)},\ldots,\overline{X}_{(k)}$ are as in definitions (1.3.13) and (1.3.14).

<u>LEMMA</u>: Let $\Theta = \{\mu\colon \mu \in \Omega_o,\ \mu_1 = \mu_{[1]}, \ldots, \mu_k = \mu_{[k]}\}$. For all

(B.2.3) $\mu \in \Omega(\neq) \cap \Theta$,

$$\lim_{n\to\infty} p(n|\mu + a\sigma/\sqrt{n}) = \lim_{n\to\infty} P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(1)} < \ldots < \overline{X}_{(k)}] = 1.$$

<u>Proof</u>: 1. Suppose that $\mu \in \Omega(\neq) \cap \Theta$. Then for all n large enough, $\mu + a\sigma/\sqrt{n} \in \Omega(\neq) \cap \Theta$. Then the $\overline{X}_{(j)}$ are independent and $\overline{X}_{(j)}$ is the sample mean of n i.i.d. $N(\mu_{[j]} + a_j\sigma/\sqrt{n}, \sigma^2)$ r.v.'s. The characteristic function of a $N(m,\sigma^2)$ r.v. is (see, e.g., Parzen (1960), p. 221) $\varphi(t) = \exp\{itm - \frac{1}{2}t^2\sigma^2\}$. Thus,

$$\varphi_{\overline{X}_{(j)}}(t) = Ee^{it\overline{X}_{(j)}} = \left[e^{i\frac{t}{n}(\mu_{[j]} + a_j\sigma/\sqrt{n}) - \frac{1}{2}\frac{t^2}{n^2}\sigma^2}\right]^n$$

$$= e^{it\mu_{[j]}} e^{ita_j\frac{\sigma}{\sqrt{n}} - \frac{1}{2}\frac{t^2}{n}\sigma^2},$$

so that $\lim_{n\to\infty} \varphi_{\overline{X}_{(j)}}(t) = e^{it\mu_{[j]}}$. It is then well-known (see, e.g., Wilks (1962), p. 124, 5.4.1a) that $\overline{X}_{(j)}$ converges in probability to $\mu_{[j]}$ $(j = 1,\ldots,k)$. Thus, since the $\overline{X}_{(j)}$ are independent, it is clear that the probability that $\{\overline{X}_{(j)}$ converges to $\mu_{[j]}$ $(j = 1,\ldots,k)\}$ approaches 1 as $n\to\infty$. However, by Lemma (B.2.1)

$$\lim_{n\to\infty} P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$$

(B.2.4) $$= \lim_{n\to\infty} P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(1)} < \ldots < \overline{X}_{(k)},$$

$$|\overline{X}_{(1)} - \mu_{[1]}| < \varepsilon, \ \ldots, |\overline{X}_{(k)} - \mu_{[k]}| < \varepsilon]$$

for any $\varepsilon > 0$. If we choose $2\varepsilon \leq \min_{1 \leq i < j \leq k} (\mu_{[j]} - \mu_{[i]})$, then the r.h.s. of (B.2.4) equals

$$\lim_{n\to\infty} P_{\mu + a\sigma/\sqrt{n}}[|\overline{X}_{(1)} - \mu_{[1]}| < \varepsilon \ , \ \ldots \ , |\overline{X}_{(k)} - \mu_{[k]}| < \varepsilon],$$

which is 1 since $P[\overline{X}_{(j)}$ converges to $\mu_{[j]}$ $(j = 1,\ldots,k)]$ approaches 1 as $n\to\infty$.

2. Suppose that $\mu \ \varepsilon [\Omega(\neq)]^c \cap \Theta$. (Eventually $\mu + a\sigma/\sqrt{n} \ \varepsilon \ \Theta \cap \Omega(\neq)$, or $\Theta \cap [\Omega(\neq)]^c$.) Then there are ℓ distinct values in $\{\mu_{[1]} + a_1\sigma/\sqrt{n}, \ldots, \mu_{[k]} + a_k\sigma/\sqrt{n}\}$ $(1 \leq \ell \leq k-1)$ and (see (1.3.14))

$$P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$$

$$= P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(i_1)} < \overline{X}_{(i_1+1)}, \ \overline{X}_{(i_2)} < \overline{X}_{(i_2+1)}, \ \ldots \ , \overline{X}_{(i_{\ell-1})} < \overline{X}_{(i_{\ell-1}+1)}].$$

However, the result will <u>not</u> follow as before since $\min_{1 \leq i < j \leq k} (\mu_{[j]} - \mu_{[i]}) = 0$ here. It can be seen (e.g., consider the case $k = 2$) that the limit $\not\to 1$ as $n\to\infty$. (In fact, it depends on a.)

