stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 58/78 JUNI
(DEPARTMENT OF MATHEMATICAL STATISTICS)

C.H. KRAFT & C. VAN EEDEN

BOUNDS FOR THE MEAN AND STANDARD DEVIATION OF LINEAR COMBINATIONS OF ORDER STATISTICS

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): Primary 62G30; Secondary 62G35

Bounds for the mean and standard deviation of linear combinations of order statistics[1) *]

by

Charles H. Kraft & Constance van Eeden
Université de Montréal[2)]

SUMMARY

In this paper bounds are given for $E_\theta\ \Sigma_{s=1}^N\ \alpha_s O_s$ and $\sigma_\theta(\Sigma_{s=1}^N\ \alpha_s O_s)$, where $O_1,\ldots,O_N$ are the order statistics of $X_1,\ldots,X_N$, a random vector with density $p_0(x_1-\theta_1,\ldots,x_N-\theta_N)$ for some $\theta = (\theta_1,\ldots,\theta_N)$ and $p_0(x_1,\ldots,x_N)$ a completely arbitrary density with respect to Lebesgue measure in R_N. If the X_i are exchangeable under p_0 the bounds depend only on $E_0\ \Sigma_{s=1}^N\ \alpha_s O_s$ and $\sigma_0(\Sigma_{s=1}^N\ \alpha_s O_s)$ plus quantities that can be computed without knowledge of p_0. The bounds are attainable for some distributions and nearly attainable for some others.

KEY WORDS & PHRASES: *Order statistics, linear combinations of order statistics, bounds for means and standard deviations.*

1) Research supported by the Ministère de l'Education du Québec, Grant from F.C.A.C., by the National Research Council of Canada, Grants No. A3038 and No. A3114 plus supplemental travel grants and by the Mathematical Centre, Amsterdam.

2) This paper was written while the authors were on leave at the Mathematical Centre, Amsterdam.

* This report will be submitted for publication elswhere.

1. INTRODUCTION

Let $p_0(x) = p_0(x_1,\ldots,x_N)$ be a given density with respect to Lebesgue measure in R_N and consider the family $p_\theta(x) = p_0(x_1-\theta_1,\ldots,x_N-\theta_N)$, $\theta = (\theta_1,\ldots,\theta_N) \in R_N$. Let $(X_1,\ldots,X_N)$ be a random vector whose density is p_θ for some θ and let $O_1,\ldots,O_N$ be the order statistics of $X_1,\ldots,X_N$.

In this paper bounds are given for $E_\theta \Sigma_{s=1}^N \alpha_s O_s$ and $\sigma_\theta(\Sigma_{s=1}^N \alpha_s O_s)$. These bounds are of the form

$$(1.1) \qquad f_1(\theta,\alpha) \le E_\theta \sum_{s=1}^{N} \alpha_s O_s - E_0 \sum_{s=1}^{N} \alpha_s O_s \le f_2(\theta,\alpha)$$

and

$$(1.2) \qquad \sigma_\theta\Big(\sum_{s=1}^{N} \alpha_s O_s\Big) \le \sqrt{\sigma_0^2\Big(\sum_{s=1}^{N} \alpha_s O_s\Big) + f_3(\theta,\alpha) + f_4(\theta,\alpha)},$$

where $\alpha = (\alpha_1,\ldots,\alpha_N)$.

The bounds (1.1) require no other assumptions than those above; (1.2) requires a further condition; exchangeability under p_0 of the X_i, for example, is sufficient (see section 3 for the exact condition and also for a bound for $\sigma_\theta(\Sigma_{s=1}^N \alpha_s O_s)$ which uses only the assumptions for (1.1)).

With the condition of exchangeability of the X_i under p_0 the functions f_1, f_2, f_3 and f_4 can be computed without knowledge of p_0.

Examples will be given to show that the bounds are in some circumstances exactly attainable and in others almost attainable.

In what follows, section 2 contains the derivation of the bounds for the mean, section 3 the derivation of the bound for the standard deviation. The computation of the functions f_1, f_2, f_3 and f_4 when the X_i are exchangeable under p_0 is discussed in section 4. Distributions for which the bounds are attained as well as examples where they are almost attained are given in section 5; section 6 contains some comparisons between actual values and their bounds.

2. BOUNDS FOR THE EXPECTATION OF $\Sigma_{s=1}^{N} \alpha_s O_s$.

It will be helpful to consider first the expectation of a single order statistic, O_s.

