MAXIMUM PROBABILITY ESTIMATORS FOR RANKED MEANS*

EDWARD J. DUDEWICZ

(Received June 23, 1969; revised July 5, 1973)

1. Introduction

Suppose that observations from populations $\pi_1, \cdots, \pi_k$ ($k \geqq 2$) are normally distributed with unknown means $\mu_1, \cdots, \mu_k$ (respectively) and a common known variance σ^2. Let $\mu_{[1]} \leqq \cdots \leqq \mu_{[k]}$ denote the ranked means. Several ranking and selection procedures take n independent observations from each population, denote the sample mean of the n observations from π_i by $\bar{X}_i$ ($i=1, \cdots, k$), and utilize the ranked sample means $\bar{X}_{[1]} \leqq \cdots \leqq \bar{X}_{[k]}$. (See Dudewicz [3] for details.) We assume throughout that both the numerical values of $\mu_1, \cdots, \mu_k$ and the pairings of the $\mu_{[1]}, \cdots, \mu_{[k]}$ with the populations $\pi_1, \cdots, \pi_k$ are completely unknown and consider problems of estimation of $\mu_{[i]}$ ($1 \leqq i \leqq k$) based on the statistics provided by the single-stage rule stated above, and utilizing recent work of Weiss and Wolfowitz.

Generalized maximum likelihood estimators (GMLE's), introduced by Weiss and Wolfowitz [5], provide (where available) asymptotically efficient estimators, whereas this is not always true for MLE's even if the latter can be found. Most classical MLE theory assumes i.i.d. observations and is therefore not applicable in our case; the theory of Weiss and Wolfowitz [5] allows for more general situations (although most of their applications are to i.i.d. "non-regular" cases). (Corrections to Weiss and Wolfowitz [5] are contained in Weiss and Wolfowitz [6], in Weiss and Wolfowitz [8], and in Dudewicz [2]. An additional example is given in Weiss and Wolfowitz [7].)

Maximum probability estimators (MPE's) were introduced by Weiss and Wolfowitz [6] for much the same reason as GMLE's were introduced by Weiss and Wolfowitz [5]. Weiss and Wolfowitz [6], 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). Below we study the extension of this result to $m \geqq 1$ parameters, first summarizing Weiss and Wolfowitz's results.

* This research was supported by ONR contracts N00014-68A-0091 and Nonr 401(53), and by NSF grants GP-8958 and GP-7798.

We then study in detail the MPE of the ranked means. We shall need the following definitions, conditions, and theorems, all of which are due to Weiss and Wolfowitz.

DEFINITION 1. Let Θ be a closed region in $\mathcal{R}^m$, $\Theta \subseteq \bar{\Theta}$ with $\bar{\Theta}$ a closed region such that every finite boundary point of Θ is an inner point of $\bar{\Theta}$.

DEFINITION 2. For each n let $X(n)$ denote the (finite) vector of r.v.'s of which the estimator is to be a function.

DEFINITION 3. 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.

DEFINITION 4. Let R be a fixed bounded set in $\mathcal{R}^m$, let $k(n)=(k_1(n),\cdots,k_m(n))$ be such that $k(n)\to\infty$, let $d=(d_1,\cdots,d_m)$, and define

$$d-R/k(n)=\{(z_1,\cdots,z_m)\in\bar{\Theta}:\ d_i-y_i/k_i(n)=z_i,\quad i=1,\cdots,m,\ (y_1,\cdots,y_m)\in R\}\ .$$

DEFINITION 5. Z_n is a *maximum probability estimator* with respect to R and $k(n)$ if (for a.e. (μ_n) value x of $X(n)$) $Z_n(x)$ equals a $d\in\bar{\Theta}$ such that

$$\int\cdots\int_{d-[k(n)]^{-1}R} K_n(x|\theta)d\theta_1\cdots d\theta_m=\sup_{t\in\bar{\Theta}}\int\cdots\int_{t-[k(n)]^{-1}R} K_n(x|\theta)d\theta_1\cdots d\theta_m\ .$$

CONDITION 1. For each $h>0$ and $\theta_0\in\Theta$

$$\lim_{n\to\infty} \mathrm{P}_\theta\,[k(n)(Z_n-\theta)\in R]=\beta$$

uniformly for all $\theta\in H=\{\theta:|k(n)(\theta-\theta_0)|\leqq h\}$.