<u>LEMMA</u>: For $\mu \ \varepsilon \ \Theta \cap \Omega(\neq)$, as $n\to\infty$

(B.2.5)
$$F^{(\mu + a\sigma/\sqrt{n})}_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)$$

$$- P_{\mu + a\sigma/\sqrt{n}}[\overline{X}_{(i)} \leq x_i \ (i = 1,\ldots,k)] \to 0.$$

<u>Proof</u>:

$$\lim_{n\to\infty} F^{(\mu + a\sigma/\sqrt{n})}_{\overline{X}_{[1]},\ldots,\overline{X}_{[k]}}(x_1,\ldots,x_k)$$

$$= \lim_{n\to\infty}\{p(n|\mu+a\sigma/\sqrt{n})\cdot$$

$$\cdot P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{[1]} \leq x_1, \ldots ,\overline{X}_{[k]} \leq x_k \mid \overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$$

$$+ (1-p(n|\mu+a\sigma/\sqrt{n}))\cdot$$

$$\cdot P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{[1]} \leq x_1, \ldots ,\overline{X}_{[k]} \leq x_k \mid \text{not } (\overline{X}_{(1)} < \ldots < \overline{X}_{(k)})]\}$$

$$= \lim_{n\to\infty} P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{[1]} \leq x_1, \ldots ,\overline{X}_{[k]} \leq x_k ; \overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$$

$$= \lim_{n\to\infty} P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k ; \overline{X}_{(1)} < \ldots < \overline{X}_{(k)}]$$

$$= \lim_{n\to\infty} P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k].$$

Here the second equality follows from Lemma (B.2.3), while the last equality follows from Lemmas (B.2.3) and (B.2.1).

LEMMA: As $n\to\infty$, if $\mu+a\sigma/\sqrt{n} \in \Theta\cap\Omega(\neq)$ then

(B.2.6) $$P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k] \rightarrow P_{\mu}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k].$$

Proof: As $n\to\infty$,

$$P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k]$$

$$= P_{\mu+a\sigma/\sqrt{n}}[\overline{X}_{(1)}-a_1\sigma/\sqrt{n} \leq x_1-a_1\sigma/\sqrt{n}, \ldots ,\overline{X}_{(k)}-a_k\sigma/\sqrt{n} \leq x_k-a_k\sigma/\sqrt{n}]$$

$$= P_{\mu}[\overline{X}_{(1)} \leq x_1-a_1\sigma/\sqrt{n}, \ldots ,\overline{X}_{(k)} \leq x_k-a_k\sigma/\sqrt{n}]$$

$$\rightarrow P_{\mu}[\overline{X}_{(1)} \leq x_1, \ldots ,\overline{X}_{(k)} \leq x_k].$$

The second equality follows because, when $\mu+a\sigma/\sqrt{n} \in \Theta\cap\Omega(\neq)$, $\overline{X}_{(i)}$ is $N(\mu_{[i]}+a_i\sigma/\sqrt{n}, \sigma^2/n)$ iff $\overline{X}_{(i)}-a_i\sigma/\sqrt{n}$ is $N(\mu_{[i]},\sigma^2/n)$ $(i=1,\ldots,k)$.

(B.2.7) DEFINITION: Let $\Phi(z_1,\ldots,z_s)$ denote the d.f. of the $1,\ldots,s$ order statistics in a sample of size s from a $N(0,1)$ population.

THEOREM: As $n\to\infty$, if $\mu \in \Theta\cap\Omega(\neq)$ then

(B.2.8)
$$F^{(\mu+a\sigma/\sqrt{n})}_{\frac{\sqrt{n}}{\sigma}(\bar{X}_{[1]}-\mu_{[1]}-a_1\sigma/\sqrt{n}),\ldots,\frac{\sqrt{n}}{\sigma}(\bar{X}_{[k]}-\mu_{[k]}-a_k\sigma/\sqrt{n})}(x_1,\ldots,x_k)$$
$$\to \prod_{i=1}^{k}\Phi(x_i).$$

Proof: This follows from Lemmas (B.2.5) and (B.2.6).