Straightforward computation gives

$$E_\theta O_s = \sum_j \int_{T_j} x_{j_s} p_\Theta(x)dx,$$

where $T_j = \{x_{j_1} \le \dots \le x_{j_N}\}$ and $j = \{j_1,\dots,j_N\}$ is a permutation of $\{1,\dots$ $\dots,N\}$. Let $y_r = x_r - \theta_r$, $r = 1,\dots,N$, then

$$E_\theta O_s = \sum_j \int_{B_j} (y_{j_s} + \theta_{j_s}) p_0(y)dy$$

with $B_j = \{y_{j_1} + \theta_{j_1} \le \dots \le y_{j_N} + \theta_{j_N}\}$. With $A_i = \{y_{i_1} \le \dots \le y_{i_N}\}$, $i = \{i_1,\dots,i_N\}$ a permutation of $\{1,\dots,N\}$, we have

(2.1) $$E_\theta O_s = \sum_{i,j} \int_{A_iB_j} (y_{j_s} + \Theta_{j_s}) p_0(y)dy,$$

where i and j range independently through the N! permutations of 1,...,N.

Concerning the sets A_iB_j we have the following lemma.

LEMMA 2.1. *For each* (i,j) *there either exists* k *such that*

(i) $k \in \{i_s,\dots,i_N\}$

and

(ii) $y_{j_s} + \theta_{j_s} \ge y_{i_s} + \theta_k$ *for all* $y \in A_iB_j$

or A_iB_j *is empty*.

PROOF. Suppose there exists j such that $A_iB_j \ne \emptyset$ and for which there is no k with (i) and (ii). Then for every $k \in \{i_s,\dots,i_N\}$ we have

$$y_{j_s} + \theta_{j_s} < y_{i_s} + \theta_k \qquad \text{for some } y \in A_iB_j,$$

which implies, because

$$y_{i_s} \le y_k \quad \text{for all } k \in \{i_s,\dots,i_N\} \text{ and all } y \in A_iB_j,$$

that for all $k \in \{i_s,\dots,i_N\}$

$$y_{j_s} + \theta_{j_s} < y_k + \theta_k \quad \text{for some } y \in A_iB_j.$$

This means that, for all $k \in \{i_s,\dots,i_N\}$, k is strictly to the right of j_s in the j-permutation, but this is a contradiction because $\{i_s,\dots,i_N\}$ contains N-s+1 integers and the number of integers strictly to the right of j_s is only N-s.

In the same way the following lemma can be proved.

LEMMA 2.2. *For each* (i,j) *there either exists* k *such that*

(i) $\qquad k \in \{i_1,\dots,i_s\}$

and

(ii) $\qquad y_{j_s} + \theta_{j_s} \le y_{i_s} + \theta_k \quad$ *for all* $y \in A_iB_j$

or A_iB_j *is empty.*

If we now define the random variables whose values on each A_i are

$$\text{(2.2)} \qquad \begin{cases} \theta_{\ell s} = \min\{\theta_{i_s},\dots,\theta_{i_N}\} \\ \theta_{us} = \max\{\theta_{i_1},\dots,\theta_{i_s}\} \end{cases} \qquad s = 1,\dots,N,$$

then we have, on each A_iB_j,

$$\text{(2.3)} \qquad y_{i_s} + \theta_{\ell s} \le y_{j_s} + \theta_{j_s} \le y_{i_s} + \theta_{us}, \qquad s = 1,\dots,N.$$

Integrating with respect to p_0 and using (2.1) gives

(2.4) $$E_0 O_s + E_0\theta_{\ell s} \le E_\theta O_s \le E_0 O_s + E_0\theta_{us} \qquad s = 1,\dots,N,$$

where $E_0\theta_{\ell s} = \Sigma_i\, P_0(A_i)\theta_{\ell s}$.
Clearly, for a monotone nondecreasing function $h(O_s)$

$$E_0 h(O_s + \theta_{\ell s}) \le E_\theta h(O_s) \le E_0 h(O_s + \theta_{us}), \qquad s = 1,\dots,N,$$

can be obtained as above. A similar result holds for a function $h(O_1,\dots,O_N)$, where h is monotone in each variable; the upperbound and lowerbound for $E_\theta h(O_1,\dots,O_N)$ are determined by the direction of the monotonicity of h in each argument. This result will now be used to find bounds for the expectation of linear combinations $\Sigma_{s=1}^N \alpha_s O_s$. First note that on $A_i B_j$ we have, for each s = 1,...,N,

(2.5) $$y_{j_s} + \theta_{j_s} = y_{i_s} + y_{j_s} - y_{i_s} + \theta_{j_s} = y_{i_s} + \Delta_{ijs}, \quad \text{say.}$$