CONDITION 2. For each $\theta_0\in\Theta$

$$\lim_{\substack{n\to\infty\\ M\to\infty}} \mathrm{P}_\theta\,[|k(n)(Z_n-\theta)|<M]=1$$

uniformly for all θ in some neighborhood of θ_0.

CONDITION 3. For each $\theta_0\in\Theta$ and $h>0$

$$\lim_{n\to\infty}\{\mathrm{P}_\theta\,[k(n)(T_n-\theta)\in R]-\mathrm{P}_{\theta_0}[k(n)(T_n-\theta_0)\in R]\}=0$$

uniformly for all $\theta\in H=\{\theta:|k(n)(\theta-\theta_0)|\leqq h\}$.

THEOREM 1. *Let $\{Z_n\}$ be an MPE with respect to R and $k(n)$. Suppose $\{Z_n\}$ satisfies Conditions* 1 *and* 2. *Let $\{T_n\}$ be any estimator which* *satisfies Condition* 3. *Then* (*for each* $\theta_0 \in \Theta$)

$$\beta \geqq \overline{\lim_{n\to\infty}} \mathrm{P}_{\theta_0} [k(n)(T_n-\theta_0) \in R] .$$

2. Maximum probability estimators for ranked means

THEOREM 2. *Let* W_n *be a GMLE* (*with respect to* $r=(r_1, \cdots, r_m)>0$) *for the estimation of* $\theta=(\theta_1, \cdots, \theta_m) \in \Theta$ $(m \geqq 1)$. *Choose* $R=\{(y_1, \cdots, y_m): -r_i/2<y_i \leqq r_i/2,\ i=1, \cdots, m\}$ *and* $k(n)$ *as for the GMLE. If the MPE* (*w.r.t. this* R *and* $k(n)$) *satisfies Conditions* 1 *and* 2, *and if the GMLE satisfies Condition* 3, *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 a condition due to Weiss and Wolfowitz (see Theorem (5.3.13) of Dudewicz [2]). Thus (for each $\theta_0 \in \Theta$)

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0} [k(n)(W_n-\theta_0) \in R] \geqq \overline{\lim_{n\to\infty}} \mathrm{P}_{\theta_0} [k(n)(Z_n-\theta_0) \in R] . \tag{1}$$

The GMLE W_n satisfies Condition 3 and thus the conclusion of Theorem 1 holds: for each $\theta_0 \in \Theta$

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0} [k(n)(Z_n-\theta_0) \in R] \geqq \overline{\lim_{n\to\infty}} \mathrm{P}_{\theta_0} [k(n)(W_n-\theta_0) \in R] . \tag{2}$$

Then (see Weiss and Wolfowitz [6], p. 198) the GMLE is (in the equivalence class of such) an MPE.

The result of Weiss and Wolfowitz [6] for the case $m=1$ is somewhat stronger than our Theorem 2 for the case $m \geqq 1$: they show that the MPE satisfies Conditions 1 and 2. (They assume, as we do, that the GMLE satisfies Condition 3, which is stronger than a condition (A′) they use elsewhere; see (5.3.5) of Dudewicz [2].) Our result (more precisely, a slight extension of our result) says that if the MPE for a problem is "good" (i.e., satisfies Conditions 1 and 2), then the GMLE (if it meets Condition 3) 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 [6], 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 $\Theta=\{\mu\colon \mu_i \in \mathcal{R}\ (i=1,\cdots,k),\ \mu_1=\mu_{[1]},\cdots,\mu_k=\mu_{[k]}\}$ and $\bar{\Theta}=\mathcal{R}^k$, and let $X(n)$, $K_n(x|\mu)$, μ_n be specified as*

$$X(n)=(\bar{X}_{[1]},\cdots,\bar{X}_{[k]})$$

$$\text{(3)}\qquad K_n(x|\theta)=K_n(x|\mu)=f^{(\mu)}_{\bar{X}_{[1]},\cdots,\bar{X}_{[k]}}(x_1,\cdots,x_k) \equiv f_{\bar{X}_{[1]},\cdots,\bar{X}_{[k]}}(x_1,\cdots,x_k)$$