COROLLARY: As $n\to\infty$, if $\mu \in \Theta\cap\Omega(\neq)$ then

(B.2.9)
$$F_{\frac{\sqrt{n}}{\sigma}(\bar{X}_{[1]}-\mu_{[1]}),\ldots,\frac{\sqrt{n}}{\sigma}(\bar{X}_{[k]}-\mu_{[k]})}(x_1,\ldots,x_k) \to \prod_{i=1}^{k}\Phi(x_i).$$

THEOREM: If $\mu \in \Theta\cap[\Omega(\neq)]^c$ then

(B.2.10)
$$\lim_{n\to\infty} F^{(\mu+a\sigma/\sqrt{n})}_{\frac{\sqrt{n}}{\sigma}(\bar{X}_{[1]}-\mu_{[1]}-a_1\sigma/\sqrt{n}),\ldots,\frac{\sqrt{n}}{\sigma}(\bar{X}_{[k]}-\mu_{[k]}-a_k\sigma/\sqrt{n})}(x_1,\ldots,x_k)$$

depends on a.

Proof: (A hint of this dependence was given in part 2 of the proof of Lemma (B.2.3).) Suppose $k = 2$, $a = (a_1,a_2)$ with $a_1 \leq a_2$, and let Y_1,Y_2 denote i.i.d. $N(0,1)$ r.v.'s. Then $\mu_{[1]} = \mu_{[2]}$ and

$$F^{(\mu+a\sigma/\sqrt{n})}_{\frac{\sqrt{n}}{\sigma}(\bar{X}_{[1]}-\mu_{[1]}-a_1\sigma/\sqrt{n}),\ \frac{\sqrt{n}}{\sigma}(\bar{X}_{[2]}-\mu_{[2]}-a_2\sigma/\sqrt{n})}(x_1,x_2)$$

$$= P_{\mu+a\sigma/\sqrt{n}}\left[\frac{\sqrt{n}}{\sigma}\left(\min(\overline{X}_1,\overline{X}_2)-\mu_{[1]}-a_1\sigma/\sqrt{n}\right) \leq x_1,\right.$$

$$\left.\frac{\sqrt{n}}{\sigma}(\max(\overline{X}_1,\overline{X}_2)-\mu_{[1]}-a_2\sigma/\sqrt{n}) \leq x_2\right]$$

$$= P[\min(Y_1,Y_2+(a_2-a_1)) \leq x_1,\ \max(Y_1-(a_2-a_1),Y_2) \leq x_2].$$

For $a_2 - a_1 = 0$, this is $\Phi(x_1,x_2)$. However, for $a_2 \gg a_1$ it is approximately $\Phi(x_1)\Phi(x_2)$, and therefore depends on a.

APPENDIX B. DISTRIBUTIONS OF VARIOUS FUNCTIONS OF

CERTAIN RANDOM VARIABLES

B-3. JOINT DISTRIBUTION OF $\overline{X}_{[k]}-\overline{X}_{[1]},\ldots,\overline{X}_{[k]}-\overline{X}_{[k-1]}$

From the joint density of $\overline{X}_{[1]},\ldots,\overline{X}_{[k]}$ given at (B.1.1), we find that (for $x_1 \leq x_2$)

$$f_{\overline{X}_{[1]},\overline{X}_{[2]}}(x_1,x_2)$$

$$= \frac{n}{2\pi\sigma^2}\left\{e^{-\frac{1}{2}\left[\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]}+e^{-\frac{1}{2}\left[\left(\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]}\right\},$$

so that (for $y \geq 0$), setting $\eta = \mu_{[2]} - \mu_{[1]}$,

$$f_{\overline{X}_{[2]}-\overline{X}_{[1]}}(y) = \int_{-\infty}^{\infty} f_{\overline{X}_{[1]},\overline{X}_{[2]}}(x,y+x)\,dx$$

$$= \int_{-\infty}^{\infty}\frac{n}{2\pi\sigma^2}\left\{e^{-\frac{1}{2}\left[\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{y+x-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]}\right.$$

$$\text{(B.3.1)}\qquad \left.+\,e^{-\frac{1}{2}\left[\left(\frac{x+y-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{x-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]}\right\}dx$$

$$= \int_{-\infty}^{\infty}\frac{1}{2\pi}\,\frac{\sqrt{n}}{\sigma}\left\{e^{-\frac{1}{2}\left[x^2+\left(\frac{y-\eta}{\sigma/\sqrt{n}}+x\right)^2\right]}\right.$$

$$\left.+\,e^{-\frac{1}{2}\left[\left(x+\frac{y}{\sigma/\sqrt{n}}\right)^2+\left(x-\frac{\eta}{\sigma/\sqrt{n}}\right)^2\right]}\right\}dx.$$