Summing (2.5) with respect to s gives

(2.6) $$\sum_{s=1}^N \Delta_{ijs} = \sum_{s=1}^N \theta_{j_s} = N\bar\theta.$$

Hence we can write

(2.7) $$\sum_{s=1}^N \alpha_s(y_{j_s} + \theta_{j_s}) = \sum_{s=1}^N \alpha_s y_{i_s} + \sum_{s=1}^N \alpha_s\Delta_{ijs} = $$
$$= \sum_{s=1}^N \alpha_s y_{i_s} + \sum_{s=1}^N (\alpha_s - \alpha_0)\Delta_{ijs} + \alpha_0 N\bar\theta,$$

where α_0 is an arbitrary constant. Since, by (2.3), Δ_{ijs} is in the interval $[\theta_{\ell s},\theta_{us}]$, we have

(2.8) $$\sum_{s=1}^N \alpha_s(y_{js} + \theta_{js}) \ge \sum_{s=1}^N \alpha_s y_{i_s} + \sum_{\alpha_s>\alpha_0} (\alpha_s - \alpha_0)\theta_{\ell s} +$$
$$+ \sum_{\alpha_s<\alpha_0} (\alpha_s - \alpha_0)\theta_{us} + \alpha_0 N\bar\theta$$

and

$$(2.9)\qquad \sum_{s=1}^{N} \alpha_s(y_{j_s} + \theta_{j_s}) \le \sum_{s=1}^{N} \alpha_s y_{i_s} + \sum_{\alpha_s>\alpha_0} (\alpha_s - \alpha_0)\theta_{us} + \sum_{\alpha_s<\alpha_0} (\alpha_s - \alpha_0)\theta_{\ell s} + \alpha_0 N\bar{\theta}.$$

Integrating with respect to p_0 gives the following lower and upper bounds

$$(2.10)\qquad E_\theta \sum_{s=1}^{N} \alpha_s O_s \ge E_0 \sum_{s=1}^{N} \alpha_s O_s + \sum_{\alpha_s>\alpha_0} (\alpha_s - \alpha_0)E_0\theta_{\ell s} + \sum_{\alpha_s<\alpha_0} (\alpha_s - \alpha_0)E_0\theta_{us} + \alpha_0 N\bar{\theta}$$

and

$$(2.11)\qquad E_\theta \sum_{s=1}^{N} \alpha_s O_s \le E_0 \sum_{s=1}^{N} \alpha_s O_s + \sum_{\alpha_s>\alpha_0} (\alpha_s - \alpha_0)E_0\theta_{us} + \sum_{\alpha_s<\alpha_0} (\alpha_s - \alpha_0)E_0\theta_{\ell s} + \alpha_0 N\bar{\theta}.$$

The distance between these bounds is

$$(2.12)\qquad \sum_{s=1}^{N} |\alpha_s - \alpha_0| E_0(\theta_{us} - \theta_{\ell s}),$$

which is minimized by choosing α_0 as the median of the α_s with respect to the probabilities

$$(2.13)\qquad \frac{E_0(\theta_{us}-\theta_{\ell s})}{\sum_{s=1}^{N} E_0(\theta_{us}-\theta_{\ell s})}, \qquad s = 1,\dots,N.$$

Sometimes a shorter bound can be found by maximizing (2.10) and minimizing (2.11) separately. Consider the upper bound (2.11) which can be written

$$\mathcal{E}_\theta \sum_{s=1}^{N} \alpha_s O_s - \mathcal{E}_0 \sum_{s=1}^{N} \alpha_s O_s \le$$

$$\le \sum_{s=1}^{N} \alpha_s \mathcal{E}_0 \theta_{us} - \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0) \mathcal{E}_0 (\theta_{us} - \theta_{\ell s}) + \alpha_0 [N\bar{\theta} - \sum_{s=1}^{N} \mathcal{E}_0 \theta_{us}]$$

$$= \sum_{s=1}^{N} \alpha_s \mathcal{E}_0 \theta_{us} + g(\alpha_0).$$

Because $g(\alpha_0)$ is continuous and has a nondecreasing derivative (except for $\alpha_0 = \alpha_s$, $s = 1,\dots,N$) which is nonpositive for small α_0 and nonnegative for large α_0, $g(\alpha_0)$ either has a minimum or an interval of minima and a minimizing α_0 is an $\alpha_{(s_0)}$ which satisfies the two inequalities

$$\begin{cases} \sum_{\alpha_{(s)} < \alpha_{(s_0)}} \mathcal{E}_0 (\theta_{us} - \theta_{\ell s}) \le \sum_{s=1}^{N} \mathcal{E}_0 \theta_{us} - N\bar{\theta} \\ \sum_{\alpha_{(s)} \le \alpha_{(s_0)}} \mathcal{E}_0 (\theta_{us} - \theta_{\ell s}) \ge \sum_{s=1}^{N} \mathcal{E}_0 \theta_{us} - N\bar{\theta}, \end{cases}$$

where $\alpha_{(1)} \le \dots \le \alpha_{(N)}$ are the ordered values of α_s.