$$\mu_n=\text{Lebesgue measure on } \mathcal{R}^k .$$

Fix $r=(r_1,\cdots,r_k)>0$ and choose $k_1(n)=\cdots=k_k(n)=\sqrt{n}/\sigma$, $R=\{(y_1,\cdots,y_k)\colon -r_i/2<y_i\leqq r_i/2,\ i=1,\cdots,k\}$. Then

$$\text{(4)}\qquad \begin{aligned} d-[k(n)]^{-1}R&=\{(z_1,\cdots,z_k)\in\bar{\Theta}\colon d_i-y_i/k_i(n)=z_i,\\ &\qquad i=1,\cdots,k,\ (y_1,\cdots,y_k)\in R\}\\ &=\left\{(z_1,\cdots,z_k)\colon d_i-\frac{r_i}{2k_i(n)}\right.\\ &\qquad\left.\leqq z_i<d_i+\frac{r_i}{2k_i(n)},\ i=1,\cdots,k\right\}, \end{aligned}$$

and

$$\text{(5)}\qquad \begin{aligned} &\sup_{t\in\bar{\Theta}}\int\cdots\int_{t-[k(n)]^{-1}R} K_n(x|\mu)d\mu_{[1]}\cdots d\mu_{[k]}\\ &=\sup_{t_1,\cdots,t_k}\int_{t_k-\{(r_k/2)\sigma\}/\sqrt{n}}^{t_k+\{(r_k/2)\sigma\}/\sqrt{n}}\cdots\int_{t_1-\{(r_1/2)\sigma\}/\sqrt{n}}^{t_1+\{(r_1/2)\sigma\}/\sqrt{n}} K_n(x|\mu)d\mu_{[1]}\cdots d\mu_{[k]} . \end{aligned}$$

For the case $k=2$, (5) becomes (when $\bar{X}_{[1]}=x_1$ and $\bar{X}_{[2]}=x_2$)

$$\text{(6)}\qquad \begin{aligned} &\sup_{t_1,t_2}\frac{n}{2\pi\sigma^2}\int_{t_2-\{(r_2/2)\sigma\}/\sqrt{n}}^{t_2+\{(r_2/2)\sigma\}/\sqrt{n}}\int_{t_1-\{(r_1/2)\sigma\}/\sqrt{n}}^{t_1+\{(r_1/2)\sigma\}/\sqrt{n}}\\ &\quad\cdot\left\{\exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_2-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right)\right.\\ &\qquad\left.+\exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2-\frac{1}{2}\left(\frac{x_1-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right)\right\}d\mu_{[1]}d\mu_{[2]}\\ &=\sup_{t_1,t_2}\left[\int_{(t_2-x_2)/(\sigma/\sqrt{n})-r_2/2}^{(t_2-x_2)/(\sigma/\sqrt{n})+r_2/2}\int_{(t_1-x_1)/(\sigma/\sqrt{n})-r_1/2}^{(t_1-x_1)/(\sigma/\sqrt{n})+r_1/2}\right.\\ &\quad\cdot\frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{1}{2}\nu_1^2\right)\right)\frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{1}{2}\nu_2^2\right)\right)d\nu_1 d\nu_2\\ &\quad+\int_{(t_2-x_1)/(\sigma/\sqrt{n})-r_2/2}^{(t_2-x_1)/(\sigma/\sqrt{n})+r_2/2}\int_{(t_1-x_2)/(\sigma/\sqrt{n})-r_1/2}^{(t_1-x_2)/(\sigma/\sqrt{n})+r_1/2} \end{aligned}$$

* Note that we take the order statistic $(\bar{X}_{[1]},\cdots,\bar{X}_{[k]})$ to be fundamental. This quantity is the naive estimator of μ, and we will be interested in seeing how MPE's modify it.