Since, via completing the square,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}[(x+a)^2+(x+b)^2]}\,dx = \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x+\frac{(a+b)}{2}}{1/\sqrt{2}}\right)^2} e^{-(a-b)^2/4}\,dx = \sqrt{\pi}\,e^{-\frac{1}{4}(a-b)^2},$$

it follows from (B.3.1) that

<u>THEOREM</u>: With $\eta = \mu_{[2]} - \mu_{[1]}$, for $y \geq 0$

(B.3.2)
$$f_{\overline{X}_{[2]}-\overline{X}_{[1]}}(y) = \frac{\sqrt{n}}{2\sigma\sqrt{\pi}}\left\{e^{-\frac{1}{4}\left(\frac{y-\eta}{\sigma/\sqrt{n}}\right)^2} + e^{-\frac{1}{4}\left(\frac{y+\eta}{\sigma/\sqrt{n}}\right)^2}\right\}.$$

REFERENCES

Abramowitz, M. and Stegun, I. A. (Editors): Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, U. S. Government Printing Office, Washington, D. C., June 1964.

Alam, K.: "A two-sample estimate of the largest mean," Annals of the Institute of Statistical Mathematics (Tokyo), Vol. 19, No. 2 (1967), pp. 271-283.

Apostol, T. M.: Mathematical Analysis, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1957 (Third Printing, December 1960).

Bahadur, R. R.: "Sufficiency and statistical decision functions," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 423-462.

Bahadur, R. R. and Goodman, L. A.: "Impartial decision rules and sufficient statistics," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 553-562.

Bechhofer, R. E.: "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 16-39.

Bechhofer, R. E.: "Design of experiments," a course offered at Cornell University, Ithaca, New York, Fall Term, 1964.

Bechhofer, R. E., Kiefer, J. and Sobel, M.: Sequential Identification and Ranking Procedures (with special reference to Koopman-Darmois populations), University of Chicago Press, Chicago, Illinois, 1968.

Bechhofer, R. E. and Sobel, M.: "Non-parametric multiple-decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observation (preliminary report)," Abstract, Annals of Mathematical Statistics, Vol. 29 (1958), p. 325.

Berk, R. H.: "Zehna, Peter W.. Invariance of maximum likelihood estimators," Review #1922, Mathematical Reviews, Vol. 33 (1967), pp. 342-343.

Blumenthal, S. and Cohen, A.: "Estimation of two ordered translation parameters," Annals of Mathematical Statistics, Vol. 39 (1968a), pp. 517-530.

Blumenthal, S. and Cohen, A.: "Estimation of the larger translation parameter," Annals of Mathematical Statistics, Vol. 39 (1968b), pp. 502-516.

Blumenthal, S. and Cohen, A.: "Estimation of the larger of two normal means," Journal of the American Statistical Association, Vol. 63 (1968), pp. 861-876.

Browne, E. T.: Introduction to the Theory of Determinants and Matrices, University of North Carolina Press, Chapel Hill, North Carolina, 1958.

Chambers, M. L. and Mack, C.: "Confidence limits for the minimum of two normal means; a new inference principle," New Journal of Statistics and Operational Research, Vol. 2 (1966), pp. 14-27.

Courant, R.: Differential and Integral Calculus, Vol. II (translated by E. J. McShane), Interscience Publishers, Inc., New York, 1936 (Reprinted 1966).

Dudewicz, E. J.: The Efficiency of a Nonparametric Selection Procedure: Largest Location Parameter Case, unpublished M.S. thesis, Cornell University, Ithaca, New York, February 1966. (Reprinted as Technical Report No. 14, Department of Operations Research, Cornell University, Ithaca, New York, December 1966.)

Dudewicz, E. J.: "The Robustness of a Selection Procedure of Bechhofer," Technical Report in preparation, Department of Operations Research, Cornell University, Ithaca, New York, 1968.