In the same way an α_0 maximizing the lower bound for $\mathcal{E}_\theta \Sigma_{s=1}^{N} \alpha_s O_s$ is an $\alpha_{(s_0)}$ which satisfies the two inequalities

$$\begin{cases} \sum_{\alpha_{(s)} < \alpha_{(s_0)}} \mathcal{E}_0 (\theta_{us} - \theta_{\ell s}) \le N\bar{\theta} - \sum_{s=1}^{N} \mathcal{E}_0 \theta_{\ell s} \\ \sum_{\alpha_{(s)} \le \alpha_{(s_0)}} \mathcal{E}_0 (\theta_{us} - \theta_{\ell s}) \ge N\bar{\theta} - \sum_{s=1}^{N} \mathcal{E}_0 \theta_{\ell s}. \end{cases}$$

An example where a shorter bound is obtained by maximizing (2.10) and minimizing (2.11) separately is the following. Let $N = 4$, $\theta_1 = 0 < \theta_2 = \theta_3 = \theta_4 = \theta$ and let the X_i be p_0-exchangeable. Then, for $\alpha_4 < \alpha_1 < \alpha_3 < \alpha_2$, the median of the α_s with respect to the probabilities (2.13) is α_3. Using this median for α_0 in the bounds for $E_\theta\ \Sigma_{s=1}^N\ \alpha_s O_s - E_0\ \Sigma_{s=1}^N\ \alpha_s O_s$ gives

$$\text{lower bound} = \theta\{\tfrac{3}{4}\alpha_1 + \tfrac{1}{4}\alpha_2 + \alpha_3 + \alpha_4\}$$

$$\text{upper bound} = \theta\{\alpha_2 + \tfrac{5}{4}\alpha_3 + \tfrac{3}{4}\alpha_4\}$$

and the length of this interval is $\theta[\frac{3}{4}(\alpha_2 - \alpha_1) + \frac{1}{4}(\alpha_3 - \alpha_4)]$. The upper bound is minimized for $\alpha_0 = \alpha_1$ and the upper bound then becomes

$$\theta\{\tfrac{1}{4}\alpha_1 + \alpha_2 + \alpha_3 + \tfrac{3}{4}\alpha_4\}.$$

The lower bound is maximized for $\alpha_0 = \alpha_3$ which is the median. The difference between these upper- and lowerbounds is $\theta[\frac{3}{4}(\alpha_2 - \alpha_1) + \frac{1}{4}(\alpha_1 - \alpha_4)]$ and the difference between the two lengths is $\frac{1}{4}\theta(\alpha_3 - \alpha_1) > 0$.

If $X_1,\ldots,X_N$ are p_0-exchangeable, then $p_0(A_i) = \frac{1}{N!}$, the distributions of $\theta_{\ell s}$ and θ_{us} do not depend on p_0 and are simple to compute. Some distributions of $\theta_{\ell s}$ and θ_{us} are given in section 4.

3. UPPERBOUND FOR THE STANDARD DEVIATION OF $\sum_{s=1}^N \alpha_s O_s$.

From (2.8) and (2.9) we have

$$(3.1)\quad \begin{cases} \sum_{s=1}^N \alpha_s(y_{j_s} + \theta_{j_s}) \quad \text{is} \\ \le \sum_{s=1}^N \alpha_s y_{i_s} + \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)\theta_{us} + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)\theta_{\ell s} + \alpha_0 N\bar{\theta} \quad \text{and} \\ \ge \sum_{s=1}^N \alpha_s y_{i_s} + \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)\theta_{\ell s} + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)\theta_{us} + \alpha_0 N\bar{\theta}. \end{cases}$$