$$\cdot \frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{1}{2}\nu_1^2\right)\right)\frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{1}{2}\nu_2^2\right)\right)d\nu_1 d\nu_2\Big]$$

$$=\sup_{t_1,t_2}\Big[\left\{\Phi\left(\frac{t_1-x_1}{\sigma/\sqrt{n}}+\frac{r_1}{2}\right)-\Phi\left(\frac{t_1-x_1}{\sigma/\sqrt{n}}-\frac{r_1}{2}\right)\right\}$$

$$\cdot\left\{\Phi\left(\frac{t_2-x_2}{\sigma/\sqrt{n}}+\frac{r_2}{2}\right)-\Phi\left(\frac{t_2-x_2}{\sigma/\sqrt{n}}-\frac{r_2}{2}\right)\right\}$$

$$+\left\{\Phi\left(\frac{t_1-x_2}{\sigma/\sqrt{n}}+\frac{r_1}{2}\right)-\Phi\left(\frac{t_1-x_2}{\sigma/\sqrt{n}}-\frac{r_1}{2}\right)\right\}$$

$$\cdot\left\{\Phi\left(\frac{t_2-x_1}{\sigma/\sqrt{n}}+\frac{r_2}{2}\right)-\Phi\left(\frac{t_2-x_1}{\sigma/\sqrt{n}}-\frac{r_2}{2}\right)\right\}\Big] .$$

LEMMA 1. *Let* $d\equiv(\sqrt{n}/\sigma)(x_2-x_1)$, $t_1=x_1+a_1\sigma/\sqrt{n}$, $t_2=x_2-a_2\sigma/\sqrt{n}$. *Then an MPE is* (t_1, t_2) *with* a_1, a_2 *which achieve*

$$(7)\qquad \sup_{a_1,a_2}[\{\Phi(a_1+r_1/2)-\Phi(a_1-r_1/2)\}\{\Phi(a_2+r_2/2)-\Phi(a_2-r_2/2)\}$$
$$+\{\Phi(a_1-d+r_1/2)-\Phi(a_1-d-r_1/2)\}\{\Phi(a_2-d+r_2/2)$$
$$-\Phi(a_2-d-r_2/2)\}] .$$

PROOF. By Definition 5, for our case as specified above (4), the MPE is (t_1, t_2) which achieves the supremum in (6). If we use $d=(\sqrt{n}/\sigma)(x_2-x_1)$ 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 $\Phi(x)=1-\Phi(-x)$ $(x\in\mathcal{R})$, this becomes as specified in the statement of the lemma.

LEMMA 2. *The supremum of Lemma* 1 *occurs only at* (a_1, a_2) *with* $0<a_1<d$, $0<a_2<d$.

PROOF. By reasoning as at (5.1.5) in Dudewicz [2], 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

$$\phi\left(a_1+\frac{r_1}{2}\right)-\phi\left(a_1-\frac{r_1}{2}\right)\begin{matrix}<0 & \text{if } a_1<d\\ >0 & \text{if } a_1>d\end{matrix}$$

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.

LEMMA 3. *By imposing a consistency criterion for estimators similar to* (5.1.4) *in Dudewicz* [2], *we may restrict ourselves to* (a_1, a_2) *with* $a_1+a_2\leqq d$.

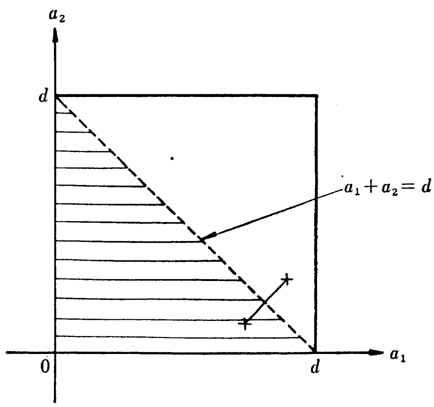

Fig. 1

PROOF. In order that we have $t_1\leqq t_2$, we must have $x_1+a_1\sigma/\sqrt{n}\leqq x_2-a_2\sigma/\sqrt{n}$, i.e., $a_1+a_2\leqq(\sqrt{n}/\sigma)(x_2-x_1)=d$.