Eaton, M. L.: "Some optimum properties of ranking procedures," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 124-137.

Faddeeva, V. N.: Computational Methods of Linear Algebra, Dover Publications, Inc., New York, 1959.

Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. I (Second Edition), John Wiley & Sons, Inc., New York, 1957.

Ferguson, T. S.: Mathematical Statistics: A Decision Theoretic Approach, Academic Press Inc., New York, 1967.

Fisz, M.: Probability Theory and Mathematical Statistics (Third Edition), John Wiley & Sons, Inc., New York, 1963.

Fraser, D. A. S.: "Confidence bounds for a set of means," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 575-585.

Graybill, F. A.: An Introduction to Linear Statistical Models, Volume I, McGraw-Hill Book Company, Inc., New York, 1961.

Gupta, S. S.: "On a decision rule for a problem in ranking means," Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina, May 1956.

Gupta, S. S.: "On some multiple decision (selection and ranking) rules," Technometrics, Vol. 7 (1965), pp. 225-245.

Hall, W. J.: "Most economical multiple-decision rules," Annals of Mathematical Statistics, Vol. 29 (1958), pp. 1079-1094.

Hall, W. J.: "The most-economical character of some Bechhofer and Sobel decision rules," Annals of Mathematical Statistics, Vol. 30 (1959), pp. 964-969.

Hancock, H.: Theory of Maxima and Minima, Dover Publications, Inc., New York, 1960.

Harter, H. L.: "Expected values of normal order statistics," Biometrika, Vol. 48 (1961), pp. 151-165.

Hodgman, C. D. (Editor): C. R. C. Standard Mathematical Tables (Twelfth Edition), Chemical Rubber Publishing Company, Cleveland, Ohio, 1959.

Hogg, R. V. and Craig, A. T.: Introduction to Mathematical Statistics (Second Edition), The Macmillan Co., New York, 1965 (Second Printing 1965).

Kaplan, W.: Advanced Calculus, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1952 (Fifth Printing, July 1959).

Katz, M. W.: "Admissible and minimax estimates of parameters in truncated spaces," Annals of Mathematical Statistics, Vol. 32 (1961), pp. 136-142.

Katz, M. W.: "Estimating ordered parameters," Annals of Mathematical Statistics, Vol. 34 (1963), pp. 967-972.

Kendall, M. G. and Stuart, A.: The Advanced Theory of Statistics, Vol. 1: Distribution Theory (Second Edition), Hafner Publishing Co., New York, 1963.

Kiefer, J.: "Invariance, minimax sequential estimation, and continuous time processes," Annals of Mathematical Statistics, Vol. 28 (1957), pp. 573-601.

Lal Saxena, K. M. and Tong, Y. L.: "Interval estimation of the largest mean of k normal populations," Abstract, Annals of Mathematical Statistics, Vol. 39 (1968), pp. 704-705.

Lawton, W. H.: "Concentration of random quotients," Annals of Mathematical Statistics, Vol. 39 (1968), pp. 466-480.

Lehmann, E. L.: "Ordered families of distributions," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 399-419.

Lehmann, E. L.: "On a theorem of Bahadur and Goodman," Annals of Mathematical Statistics, Vol. 37 (1966), pp. 1-6.

Loève, M.: Probability Theory (Third Edition), D. Van Nostrand Co., Inc., Princeton, New Jersey, 1963.

Mahamunulu, D. M.: "Some fixed-sample ranking and selection problems," Annals of Mathematical Statistics, Vol. 38 (1967), pp. 1079-1091.

Narayan, S.: A Text Book of Matrices (Fourth Edition), S. Chand & Co., Delhi, India, 1962.

Parzen, E.: Modern Probability Theory and Its Applications, John Wiley & Sons, Inc., New York, 1960 (Fourth Printing, March 1963).

Paulson, E.: "A sequential procedure for selecting the population with the largest mean from k normal populations," Annals of Mathematical Statistics, Vol. 35 (1964), pp. 174-180.

Reitsma, A.: "On approximations to sampling distributions of the mean for samples from non-normal populations," Annals of Mathematical Statistics, Vol. 34 (1963), pp. 1308-1314.

Robertson, T. and Waltman, P.: "On estimating monotone parameters," Annals of Mathematical Statistics, Vol. 39 (1968), pp. 1030-1039.