Then, if

$$Y = \sum_{s=1}^N \alpha_s(y_{j_s} + \theta_{j_s}) - E_0 \sum_{s=1}^N \alpha_s O_s,$$

$$X = \sum_{s=1}^{N} \alpha_s y_{i_s} - E_0 \sum_{s=1}^{N} \alpha_s O_s,$$

$$Z_1 = \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)(\theta_{\ell s} - E_0\theta_{\ell s}) + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)(\theta_{us} - E_0\theta_{us}),$$

$$Z_2 = \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)(\theta_{us} - E_0\theta_{us}) + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)(\theta_{\ell s} - E_0\theta_{\ell s}),$$

$$A = \alpha_0 N\bar{\theta} - E_\theta \sum_{s=1}^{N} \alpha_s O_s + E_0 \sum_{s=1}^{N} \alpha_s O_s + \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)E_0\theta_{\ell s} + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)E_0\theta_{us},$$

$$B = \alpha_0 N\bar{\theta} - E_\theta \sum_{s=1}^{N} \alpha_s O_s + E_0 \sum_{s=1}^{N} \alpha_s O_s + \sum_{\alpha_s > \alpha_0} (\alpha_s - \alpha_0)E_0\theta_{us} + \sum_{\alpha_s < \alpha_0} (\alpha_s - \alpha_0)E_0\theta_{\ell s},$$

(3.1) becomes

(3.2) $$X + Z_1 + A \le Y \le X + Z_2 + B.$$

To find an upper bound for $\sigma(Y) = \sigma(\Sigma_{s=1}^{N} \alpha_s O_s)$ consider parabolas

$$k(X + \frac{Z_1+Z_2+A+B}{2})^2 + \ell$$

with k and ℓ such that

$$\max\{(X+Z_1+A)^2, (X+Z_2+B)^2\} \le k\{X + \frac{Z_1+Z_2+A+B}{2}\}^2 + \ell.$$

With a little algebra one sees that $k > 1$ and

$$\ell \ge \frac{k}{k-1}(\frac{Z_2+B-Z_1-A}{2})^2$$

are necessary. Let

$$\ell = \frac{k}{k-1}\left(\frac{Z_2+B-Z_1-A}{2}\right)^2$$

and k will be chosen after integration. Integrating with respect to p_0 gives

(3.3) $$\sigma_\theta^2\left(\sum_{s=1}^{N} \alpha_s O_s\right) \le kE_0\left(X + \frac{Z_1+Z_2+A+B}{2}\right)^2 + \frac{k}{k-1} E_0\left(\frac{Z_2+B-Z_1-A}{2}\right)^2$$

for all $k > 1$.

Minimizing the right hand side of (3.3) with respect to $k > 1$ one finds for the minimizing value of k

$$k = 1 + \left\{\frac{E_0\left(\frac{Z_2+B-Z_1-A}{2}\right)^2}{E_0\left(X + \frac{Z_1+Z_2+A+B}{2}\right)^2}\right\}^{\frac{1}{2}} .$$

Substitution of this value in (3.3) gives

(3.4) $$\sigma_\theta\left(\sum_{s=1}^{N} \alpha_s O_s\right) \le \left\{E_0\left(X + \frac{Z_1+Z_2+A+B}{2}\right)^2\right\}^{\frac{1}{2}} + \left\{E_0\left(\frac{Z_2+B-Z_1-A}{2}\right)^2\right\}^{\frac{1}{2}}.$$

$$= \left\{E_0\left(X + \frac{Z_1+Z_2}{2}\right)^2 + \left(\frac{A+B}{2}\right)^2\right\}^{\frac{1}{2}} + \left\{E_0\left(\frac{Z_2+B-Z_1-A}{2}\right)^2\right\}^{\frac{1}{2}}.$$

This bound for $\sigma_\theta(\Sigma_{s=1}^N \alpha_s O_s)$ contains $E_\theta\ \Sigma_{s=1}^N \alpha_s O_s$; also, it depends on the correlation between X and $Z_1 + Z_2$. Using (2.10) and (2.11) one obtains

$$\frac{|A+B|}{2} = |\alpha_0 N\bar{\theta} - E_\theta \sum_{s=1}^{N} \alpha_s O_s + E_0 \sum_{s=1}^{N} \alpha_s O_s + \frac{1}{2} \sum_{s=1}^{N} (\alpha_s - \alpha_0)E_0(\theta_{us} + \theta_{\ell s})|$$

$$\le \frac{1}{2} \sum_{s=1}^{N} |\alpha_s - \alpha_0| E_0(\theta_{us} - \theta_{\ell s}).$$

Substituting this bound in (3.4) gives the following bound for $\sigma_\theta(\Sigma_{s=1}^N \alpha_s O_s)$

(3.5)
$$\sigma_\theta\left(\sum_{s=1}^N \alpha_s O_s\right) \le$$
$$\le \{\sigma_0^2\left(\sum_{s=1}^N \alpha_s O_s + \frac{1}{2}\sum_{s=1}^N (\alpha_s - \alpha_0)(\theta_{us} + \theta_{\ell s})\right) + $$
$$+ \frac{1}{4}\left(\sum_{s=1}^N |\alpha_s - \alpha_0| E_0(\theta_{us} - \theta_{\ell s})\right)^2\}^{\frac{1}{2}} +$$
$$+ [\frac{1}{4} E_0\{\sum_{s=1}^N |\alpha_s - \alpha_0|(\theta_{us} - \theta_{\ell s})\}^2]^{\frac{1}{2}}.$$

If X and $Z_1 + Z_2$ are uncorrelated under p_0, for example if the X_i are p_0-exchangeable, the bound (3.5) becomes

(3.6)
$$\sigma_\theta\left(\sum_{s=1}^N \alpha_s O_s\right) \le$$
$$\le \{\sigma_0^2\left(\sum_{s=1}^N \alpha_s O_s\right) + \frac{1}{4}\sigma_0^2\left(\sum_{s=1}^N (\alpha_s - \alpha_0)(\theta_{us} + \theta_{\ell s})\right)$$
$$+ \frac{1}{4}\left(\sum_{s=1}^N |\alpha_s - \alpha_0| E_0(\theta_{us} - \theta_{\ell s})\right)^2\}^{\frac{1}{2}} +$$
$$+ [\frac{1}{4} E_0\{\sum_{s=1}^N |\alpha_s - \alpha_0|(\theta_{us} - \theta_{\ell s})\}^2]^{\frac{1}{2}}.$$

4. THE DISTRIBUTION OF $(\theta_{\ell s}, \theta_{us})$ WHEN THE X_i ARE p_0-EXCHANGEABLE.

Let $\theta_1' < \ldots < \theta_M'$ be the ordered values of the θ's, let, for $k = 1,\ldots, M$, N_k be the number of $\theta_j' \le \theta_k'$ and let $N_0 = 0$. If the X_i are p_0-exchangeable, then $P_0(A_i) = \frac{1}{N!}$ and a somewhat lengthy but straightforward combinatorial argument gives, for the joint distribution of $\theta_{\ell s}$ and θ_{us},

$$(4.1)\qquad \frac{N!}{(s-1)!(N-s)!}\,P_0(\theta_{\ell s}=\theta_j',\ \theta_{us}=\theta_k') = \begin{cases} \dfrac{(N_k-N_{k-1})!}{(s-1-N_{k-1})!(N_k-s)!} & \text{if } j=k \text{ and } N_{k-1}+1 \le s \le N_k, \\ A-B-C+D & \text{if } j<k \text{ and } N_{j-1}+1 \le s \le N_k, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$A = \frac{(N_k-N_{j-1})!}{(s-1-N_{j-1})!(N_k-s)!}$$

$$B = \begin{cases} \dfrac{(N_k-N_j)!}{(s-1-N_j)!(N_k-s)!} & \text{if } s \ge N_j+1 \\ 0 & \text{if not} \end{cases}$$

$$C = \begin{cases} \dfrac{(N_{k-1}-N_{j-1})!}{(s-1-N_{j-1})!(N_{k-1}-s)!} & \text{if } s \le N_{k-1} \\ 0 & \text{if not} \end{cases}$$

$$D = \begin{cases} \dfrac{(N_{k-1}-N_j)!}{(s-1-N_j)!(N_{k-1}-s)!} & \text{if } N_j+1 \le s \le N_{k-1} \\ 0 & \text{if not.} \end{cases}$$

From (4.1), or by direct computation, one obtains for the marginal distributions of $\theta_{\ell s}$ and θ_{us}

$$(4.2)\qquad \binom{N}{N-s+1}P_0(\theta_{\ell s}=\theta_k') = \binom{N-N_{k-1}}{N-s+1} - \binom{N-N_k}{N-s+1}$$

and

(4.3) $$\binom{N}{s} P_0(\theta_{us} = \theta'_k) = \binom{N_k}{s} - \binom{N_{k-1}}{s}.$$

If the θ's are all unequal then $N_k = k$ and the joint and marginal distributions of $\theta_{\ell s}$ and θ_{us} become

(4.4) $$\frac{N!}{(s-1)!(N-s)!} P_0(\theta_{\ell s} = \theta'_j,\ \theta_{us} = \theta'_k) = \begin{cases} (k-s)\binom{k-j-1}{s-j-1} + \binom{k-j}{s-j} & \text{if } j \le s \le k \\ 0 & \text{otherwise,} \end{cases}$$

(4.5) $$P_0(\theta_{\ell s} = \theta'_k) = \frac{\binom{N-k}{N-s}}{\binom{N}{N-s+1}} = \frac{N-s+1}{s} \frac{\binom{s}{k}}{\binom{N}{k}}$$

and

(4.6) $$P_0(\theta_{us} = \theta'_k) = \frac{\binom{k-1}{s-1}}{\binom{N}{s}} = \frac{s}{N-s+1} \frac{\binom{N-s+1}{N-k+1}}{\binom{N}{N-k+1}}.$$

Two simple limiting distributions for $\theta_{\ell s}$, when the θ's are all unequal, are the following. If $\frac{s}{N} \to \lambda$, $0 < \lambda < 1$, then

(4.7) $$P_0(\theta_{\ell s} = \theta'_k) \to \frac{1-\lambda}{\lambda} \lambda^k, \qquad k = 1,2,\ldots .$$

If $s = N - a$, a fixed, then

(4.8) $$P_0(\theta_{\ell s} \le \theta'_{[Nx]}) \to (a+1) \int_0^x (1-t)^a dt \qquad 0 < x < 1.$$

Similar results hold for the distribution of θ_{us}.

If M = 2, i.e. the two sample problem, then the joint distribution of $(\theta_{\ell s}, \theta_{us})$ is

$$(4.9)\quad \begin{cases} P_0(\theta_{\ell s} = \theta_{us} = \theta_1') = \dfrac{\binom{n}{s}}{\binom{N}{s}}, \\[2ex] P_0(\theta_{\ell s} = \theta_1', \theta_{us} = \theta_2') = 1 - \dfrac{\binom{n}{s}}{\binom{N}{s}} - \dfrac{\binom{m}{N-s+1}}{\binom{N}{N-s+1}}, \\[2ex] P_0(\theta_{\ell s} = \theta_{us} = \theta_2') = \dfrac{\binom{m}{N-s+1}}{\binom{N}{N-s+1}}, \end{cases}$$

where $n = N_1$ and $m = N - N_1$. The marginal distributions are

$$(4.10)\quad \begin{cases} P_0(\theta_{\ell s} = \theta_1') = 1 - \dfrac{\binom{m}{N-s+1}}{\binom{N}{N-s+1}} \\[2ex] P_0(\theta_{\ell s} = \theta_2') = \dfrac{\binom{m}{N-s+1}}{\binom{N}{N-s+1}} \end{cases}$$

and

$$(4.11)\quad \begin{cases} P_0(\theta_{us} = \theta_1') = \dfrac{\binom{n}{s}}{\binom{N}{s}} \\[2ex] P_0(\theta_{us} = \theta_2') = 1 - \dfrac{\binom{n}{s}}{\binom{N}{s}}. \end{cases}$$

Note that, for the two sample case, one of the distributions of $\theta_{\ell s}$ and θ_{us} is degenerate, so $\theta_{\ell s}$ and θ_{us} are independent.

5. ATTAINABILITY

In this section distributions are given for which the bounds are attained, as well as examples where they are almost attained.

EXAMPLE a. If p_0 is concentrated on a single A_i, say $i = \{1,\ldots,N\}$, and the order of the θ's is the same as i, here $\theta_1 \leq \ldots \leq \theta_N$, then $\theta_{\ell s} = \theta_{us} = \theta_{is}$, here θ_s, with probability 1 for every s and the inequalities (2.10) and (2.11) for the mean and the inequalities (3.5) and (3.6) for the standard deviation become equalities.

EXAMPLE b. If the α_s are all equal, say equal to $\frac{1}{N}$, and α_0 is the common value then the inequalities (2.10), (2.11), (3.5) and (3.6) are equalities.

EXAMPLE c. The inequalities (2.10), (2.11), (3.5) and (3.6) are continuous as the θ's approach equality and hence can be made arbitrary close to equality for $\theta_{max} - \theta_{min}$ sufficiently small.

EXAMPLE d. $\theta_{max} - \theta_{min}$ does not have to be close to zero to get approximate equality. For N odd let $X_1,\ldots,X_{N+2}$ be independent. For large N there is practically no difference between the distribution of the median of $X_1,\ldots,X_{N+2}$ from U(0,1) and the median of $X_1,\ldots,X_N$ from U(0,1). Hence, for $\theta > 1$, there is, for large N, practically no difference between the distribution of the median of $X_1,\ldots,X_{N+2}$ from U(0,1) and the median of $X_1,\ldots,X_{N+2}$ for $X_1,\ldots,X_N$ from U(0,1) and X_{N+1}, X_{N+2} from $U(\theta,\theta+1)$ and the inequalities will be almost equalities.

6. SOME COMPARISONS BETWEEN ACTUAL VALUES AND THEIR BOUNDS

Let X_1, X_2, X_3 be independent with X_1 and X_2 double exponential with mean 0 and X_3 double exponential with mean $\theta \geq 0$. Tables 6.1, 6.2 and 6.3 below give, for several values of θ, ($\sigma = \sigma(X_i) = \sqrt{2}$) the values of $E_\theta O_s$ and $\sigma_\theta(O_s)$, s = 1,2,3, together with the bounds (2.10), (2.11) and (3.6). For all three tables and all values of θ the value of α_0 used for the bounds is $\alpha_0 = 0$. This value of α_0 maximizes the lower bound (2.10), minimizes the upper bound (2.11) and is the median of the α_s with respect to the weights (2.13).

Table 6.1

Comparison of $E_\theta O_1$ and $\sigma_\theta(O_1)$ with their bounds

	$E_\theta O_1$			$\sigma_\theta(O_1)$	
θ	lower bound	value	upper bound	value	upper bound
0	-1.125	-1.125	-1.125	1.1895	1.1895
$.04\sigma$	-1.125	-1.1065	-1.1061	1.1895	1.2060
$.1\ \sigma$	-1.125	-1.0803	-1.0779	1.1896	1.2310
$.4\ \sigma$	-1.125	-.9732	-.9364	1.1904	1.3640
σ	-1.125	-.8488	-.6536	1.1932	1.6659

Table 6.2

Comparison of $E_\theta O_2$ and $\sigma_\theta(O_2)$ with their bounds

	$E_\theta O_2$			$\sigma_\theta(O_2)$	
θ	lower bound	value	upper bound	value	upper bound
0	0	0	0	.7993	.7993
$.04\sigma$	0	.0189	.0377	.7995	.8227
$.1\ \sigma$	0	.0471	.0943	.8003	.8591
$.4\ \sigma$	0	.1838	.3771	.8154	1.0629
σ	0	.4111	.9428	.8852	1.5634

Table 6.3

Comparison of $E_\theta O_3$ and $\sigma_\theta(O_3)$ with their bounds

θ	$E_\theta O_3$			$\sigma_\theta(O_3)$	
	lower bound	value	upper bound	value	upper bound
0	1.125	1.125	1.125	1.1895	1.1895
.04σ	1.1439	1.1443	1.1816	1.1895	1.2128
.1 σ	1.1721	1.1747	1.2664	1.1896	1.2486
.4 σ	1.3136	1.3551	1.6907	1.1920	1.4427
σ	1.5964	1.8520	2.5392	1.2138	1.8996

As a second example, let X_1, X_2, X_3, X_4 be independent with X_1 and X_2 U(0,1) and X_3 and X_4 U(θ,θ+1), $0 \le \theta \le 1$. Tables 6.4 and 6.5 below give, for several values of θ, ($\sigma = \sigma(X_i) = \frac{1}{\sqrt{12}}$) the values of $E_\theta O_1$, $\sigma_\theta(O_1)$, $E_\theta \frac{O_1+O_4}{2}$ and $\sigma_\theta(\frac{O_1+O_4}{2})$. In both tables and for all values of θ the value of α_0 used for the bounds in $\alpha_0 = 0$; this value maximizes the lower bound (2.10), minimizes the upper bound (2.11) and is the median of the α_s with respect to the weights (2.13).

Table 6.4

Comparison of $E_\theta O_1$ and $\sigma_\theta(O_1)$ with their bounds

θ	$E_\theta O_1$			$\sigma_\theta(O_1)$	
	lower bound	value	upper bound	value	upper bound
0	.2	.2	.2	.1633	.1633
.04σ	.2	.2057	.2058	.1634	.1674
.1 σ	.2	.2139	.2144	.1638	.1738
.4 σ	.2	.2494	.2577	.1700	.2091
σ	.2	.2967	.3443	.1937	.2046

Table 6.5

Comparison of $E_\theta \frac{O_1+O_4}{2}$ and $\sigma_\theta\left(\frac{O_1+O_4}{2}\right)$ with their bounds

	$E_\theta \frac{O_1+O_4}{2}$			$\sigma_\theta\left(\frac{O_1+O_4}{2}\right)$	
θ	lower bound	value	upper bound	value	upper bound
0	.5	.5	.5	.1291	.1291
.04σ	.5029	.5058	.5087	.1291	.1329
.1 σ	.5072	.5144	.5217	.1294	.1387
.4 σ	.5289	.5577	.5866	.1327	.1706
σ	.5722	.6443	.7165	.1451	.2468

Département de Mathématiques et de Statistique
Université de Montréal
Montréal, Québec, Canada.