Note that, in the region of (a_1, a_2)-space in which Lemma 2 tells us the supremum of Lemma 1 must lie, we have symmetry (of values of Lemma 1) about the line $a_1+a_2=d$; see Figure 1. Thus, our consistency criterion only eliminates an illogical duplicate maximizing point.

LEMMA 4. *For any fixed* $\delta>0$, *there is a* $K(r_1, r_2, \delta)$ *such that if* $d\geqq K(r_1, r_2, \delta)$ *then* (7) *is maximized* (*in the shaded region* I: $a_1>0$, $a_2>0$, $a_1+a_2\leqq d$ *of Figure* 1) *inside the disk* D: $a_1^2+a_2^2\leqq\delta$.

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 (7) is $\sup_{(a_1,a_2)\text{ in } I}(f_1+f_2)$.

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

$$\begin{aligned}\sup_{\substack{a_1^2+a_2^2=\delta\\(a_1,a_2)\text{ in } I}} f_1((a_1, a_2)) &= \sup_{0\leqq a_1\leqq\delta}\{\Phi(a_1+r_1/2)-\Phi(a_1-r_1/2)\}\\ &\quad\cdot\{\Phi(\sqrt{\delta-a_1^2}+r_2/2)-\Phi(\sqrt{\delta-a_1^2}-r_2/2)\}\\ &\leqq\{\Phi(c_1\delta+r_1/2)-\Phi(c_1\delta-r_1/2)\}\\ &\quad\cdot\{\Phi(r_2/2)-\Phi(-r_2/2)\}\ ,\end{aligned}$$

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

$$\begin{aligned}&\{\Phi(r_1/2)-\Phi(-r_1/2)\}\{\Phi(r_2/2)-\Phi(-r_2/2)\}\\&\quad-\{\Phi(c_1\delta+r_1/2)-\Phi(c_1\delta-r_1/2)\}\{\Phi(r_2/2)-\Phi(-r_2/2)\}\\&\quad=\{\Phi(r_2/2)-\Phi(-r_2/2)\}[\{\Phi(r_1/2)-\Phi(-r_1/2)\}\\&\qquad-\{\Phi(c_1\delta+r_1/2)-\Phi(c_1\delta-r_1/2)\}]\\&\quad=c_2(r_2)c_3(r_1, r_2, \delta)\qquad\text{(say)}\,.\end{aligned}$$

The gain in f_2 (which increases as a_1 and a_2 increase in I) is less than

$$\begin{aligned}&\sup_{(a_1,a_2)\text{ in }I}\Phi(a_1-d+r_1/2)\Phi(a_2-d+r_2/2)\\&\quad\leqq\sup_{(a_1,a_2)\text{ in }I}\Phi(a_1-d+\max(r_1, r_2))\Phi(a_2-d+\max(r_1, r_2))\\&\quad=\sup_{\substack{a_1+a_2=d\\a_1,a_2\geqq 0}}\Phi(a_1-d+\max(r_1, r_2))\Phi(a_2-d+\max(r_1, r_2))\\&\quad=\sup_{0\leqq a_1\leqq d}\Phi(a_1-d+\max(r_1, r_2))\Phi(-a_1+\max(r_1, r_2))\,.\end{aligned}$$

We will show that

$$(8)\qquad \lim_{d\to\infty}\sup_{0\leqq a_1\leqq d}\Phi(a_1-d+\max(r_1, r_2))\Phi(-a_1+\max(r_1, r_2))=0\,.$$

Thus, there will exist a $K(r_1, r_2, \delta)$ such that $d\geqq K(r_1, r_2, \delta)$ 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 (8) is equal to

$$(9)\qquad \lim_{d\to\infty}\sup_{0\leqq a_1\leqq d}\mathrm{P}[X\leqq a_1-d+\max(r_1, r_2),\ Y\leqq -a_1+\max(r_1, r_2)]\,,$$

which involves the probability in a certain rectangle in $\mathcal{R}^2$, as illustrated in Figure 2. Thus, (9) is less than or equal to the limit of the supremum of the probability to the left of the line $X+Y=-d+\max(r_1, r_2)$,

$$\begin{aligned}&\lim_{d\to\infty}\sup_{0\leqq a_1\leqq d}\mathrm{P}[X+Y\leqq -d+\max(r_1, r_2)]\\&\qquad=\lim_{d\to\infty}\mathrm{P}[X+Y\leqq -d+\max(r_1, r_2)]=0\,.\end{aligned}$$

THEOREM 3. *For* $\mu\in\Theta(\eta^*)=\{\mu\colon\ \mu\in\Theta,\ \mu_2-\mu_1\geqq\eta^*,\ \text{some }\eta^*>0\}$ *the MPE* (t_1, t_2) *is equivalent to the GMLE* $(\bar{X}_{[1]}, \bar{X}_{[2]})$, *found in Dudewicz* [2], *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 $\delta>0$,

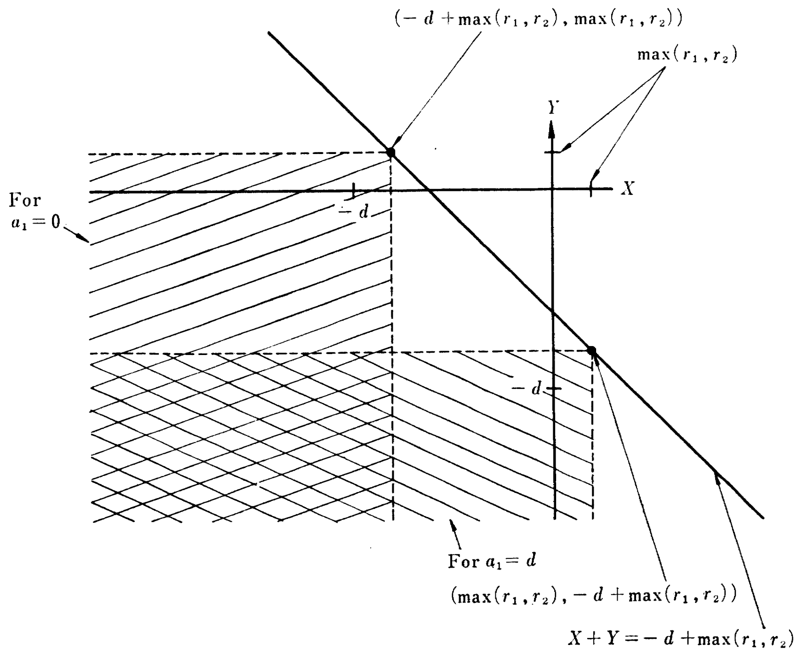

Fig. 2

$$\begin{aligned}1&=\lim_{n\to\infty} \mathrm{P}_\mu\,[k(n)\max(|t_1-\bar{X}_{[1]}|,\,|t_2-\bar{X}_{[2]}|)<\delta]\\&=\lim_{n\to\infty} \mathrm{P}_\mu\left[\frac{\sqrt{n}}{\sigma}\max(|a_1|\sigma/\sqrt{n},\,|a_2|\sigma/\sqrt{n})<\delta\right]\\&=\lim_{n\to\infty} \mathrm{P}_\mu\,[\max(a_1,a_2)<\delta]\ ,\end{aligned}$$

where the last equality uses Lemma 2. Now by Theorem A-1, the density of $d=(\sqrt{n}/\sigma)(\bar{X}_{[2]}-\bar{X}_{[1]})$ for $y\geqq 0$ is

$$\frac{1}{2\sqrt{\pi}}\left\{\exp\left(-\frac{1}{4}\left(y-\frac{\eta}{\sigma/\sqrt{n}}\right)^2\right)+\exp\left(-\frac{1}{4}\left(y+\frac{\eta}{\sigma/\sqrt{n}}\right)^2\right)\right\}$$

where $\eta=\mu_{[2]}-\mu_{[1]}$. Thus $\lim_{n\to\infty}\mathrm{P}_\mu\,[d\geqq K(r_1,r_2,\delta)]=1$, so using Lemma A-1,

$$\lim_{n\to\infty}\mathrm{P}_\mu\,[\max(a_1,a_2)<\delta]=\lim_{n\to\infty}\mathrm{P}_\mu\,[\max(a_1,a_2)<\delta\,|\,d\geqq K(r_1,r_2,\delta)]=1\ ,$$

where the last step uses Lemma 4.

3. Suggestions for future work

The present paper has concerned itself with showing that for $m\geqq 1$ parameters any GMLE is an MPE (Theorem 2 above, which extends a result Weiss and Wolfowitz gave for the case $m=1$), and with studying in detail MPE's for ranked means. In the latter case, and with $m=2$, it would be interesting to perform numerical comparisons with other estimators along the lines of Blumenthal and Cohen [1]. It would also be of interest to study the behavior of the MPE (t_1, t_2) (see Lemma 1) as n, r_1, r_2 change.

Acknowledgements

Thanks are especially due to Professor Lionel Weiss for his suggestions for and guidance of Dudewicz [2], which contains these results. Thanks are also due to a referee for his comments.

Appendix

The joint density of $\bar{X}_1, \cdots, \bar{X}_k$ is

$$f_{\bar{X}_1,\cdots,\bar{X}_k}(y_1,\cdots,y_k)=f_{\bar{X}_1}(y_1)\cdots f_{\bar{X}_k}(y_k) \qquad (y_i\in\mathcal{R};\ i=1,\cdots,k)$$

where $f_{\bar{X}_i}(\cdot)$ is the $N(\mu_i, \sigma^2/n)$ density function $(i=1,\cdots,k)$.

It is well-known (e.g. see Dudewicz [4]) that then the joint density of the ordered $\bar{X}_i$ $(i=1,\cdots,k)$, i.e. of $\bar{X}_{[1]}\leqq\cdots\leqq\bar{X}_{[k]}$, is

$$\text{(A-1)}\quad f_{\bar{X}_{[1]},\cdots,\bar{X}_{[k]}}(x_1,\cdots,x_k)$$

$$=\begin{cases}\sum_{\beta\in S_k} f_{\bar{X}_1,\cdots,\bar{X}_k}(x_{\beta(1)},\cdots,x_{\beta(k)})\,, & x_1\leqq\cdots\leqq x_k\\ 0\,, & \text{otherwise}\end{cases}$$

$$=\begin{cases}\sum_{\beta\in S_k} (\sqrt{n}/\sigma)^k\phi\left(\frac{x_{\beta(1)}-\mu_1}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-\mu_k}{\sigma/\sqrt{n}}\right)\,, & x_1\leqq\cdots\leqq x_k\\ 0\,, & \text{otherwise}\end{cases}$$

$$=\begin{cases}\sum_{\beta\in S_k} (\sqrt{n}/\sigma)^k\phi\left(\frac{x_{\beta(1)}-\mu_{[1]}}{\sigma/\sqrt{n}}\right)\cdots\phi\left(\frac{x_{\beta(k)}-\mu_{[k]}}{\sigma/\sqrt{n}}\right)\,, & x_1\leqq\cdots\leqq x_k\\ 0\,, & \text{otherwise}\,.\end{cases}$$

From the joint density of $\bar{X}_{[1]},\cdots,\bar{X}_{[k]}$ given at (A-1), we find that (for $x_1\leqq x_2$)

$$f_{\bar{X}_{[1]},\bar{X}_{[2]}}(x_1,x_2)$$

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

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

so that (for $y \geqq 0$), setting $\eta = \mu_{[2]} - \mu_{[1]}$,

$$\begin{aligned} f_{\bar{X}_{[2]}-\bar{X}_{[1]}}(y) &= \int_{-\infty}^{\infty} f_{\bar{X}_{[1]}, \bar{X}_{[2]}}(x, y+x)dx \\ &= \int_{-\infty}^{\infty} \frac{n}{2\pi\sigma^2}\left\{\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{y+x-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]\right)\right. \\ &\qquad \left. +\exp\left(-\frac{1}{2}\left[\left(\frac{x+y-\mu_{[1]}}{\sigma/\sqrt{n}}\right)^2+\left(\frac{x-\mu_{[2]}}{\sigma/\sqrt{n}}\right)^2\right]\right)\right\}dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi}\frac{\sqrt{n}}{\sigma}\left\{\exp\left(-\frac{1}{2}\left[x^2+\left(\frac{y-\eta}{\sigma/\sqrt{n}}+x\right)^2\right]\right)\right. \\ &\qquad \left. +\exp\left(-\frac{1}{2}\left[\left(x+\frac{y}{\sigma/\sqrt{n}}\right)^2+\left(x-\frac{\eta}{\sigma/\sqrt{n}}\right)^2\right]\right)\right\}dx . \end{aligned}$$

Since, via completing the square,

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}[(x+a)^2+(x+b)^2]\right)dx \\ &\qquad = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x+(a+b)/2}{1/\sqrt{2}}\right)^2\right)e^{-(a-b)^2/4}dx = \sqrt{\pi}\, e^{-(a-b)^2/4} , \end{aligned}$$

it follows that

THEOREM A-1. *With* $\eta = \mu_{[2]} - \mu_{[1]}$, *for* $y \geqq 0$

$$f_{\bar{X}_{[2]}-\bar{X}_{[1]}}(y) = \frac{\sqrt{n}}{2\sigma\sqrt{\pi}}\left\{\exp\left(-\frac{1}{4}\left(\frac{y-\eta}{\sigma/\sqrt{n}}\right)^2\right)+\exp\left(-\frac{1}{4}\left(\frac{y+\eta}{\sigma/\sqrt{n}}\right)^2\right)\right\} .$$

Let $\{A_n, n \geqq 1\}$ and $\{B_n, n \geqq 1\}$ be sequences of events on some probability space (which may depend on n).

LEMMA A-1. *If* $\lim_{n\to\infty} \mathrm{P}_n(B_n) = 1$, *then* (*if either of the following limits exists*) $\lim_{n\to\infty} \mathrm{P}_n(A_n B_n) = \lim_{n\to\infty} \mathrm{P}_n(A_n)$.

PROOF. Suppose $\lim_{n\to\infty} \mathrm{P}_n(B_n) = 1$. Then by taking limits in $\mathrm{P}_n(B_n) \leqq \mathrm{P}_n(A_n \cup B_n) \leqq 1$ we find $\lim_{n\to\infty} \mathrm{P}_n(A_n \cup B_n) = 1$, and hence $\lim_{n\to\infty} \{\mathrm{P}_n(B_n) - \mathrm{P}_n(A_n \cup B_n)\} = 0$. Taking limits in $\mathrm{P}_n(A_n B_n) = \mathrm{P}_n(A_n) + \{\mathrm{P}_n(B_n) - \mathrm{P}_n(A_n \cup B_n)\}$ yields our result.

THE OHIO STATE UNIVERSITY

REFERENCES

[1] Blumenthal, S. and Cohen, A. (1968). Estimation of the larger of two normal means, *J. Amer. Statist. Ass.*, **63**, 861-876.

[2] Dudewicz, E. J. (1969). Estimation of ordered parameters, *Technical Report*, No. 60, Department of Operations Research, Cornell University, Ithaca, New York.

[3] Dudewicz, E. J. (1970). Confidence intervals for ranked means, *Naval Research Logistics Quarterly*, **17**, 69-78.

[4] Dudewicz, E. J. (1971). Maximum likelihood estimators for ranked means, *Z. Wahrsch. verw. Geb.*, **19**, 29-42.

[5] Weiss, L. and Wolfowitz, J. (1966). Generalized maximum likelihood estimators, *Teor. Veroyat. Primen.*, **11**, 68-93.

[6] Weiss, L. and Wolfowitz, J. (1967a). Maximum probability estimators, *Ann. Inst. Statist. Math.*, **19**, 193-206.

[7] Weiss, L. and Wolfowitz, J. (1967b). Estimation of a density function at a point, *Z. Wahrsch. verw. Geb.*, **7**, 327-335.

[8] Weiss, L. and Wolfowitz, J. (1968). Generalized maximum likelihood estimators in a particular case, *Teor. Veroyat. Primen.*, **13**, 657-662.

[9] Weiss, L. and Wolfowitz, J. (1969). Maximum probability estimators with a general loss function, *Probability and Information Theory* (Vol. 89 of Lecture Notes in Mathematics), Springer-Verlag, Berlin, 232-256.