Teichroew, D.: Probabilities Associated with Order Statistics in Samples from Two Normal Populations with Equal Variance, ENASR no. ES-3, Chemical Corps Engineering Agency, Engineering Statistics Unit, Army Chemical Center, Maryland, December 7, 1955.

Teichroew, D.: "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," Annals of Mathematical Statistics, Vol. 27 (1956), pp. 410-426.

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

Wadsworth, G. P. and Bryan, J. G.: Introduction to Probability and Random Variables, McGraw-Hill Book Company, Inc., New York, 1960.

Wedderburn, J. H. M.: Lectures on Matrices, Dover Publications, Inc., New York, 1964.

Weiss, L.: "On estimating scale and location parameters," Journal of the American Statistical Association, Vol. 58 (1963), pp. 658-659.

Weiss, L. and Wolfowitz, J.: "Generalized maximum likelihood estimators," Teorija Verojatnostei i ee Primenenija, Vol. 11 (1966), pp. 68-93.

Weiss, L. and Wolfowitz, J.: "Generalized maximum likelihood estimators in a particular case," Teorija Verojatnostei i ee Primenenija, Vol. 12 (1967a), to appear.

Weiss, L. and Wolfowitz, J.: "Maximum probability estimators," Annals of the Institute of Statistical Mathematics (Tokyo), Vol. 19, No. 2 (1967b), pp. 193-206.

Weiss, L. and Wolfowitz, J.: "Estimation of a density function at a point," Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, Vol. 7 (1967c), pp. 327-335.

Weiss, L. and Wolfowitz, J.: "Maximum probability estimators with a general loss function," to appear in the Proceedings of the International Symposium on Probability and Information Theory, held April 4-5 (1968) at McMaster University, Hamilton, Ontario, Canada.

Wilks, S. S.: Mathematical Statistics, John Wiley & Sons, Inc., New York, 1962 (Second Printing with Corrections, 1963).

Zehna, P. W.: "Invariance of maximum likelihood estimators," Annals of Mathematical Statistics, Vol. 37 (1966), p. 744.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Department of Operations Research College of Engineering, Cornell University Ithaca, New York 14850	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ESTIMATION OF ORDERED PARAMETERS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, February 1969

5. AUTHOR(S) (Last name, first name, initial)

Edward J. Dudewicz

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 1969	154	62

8a. CONTRACT OR GRANT NO. Nonr 401(53) b. ~~XXXXXXXXX~~ N00014-68A-0091 c. GP-7798 d. GP-8958	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 60 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. ~~SUPPLEMENTARY NOTES~~ Sponsoring military activity Logistics and Mathematical Statistics Branch, Office of Naval Research Washington, D.C. 20360	12. SPONSORING MILITARY ACTIVITY Center for Naval Analysis University of Rochester Rochester, New York 14627

13. ABSTRACT

Suppose given $k \geq 2$ normal populations $\pi_1,\ldots,\pi_k$; π_i has unknown mean μ_i and variance σ^2 $(i=1,\ldots,k)$. We assume throughout that $\mu_1,\ldots,\mu_k$ and the pairings of $\mu_{[1]}\leq\ldots\leq\mu_{[k]}$ with $\pi_1,\ldots,\pi_k$ are completely unknown (although we vary the distribution from normality) and consider the problem: estimate $\mu_{[1]},\ldots,\mu_{[k]}$ based on $\overline{X}_1,\ldots,\overline{X}_k$, where $\overline{X}_i$ is the average of n independent observations on π_i $(i = 1,\ldots,k)$. Applications to ranking and selection problems are noted.

$\overline{X}_{[i]}$, the ith smallest of $\overline{X}_1,\ldots,\overline{X}_k$, is a natural estimator of $\mu_{[i]}$ $(1 \leq i \leq k)$ and is studied with regard to bias, asymptotic unbiasedness, strong consistency, mean squared error, and minimax |bias| estimator of type $\overline{X}_{[i]} + a$. Results for the location parameter case are extended in the normal case. Maximum likelihood estimation, MLE's for non-1-1 functions, iterated MLE's, generalized MLE's, and maximum probability estimation are studied. Confidence intervals on $\mu_{[i]}$ $(1 \leq i \leq k)$ are found for location parameter populations.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification