The Annals of Mathematical Statistics
1971, Vol. 42, No. 3, 976-990

UPPER AND LOWER POSTERIOR PROBABILITIES FOR TRUNCATED MEANS[1]

BY ROBERT KLEYLE
University of Massachusetts

1. Introduction and summary. In the recent literature a system of inference which leads to upper and lower posterior distributions based on sample data has been proposed by Dempster (1966, 1967, 1968). The purpose of this paper is to apply this method to the problem of finding the upper and lower probabilities that the mean of a distribution falls within a given interval, the only prior information being that the distribution is continuous. Included in the class of all continuous distributions are those having the property that $0 < F(y) < 1$ for all real y. For distributions of this type the approach used here leads to trivial results. The reason for this is discussed in Section 2. To circumvent this difficulty the mean of the truncated distribution is used. That is, for given ε_1 and ε_2 such that $0 < \varepsilon_1 < 1-\varepsilon_2 < 1$,

$$\mu(\varepsilon_1, \varepsilon_2) = \frac{1}{1-\varepsilon_1-\varepsilon_2}\int_{\xi_1}^{\xi_2} t dF(t) \tag{1.1}$$

where F is a distribution function and $\xi_1 < \xi_2$ are real constants such that $F(\xi_1) = \varepsilon_1$ while $F(\xi_2) = 1-\varepsilon_2$.

The device of truncating the mean is to some extent artificial and is unnecessary when the continuous distribution function is constant except on an interval of finite length. However, it does offer an approach to the more general (and in the author's opinion more interesting) problem where no restriction is placed on the interval on which the distribution function is neither zero nor one.

Section 2 contains a brief restatement of some of Dempster's basic definitions in the context of this particular problem along with an exposition of the type of reasoning required whenever this method is applied to problems involving truncated moments. Exact expressions for the upper and lower probabilities are derived in Section 3, and some asymptotic results are presented in Section 4. In particular, it is shown that if the upper and lower probabilities converge, they converge to the same limit. All relevant distribution theory is presented in an appendix.

2. The method of inference. The basic components of this inference system are a probability space $(X, \mathscr{F}, \mu)$ and a class $\mathscr{M}$ of measurable mappings of $(X, \mathscr{F})$ into the Borel subsets of the real line. These mappings along with initial probability

Received April 29, 1970; revised November 16, 1970.
[1] This research was supported in part by the Office of Naval Research, Contract No. 1866 (37) and by the National Institutes of Health, Grant No. 1-F1-GM-21, 289.

measure μ induce a class of univariate distribution functions $\mathscr{C} = \{F_\theta : \theta \in \Omega\}$, where θ is an indexing parameter and Ω the parameter space. The only restriction on class $\mathscr{M}$ is that it satisfy the two postulates given by Dempster (1966) which set up a one to one correspondence between $\mathscr{M}$ and $\mathscr{C}$ or equivalently between $\mathscr{M}$ and Ω. Assuming that X contains an infinite number of elements, the random sampling process consists of a drawing from the product space $(X^n, \mathscr{F}^n)$ governed by the product measure μ^n. The observer, however, identifies only the corresponding point $(y_1, \cdots, y_n)$ and must on the basis of these observations infer which mapping is the true mapping, or equivalently, which $\theta \in \Omega$ is the true θ.

Let $\mathscr{C}$ denote the class of all continuous univariate distributions, and let Ω be any set which can be put into one to one correspondence with the distributions in $\mathscr{C}$. Because of the well-known property of the transformation $x = F(y)$, where F is a continuous distribution function, there is no loss of generality in letting $(X, \mathscr{F}, \mu)$ represent the uniform probability measure over the Borel subsets of the unit interval $(0, 1)$.

Let Γ denote a multivalued mapping of X^n into Ω in accordance with Dempster's consistency principle (cf. Dempster (1966)). That is, each sample point $(x_1, \cdots, x_n) \in X^n$ maps into a subset $\Gamma(x_1, \cdots, x_n) \subset \Omega$ where $\Gamma(x_1, \cdots, x_n)$ consists of all points θ for which $y_i = F_\theta^{-1}(x_i)$, $i = 1, 2, \cdots, n$. Let $\mathscr{E}$ denote the class of all subsets of Ω of the type

$$\Sigma(\varepsilon_1, \varepsilon_2, \lambda_1, \lambda_2) = \{\theta : \lambda_1 < \mu(\varepsilon_1, \varepsilon_2) < \lambda_2\}, \tag{2.1}$$

where $0 < \varepsilon_1 < 1 - \varepsilon_2 < 1$ and $\lambda_1 < \lambda_2$. For any $\Sigma \in \mathscr{E}$ define

$$\bar{S}_n(\Sigma) = \{(x_1, \cdots, x_n) \in X^n, \Gamma(x_1, \cdots, x_n) \cap \Sigma \neq \phi\} \tag{2.2}$$

and

$$\underline{S}_n(\Sigma) = \{(x_1, \cdots, x_n) \in X^n, \Gamma(x_1, \cdots, x_n) \neq \phi, \Gamma(x_1, \cdots, x_n) \subset \Sigma\}. \tag{2.3}$$

Finally if both upper and lower inverse sets are Borel subsets of the n-dimensional unit cube X^n for all $\Sigma \in \mathscr{E}$, then upper and lower probabilities can be defined as

$$\bar{P}_n(\Sigma) = \mu^n(\bar{S}_n(\Sigma))/\mu^n(S_n) \tag{2.4}$$

and

$$\underline{P}_n(\Sigma) = \mu^n(\underline{S}_n(\Sigma))/\mu^n(S_n), \tag{2.5}$$

where $S_n = \bar{S}_n(\Omega) = \underline{S}_n(\Omega)$ is the domain of Γ and $\mu^n(S_n) > 0$. (cf. Dempster (1967)). Before proceeding further sets $\bar{S}_n(\Sigma)$ and $\underline{S}_n(\Sigma)$ must be explicitly identified. This presents special difficulties for events in class $\mathscr{E}$ as can be seen from the following discussion.

Since the order in which the data are observed is irrelevant when making inferences concerning means, it is sufficient to know only the ordered data $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$. Furthermore, since the transformation $x = F(y)$ is order-preserving for all $F \in \mathscr{C}$, attention can be restricted to the ordered subset

$$S_n = \{0 < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < 1\}. \tag{2.6}$$

Thus $\Gamma(x_1, \cdots, x_n) = \phi$ for all $(x_1, \cdots, x_n) \notin S_n$ and $S_n = \bar{S}_n(\Omega) = \underline{S}_n(\Omega)$ as required. Now consider an arbitrary point $(x_{(1)}, \cdots, x_{(n)}) \in S_n$. For any such point, $x_{(i)} = F_\theta(y_{(i)})$, $i = 1, 2, \cdots, n$ for some $\theta \in \Omega$. Thus the event $x_{(i)} < \varepsilon_1$ corresponds to the event $y_{(i)} < \xi_1$, while $x_{(j)} > 1-\varepsilon_2$ corresponds to $y_{(j)} > \xi_2$. From the foregoing it is clear that observations which lie in the tails (i.e., values below ξ_1 or above ξ_2) should not be involved in inferences concerning the truncated mean $\mu(\varepsilon_1, \varepsilon_2)$. But since the values of ξ_1 and ξ_2 are unknown, a conditional argument must be employed.

Let I^* denote the *largest* integer such that $x_{(I^*)} < \varepsilon_1$. If no such integer exists, let $I^* = 0$. Let J^* denote the *smallest* integer such that $x_{(J^*)} > 1-\varepsilon_2$. If no such integer exists, let $J^* = n+1$. Since the values of $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are unknown, I^* and J^* are random variables whose joint distribution is given by (3.23). Now for any given ε_1 and ε_2 if $S(I, J)$ denotes the subspace of S_n for which

$$(2.7)\quad x_{(1)} < \cdots < x_{(I)} < \varepsilon_1 < x_{(I+1)} < \cdots < x_{(J-1)} < 1-\varepsilon_2 < x_{(J)} < \cdots < x_{(n)},$$

then $(x_1, \cdots, x_n) \in S(I, J)$ if and only if $I^* = I$ and $J^* = J$. Notice also that

$$(2.8)\qquad S_n = \textstyle\bigcup_{I=0}^{n} \bigcup_{J=I+1}^{n+1} S(I, J),$$

and the components of the union are mutually disjoint. The problem of finding upper and lower inverse sets for each $\Sigma \varepsilon \mathscr{E}$ can now be attacked separately on each disjoint subset.

Let $(x_{(1)}, \cdots, x_{(n)})$ denote an arbitrary point in $S(I, J)$, and define $\Gamma(I, J) = \Gamma(I, J; x_{(1)}, \cdots, x_{(n)})$ to be that subset of Ω which is consistent with the data for the given sample point. That is, given data $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$, $\theta \varepsilon \Gamma(I, J; x_{(1)}, \cdots, x_{(n)})$ if and only if $x_{(i)} = F_\theta(y_{(i)})$ for all $i = 1, \cdots, n$. Therefore, if

$$(2.9)\qquad L(I, J) = L(I, J; x_{(1)}, \cdots, x_{(n)}) = \inf_{\theta \in \Gamma(I,J)} \textstyle\int_{\xi_1}^{\xi_2} t\, dF\theta(t),$$

and

$$(2.10)\qquad U(I, J) = U(I, J, x_{(1)}, \cdots, x_{(n)}) = \sup_{\theta \in \Gamma(I,J)} \textstyle\int_{\xi_1}^{\xi_2} t\, dF_\theta(t),$$

it is clear that $\Gamma(I, J) \cap \Sigma \neq \phi$ if and only if the intervals $(L(I, J), U(I, J))$ and $((1-\varepsilon_1-\varepsilon_2)\lambda_1, (1-\varepsilon_1-\varepsilon_2)\lambda_2)$ are *not* disjoint, while $\Gamma(I, J) \subset \Sigma$ if and only if the first interval is contained in the second. Thus when attention is restricted to subspace $S(I, J)$, the upper and lower probabilities are generated by random intervals. Some problems of this type havé been discussed by Dempster (1968).

Now if $T_*(I, J; \lambda)$ and $T^*(I, J; \lambda)$ are defined to be those subsets of $S(I, J)$ for which $(1-\varepsilon_1-\varepsilon_2)\lambda < L(I, J)$ and $(1-\varepsilon_1-\varepsilon_2)\lambda > U(I, J)$ respectively, then

$$(2.11)\qquad \bar{S}(I, J; \Sigma) = S(I, J) - (T_*(I, J; \lambda_2) \cup T^*(I, J; \lambda_1))$$

and

$$(2.12)\qquad \underline{S}(I, J; \Sigma) = T_*(I, J; \lambda_1) \cap T^*(I, J; \lambda_2).$$

Given the data $y_{(1)} < \cdots < y_{(n)}$, however, it can be shown that for some combinations of I, J, and λ, $T_*(I, J; \lambda)$ is the entire subspace $S(I, J)$ while for other combinations $T_*(I, J; \lambda)$ is the null set ϕ. The same is true of $T^*(I, J; \lambda)$. Thus if $y_{(p)} < \lambda_1 < y_{(p+1)}$ and $y_{(q)} < \lambda_2 < y_{(q+1)}$, where $0 \leqq p \leqq q \leqq n$,

$$\bar{P}_n(\Sigma) = 1 - \textstyle\sum_{I=1}^{n} \sum_{J=\max(I+1,\, q+2)}^{n+1} P(T_*(I, J; \lambda_2) \mid S(I, J)) P(S(I, J) \\ - \textstyle\sum_{I=0}^{p-1} \sum_{J=I+1}^{n} P(T^*(I, J; \lambda_1) \mid S(I, J)) P(S(I, J)) \tag{2.13}$$

and

$$\begin{aligned} \underline{P}_n(\Sigma) = & \textstyle\sum_{I=p+1}^{q-1} \sum_{J=I+1}^{q} P(S(I, J)) \\ & + \textstyle\sum_{I=1}^{p} \sum_{J=p+2}^{q} P(T_*(I, J; \lambda_1) \mid S(I, J)) P(S(I, J)) \\ & + \textstyle\sum_{I=p+1}^{q-1} \sum_{J=q+1}^{n} P(T^*(I, J; \lambda_2) \mid S(I, J)) P(S(I, J)) \\ & + \textstyle\sum_{I=1}^{p} \sum_{J=q+1}^{n} P(T_*(I, J; \lambda_1) \cap T^*(I, J; \lambda_2) \mid S(I, J)) P(S(I, J)), \end{aligned} \tag{2.14}$$

where probability measure P is defined such that $P(B) = \mu^n(B)$ for all Borel subsets $B \subset (0, 1)^n$ and probability measure μ^n is uniform over the Borel subsets of $(0, 1)^n$.

Explicit expressions for the component probabilities of (2.13) and (2.14) will be derived in the next section. For the moment, however, notice that if the range of the class of measurable mappings is the entire real line, the "non-truncated" mean is

$$\mu = \textstyle\int_{-\infty}^{\infty} t \, dF_\theta(t), \qquad \theta \in \Omega, \tag{2.15}$$

and $I = 0$ and $J = n+1$ are the only possible values of I and J. However, $T_*(0, n+1; \lambda) = T^*(0, n+1; \lambda) = \phi$ for all values of λ. Thus, for all $\lambda_1 < \lambda_2$ the upper and lower probabilities of events of the type $\lambda_1 < \mu < \lambda_2$ are one and zero respectively. This result also follows immediately from (2.13) and (2.14).

3. Exact upper and lower probabilities. In order to get explicit expressions for the upper and lower probabilities of events in class $\mathscr{E}$ it is first necessary to obtain explicit expressions for $L(I, J)$ and $U(I, J)$ which are given by the following lemma:

LEMMA 3.1. *Given sample point* $(x_{(1)}, \cdots, x_{(n)})$ *and data* $y_{(1)} < \cdots < y_{(n)}$,

$$L(I, J) = y_{(I)}(x_{(I+1)} - \varepsilon_1) + \textstyle\sum_{j=I+1}^{J-2} y_{(j)}(x_{(j+1)} - x_{(j)}) + y_{(J+1)}(1 - \varepsilon_2 - x_{(J-1)}), \tag{3.1}$$

$$U(I, J) = y_{(I+1)}(x_{(I+1)} - \varepsilon_1) + \textstyle\sum_{I+1}^{J-2} y_{(j+1)}(x_{(j+1)} - x_{(j)}) + y_{(J)}(1 - \varepsilon_2 - x_{(J-1)}) \tag{3.2}$$

when $J \geqq I+2$, *and*

$$L(I, I+1) = y_{(I)}(1 - \varepsilon_1 - \varepsilon_2), \tag{3.3}$$

$$U(I, I+1) = y_{(I+1)}(1 - \varepsilon_1 - \varepsilon_2). \tag{3.4}$$

PROOF. Suppose that $J \geqq I+2$ and consider the intervals $(\xi_1, y_{(I+1)}), (y_{(I+1)}, y_{(I+2)}), \cdots, (y_{(J-1)}, \xi_2)$ on R_1. Now for all $\theta \in \Gamma(I,J; x_{(i)}, \cdots, x_{(n)})$,

$$\int_{\xi_1}^{\xi_2} t\, dG_L(t) < \int_{\xi_1}^{\xi_2} t\, dF\theta(t) < \int_{\xi_1}^{\xi_2} t\, dG_U(t), \tag{3.5}$$

where G_L and G_U are step functions which put all the probability associated with the above intervals on their lower and upper end points respectively. Now since $\theta \in \Gamma(I, J)$ implies that $x_{(i)} = F_\theta(y_{(i)})$ for all $i = 1, 2, \cdots, n$, and that (2.7) holds, the corresponding intervals on (0, 1) are

$$(\varepsilon_1, x_{(I+1)}), (x_{(I+1)}, x_{(I+2)}), \cdots, (x_{(J-1)}1-\varepsilon_2).$$

Thus since the probability measure on (0, 1) is uniform,

$$\begin{aligned} G_U(y) &= \varepsilon_1 && \xi_1 \leqq y < y_{(I+1)} \\ &= x_{(j)} && y_{(j)} \leqq y < y_{(j+1)} \qquad j = I+1, \cdots, J-2, \\ &= x_{(J-1)} && y_{(J-1)} \leqq y \leqq \xi_2, \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} G_L(y) &= x_{(I+1)} && \xi_1 \leqq y \leqq y_{(I+1)} \\ &= x_{(j+1)} && y_{(j)} < y \leqq y_{(j+1)} \qquad j = I+1, \cdots, J-2. \\ &= 1-\varepsilon_2 && y_{(J-1)} < y \leqq \xi_2, \end{aligned} \tag{3.7}$$

Finally since $(x_{(1)}, \cdots, x_{(n)}) \in S(I, J)$ implies $y_{(I)} < \xi_1 < y_{(I+1)}$ and $y_{(J+1)} < \xi_2 < y_{(J)}$, the expression given by the right-hand side of (3.1) is the g.l.b. of $(1-\varepsilon_1-\varepsilon_2)\mu(\varepsilon_1, \varepsilon_2)$ for $\theta \varepsilon \Gamma_{(I, J)}$, while the right-hand side of (3.2) gives the l.u.b.

When $J = I+1$, there are no observations between ξ_1 and ξ_2, and

$$\int_{\xi_1}^{\xi_2} t\, dG_L(t) = \xi_1(1-\varepsilon_1-\varepsilon_2) > y_{(I)}(1-\varepsilon_1-\varepsilon_2), \tag{3.8}$$

while

$$\int_{\xi_1}^{\xi_2} t\, dG_U(t) = \xi_2(1-\varepsilon_1-\varepsilon_2) < y_{(I+1)}(1-\varepsilon_1-\varepsilon_2). \tag{3.9}$$

Expressions for the conditional probabilities which appear in lines (2.13) and (2.14) can now be derived by the following argument. Conditioning on subspace $S(I, J)$ is equivalent to conditioning on the event $\{I^* = I, J^* = J\}$. Now for given I and J set $T_*(I, J; \lambda)$ is defined in terms of $L(I, J)$, where $L(I, J)$ is given by (3.1) and (3.3). From (3.3) it is clear that $T_*(I, I+1; \lambda)$ is either $S(I, I+1)$ or ϕ depending on whether $\lambda \leqq y_{(I)}$ or $\lambda > y_{(I)}$. Now when $J \geqq I+2$, define

$$\begin{aligned} v_1 &= (x_{(I+1)}-\varepsilon_1)/(1-\varepsilon_1-\varepsilon_2), \\ v_i &= (x_{(I+i)}-x_{(I+i-1)})/(1-\varepsilon_1-\varepsilon_2), \quad i = 2, 3, \cdots, J-I-1. \end{aligned} \tag{3.10}$$

Thus,

$$T_*(I, J; \lambda) = \{\textstyle\sum_1^{J-I-1} (y_{(J-1)}-y_{(I+i-1)})v_i \leqq y_{(J-1)}-\lambda\}. \tag{3.11}$$

Now when $(x_1, \cdots, x_n) \in S(I, J)$, $x_{(I+1)} < x_{(I+2)} < \cdots < x_{(J+1)}$ are the order statistics of a random sample of size $J-I-1$ from the uniform distribution over $(\varepsilon_1, 1-\varepsilon_2)$. It follows, therefore, from (3.10) that given $S(I, J)$, $v_1, \cdots, v_{J-I-1}$ are jointly uniform on the simplex

$$V(I, J) = \{(v_1, \cdots, v_{J-I-1}) : v_i \geqq 0, i = 1, 2, \cdots, J-I-1, \textstyle\sum_1^{J-I-1} v_i \leqq 1\}. \tag{3.12}$$

Thus when $\lambda \geqq y_{(J-1)}$ the conditional probability given (I, J) of $T_*(I, J; \lambda)$ is zero, and when $y_{(p)} < \lambda < y_{(p+1)}$, $0 \leqq p \leqq J-2$, it is an immediate consequence of Theorem A.1 (cf. Appendix) that

$$P(T_*(I, J; \lambda) \mid S(I, J)) = A(I, J; p, \lambda) \tag{3.13}$$

where

$$A(I, J; p, \lambda) = 1 - \textstyle\sum_1^p [(\lambda - y_{(i)})^{J-I-1} / \prod_{1, j \neq i}^{J-1} (y_{(j)} - y_{(i)})] \tag{3.14}$$

or alternatively

$$A(I, J; p, \lambda) = \textstyle\sum_{p+1}^{J-1} [(\lambda - y_{(i)})^{J-I-1} / \prod_{1, j \neq i}^{J-1} (y_{(j)} - y_{(i)})]. \tag{3.15}$$

From (3.15) it is clear that if $p \geqq J-1$, $A(I, J; p, \lambda) = 0$, which is consistent with what happens when $\lambda \geqq y_{(J-1)}$.

By a completely analogous argument it can be shown that for $0 \leqq I < J \leqq n$ and $y_{(p)} < \lambda < y_{(p+1)}$, $(0 \leqq p \leqq n)$,

$$P(T^*(I, J; \lambda) \mid S(I, J)) = 1 - A(I+1, J+1; p, \lambda). \tag{3.16}$$

The intersection which defines $\underline{S}(I, J; \Sigma)$ can be written

$$T_*(I, J; \lambda_1) \cap T^*(I, J; \lambda_2) = \{\textstyle\sum_1^{J-I-1} (y_{(J-1)} - y_{(I+i-1)}) v_i \leqq y_{(J-1)} - \lambda_1, \quad \sum_1^{J-I-1} (y_{(J)} - y_{(I+i)}) v_i > y_{(J)} - \lambda_2\}. \tag{3.17}$$

Now if $y_{(p)} < \lambda_1 < y_{(p+1)}$ and $y_{(q)} < \lambda_2 < y_{(q+1)}$ where $I \leqq p \leqq q \leqq J-1$, Theorem A.2 is applicable with $r = p-I+1$ and $s = q-I$. The expression obtained from Theorem A.2 is a complicated function of $I, J, p, q, \lambda_1, \lambda_2$ and the ordered observations. It can be written in several alternate forms, only one of which will be given here. It is convenient to write this expression in terms of the following component functions:

$$\begin{aligned} \delta_{ik} &= (y_{(I+i)} - y_{(I+k)}) / (y_{(I+i-1)} - y_{(I+k-1)}) \quad \text{for } i \neq k, \\ \delta_{ii} &= (y_{(J)} - y_{(I+i)}) / (y_{(J-1)} - y_{(I+i-1)}), \\ \delta_i &= (y_{(I+i)} - \lambda_2) / (y_{(I+i-1)} - \lambda_1). \end{aligned} \tag{3.18}$$

When $\delta_i > 0$ let $\Delta(ij)$ denote the jth largest δ_{ik}, and m_i the integer such that $\Delta(im_i) \geqq \delta_i > \Delta(im_{i+1})$, and when $\delta_i < 0$. let $\Delta(ij)$ denote the jth smallest δ_{ik}, and m_i the integer such that $\Delta(im_i) \leqq \delta_i < \Delta(im_{i+1})$. Also define

$$b_i(\lambda : I, J) = (|\lambda - y_{(I+i-1)}|^{J-I-1} / \textstyle\prod_{1, j \neq i}^{J-I} |y_{(I+i-1)} - y_{(I+j-1)}|), \tag{3.19}$$

$$A_i^*(\delta_i, m_i : I, J) = 1 - \textstyle\sum_{j=1}^{m_i} \{(\Delta(ij) - \delta_i)^{J-I-1} \times [\Delta(ij) \prod_{1, k \neq j}^{J-I-1} (\Delta(ij) - \Delta(ik))]^{-1}\}, \tag{3.20}$$

and

$$B_i(\delta_i, m_i; I, J) = A_i^*(\delta_i, m_i: I, J) \qquad i \leqq s, \tag{3.21}$$
$$= 1 - A_i^*(\delta_i, m_i, I, J) \qquad i > s.$$

Finally, having specified all component functions,

$$P(T_*(I, J; \lambda_1) \cap T^*(I, J: \lambda_2) \mid S(I, J)) = 1 - A(I+1, J+1; q, \lambda_2) + \textstyle\sum_{i=1}^{r} (-1)^i b_i(\lambda_1; I, J) B_i(\delta_i, m_i, I, J), \tag{3.22}$$

whenever $y_{(p)} < \lambda_1 < y_{(p+1)}, y_{(q)} < \lambda_2 < y_{(q+1)}$, and $I \leqq p \leqq q \leqq J-1$. Notice from (2.14) that the conditional probability of the intersection occurs only when $I \leqq p \leqq q \leqq J-1$.

The weighting probabilities $P(S(I, J))$ which appear in (2.13) and (2.14) are clearly trinomial since $P(S(I, J))$ is simply the probability that in a random sample of size n from the uniform distribution over (0, 1) exactly I observations fall in the interval $(0, \varepsilon_1)$, $J-I-1$ in the interval $(\varepsilon_1, 1-\varepsilon_2)$, and the remaining $n-J+1$ in $(1-\varepsilon_2, 1)$. Thus,

$$P(S(I, J)) = \frac{n!}{I!(J-I-1)!\,(n-J+1)!} \varepsilon_1^{I}(1-\varepsilon_1-\varepsilon_2)^{J-I-1} \varepsilon_2^{n-J+1}. \tag{3.23}$$

Explicit formulae for the upper and lower probabilities of event Σ can now be obtained by substituting the expressions given by (3.13), (3.16), (3.22), and (3.23) into (2.13) and (2.14) respectively.

Now let clsss $\mathscr{E}$ be extended to include "one-sided" events of the types

$$\Sigma_a = \{\theta: \mu(\varepsilon_1, \varepsilon_2) < \lambda\}, \tag{3.24}$$

and

$$\Sigma_b = \{\theta: \lambda < \mu(\varepsilon_1, \varepsilon_2)\}. \tag{3.25}$$

Since $\Sigma \to \Sigma_a$ as $\lambda_1 \to -\infty$ and $\Sigma \to \Sigma_b$ as $\lambda_2 \to \infty$, and since $p \to 0$ and $q \to n$ as λ_1 and λ_2 approach $-\infty$ and $+\infty$ respectively, expressions for the upper probabilities of events Σ_a and Σ_b can be obtained from (2.13) by setting $p = 0$ and $q = n$. However, the lower probabilities of Σ_a and Σ_b *cannot* be determined by considering the respective limits of $\underline{P}_n(\Sigma)$ as $\lambda_1 \to -\infty$ and $\lambda_2 \to +\infty$. This is because the cases where $I = 0$ or $J = n+1$ contribute nothing to the lower probability of a two-sided event, but the case $I = 0$ contributes to the lower probability of Σ_a, while the case $J = n+1$ contributes to the lower probability of Σ_b. However, since Σ_a and Σ_b are complementary events, the lower probability of one can be found by subtracting the upper probability of the other from unity (cf. Dempster, (1967)).

If $\mathscr{C}$ were to denote the class of continuous distribution functions having the property that $F(y) = 0$ for all $\theta \in \Omega$ whenever $y \leqq a$, it is necessary to truncate only in the right-hand tail. This has the effect of fixing the value of I at zero and letting $y_{(0)} = a$. The conditional argument is still necessary, but now the conditioning is with respect to J only, and expressions for the upper and lower probabilities of one and two sided events can be derived in a manner completely analogous to the above approach. If $F_\theta(y) = 1$ for all $\theta \in \Omega$ whenever $y \geqq b$, truncation need take place only in the left-hand tail, J is fixed at $n+1$ and the conditioning is on I.

Finally, if $\mathscr{C}$ were to denote the subclass of continuous distributions for which $0 < F_\theta(y) < 1$ only when $y \in (a, b)$ for all $\theta \in \Omega$, the untruncated mean

$$\mu = \textstyle\int_a^b t\, dF_\theta(t) < \infty \quad \text{for all} \quad \theta \in \Omega. \tag{3.26}$$

Thus both truncation and the conditional argument are unnecessary and the derivation of the upper and lower probabilities is straightforward.

4. Approximate upper and lower probabilities. From (2.13) and (2.14) it is seen that the upper and lower probabilities of events in class $\mathscr{E}$ are weighted averages of certain conditional probabilities. Exact expressions for these conditional probabilities are extremely complicated functions of the ordered observations and parameters $\lambda_1 < \lambda_2$. It is desirable, therefore, to approximate these exact expressions with more tractable functions.

Recall that the weights are trinomial probabilities given by (3.23). Thus for any positive integer k it is clear that

$$P(J^*-I^* \leqq k) = \textstyle\sum\sum_{J-I \leqq k} P(S(I, J)) = B(n, 1-\varepsilon_1-\varepsilon_2; k-1), \tag{4.1}$$

where $B(n, p\colon x)$ denotes the cumulative binomial distribution function. From (4.1) it is clear that for small ε_1 and ε_2 (i.e., max $(\varepsilon_1, \varepsilon_2) \leqq .05$), $B(n, 1-\varepsilon_1-\varepsilon_2; k-1)$ will be close to zero unless k is relatively large with respect to n. Thus, since almost all the weight is concentrated on those values of I and J for which the difference $J-I$ is large, approximate conditional probabilities will be derived for large values of $J-I$.

From the definitions of $T_*(I, J\colon \lambda)$ and $T^*(I, J\colon \lambda)$ it follows that

$$\begin{aligned} T_*(I, J; \lambda) &= \{\lambda \leqq \textstyle\sum_1^{J-I} y_{(I+i-1)} v_i\}, \\ T^*(I, J; \lambda) &= \{\textstyle\sum_1^{J-I} y_{(I+i)} v_i \leqq \lambda\}, \end{aligned} \tag{4.2}$$

where random variables $v_1, v_2, \cdots, v_{J-I-1}$ (defined by (3.10)) are uniformly distributed over the simplex given by (3.12), and $v_{J-I} = 1 - \sum_1^{J-I-1} v_i$.
Now if

$$z(I, J) = \frac{(J-I)^{\frac{1}{2}}}{s(I, J)} \left[\sum_I^{J-1} y_{(i)} v_{i-I+1} - \bar{y}(I, J) \right], \tag{4.3}$$

where

$$\bar{y}(I, J) = \frac{1}{J-I} \sum_I^{J-1} y_{(i)}, \tag{4.4}$$

and

(4.5) $$s^2(I, J) = \frac{1}{J-I} \sum_I^{J-1} (y_{(i)} - \bar{y}(I, J))^2,$$

and if

(4.6) $$t(I, J: \lambda) = (J-I)^{\frac{1}{2}}(\lambda - \bar{y}(I, J))/s(I, J),$$

sets $T_*(I, J; \lambda)$ and $T^*(I, J; \lambda)$ can be rewritten

(4.7) $$T_*(I, J; \lambda) = \{t(I, J; \lambda) \leqq z(I, J)\},$$
$$T^*(I, J; \lambda) = \{z(I+1, J+1) \leqq t(I+1, J+1; \lambda)\}.$$

Now for all $I < J$, let

(4.8) $$Q(I, J) = \max_{I \leqq i \leqq J-1} y_{(i)}^2 / \sum_I^{J-1} y_{(i)}^2.$$

If $y_{(I)}^2 < y_{(J-1)}^2$,

(4.9) $$Q(I, J) = y_{(J-1)}^2 / \sum_I^{J-1} y_{(i)}^2 = 1 - \sum_I^{J-2} y_{(i)}^2 / \sum_I^{J-1} y_{(i)}^2.$$

Now

(4.10) $$\lim_{J-I\to\infty} \frac{1}{J-I} \sum_I^{J-1} y_{(i)}^2 = \lim_{J-1\to\infty} \frac{1}{J-I-1} \sum_I^{J-2} y_{(i)}^2$$
$$= \frac{1}{1-\varepsilon_1-\varepsilon_2} \int_{\xi_1}^{\xi_2} t^2\, dF(t) \neq 0,$$

and thus

(4.11) $$\lim_{J-I\to\infty} Q(I, J) = 0.$$

It can be easily shown that (4.11) also holds if $y_{(I)}^2 > y_{(J-1)}^2$. Thus condition (A.9) is satisfied and by Theorem A.3 (cf. Appendix) $z(I, J)$ converges in distribution to $N(0, 1)$ as $J-I \to \infty$. Thus when the difference $J-I$ is large,

(4.12) $$P(T_*(I, J; \lambda) \mid S(I, J)) \simeq 1 - \Phi(t(I, J; \lambda))$$

(4.13) $$P(T^*(I, J; \lambda) \mid S(I, J)) \simeq \Phi(t(I+1, J+1; \lambda)),$$

and for $\lambda_1 < \lambda_2$,

(4.14) $$P(T_*(I, J; \lambda_1) \cap T^*(I, J; \lambda_2) \mid S(I, J)) \simeq \Phi(t(I+1, J+1; \lambda_2)) - \Phi(t(I, J; \lambda_1)),$$

where Φ denotes the cumulative distribution function of the $N(0, 1)$ distribution.

The trinomial weight functions can be approximated by the product of independent Poissons. This follows since

(4.15) $$P(S(I, J)) = \binom{n}{I} \varepsilon_1^I (1-\varepsilon_1)^{n-I} \binom{n-I}{n-J+1} \left(\frac{\varepsilon_2}{1-\varepsilon_1}\right)^{n-J+1} \left(1 - \frac{\varepsilon_2}{1-\varepsilon_1}\right)^{J-I-1}$$

and since it is well known that the binomial distribution converges to the Poisson distribution as $n \to \infty$ and $np \to \lambda$. Thus for large n and small ε_1 and ε_2, each of the binomial terms of (4.15) can be approximated by Poisson probabilities, and

$$(4.16) \qquad P(S(I, J)) \simeq \exp(-n\varepsilon_1)\frac{(n\varepsilon_1)^I}{I!}\exp(-n\varepsilon_2/(1-\varepsilon_1))\frac{(n\varepsilon_2/(1-\varepsilon_1))^{n-J+1}}{(n-J+1)!}$$

Approximate upper and lower probabilities can now be computed for large n and small ε_1, ε_2 by determining which values of I and J carry essentially all of the weight and summing over these values only, replacing the exact probabilities of (2.13) and (2.14) with the approximations developed in this section.

The difficulty with this approach is that it attacks the problem in a piecemeal fashion. Although asymptotic theory is used to develop approximations for each piece of (2.13) and (2.14), the argument does not yield explicit limiting forms for either $\bar{P}_n$ or $\underline{P}_n$. Actually there would be only one limiting form since if either sequence converges, the other converges to the same limit.

THEOREM 4.1. *For events* $\Sigma \in \xi$,

$$\lim_{n\to\infty}[\bar{P}_n(\Sigma)-\underline{P}_n(\Sigma)] = 0.$$

PROOF. It follows immediately from Lemma 3.1 that for all values of $I < J$,

$$(4.17) \qquad U(I, J)-L(I, J) < \textstyle\sum_I^{J-1}(y_{(i+1)}-y_{(i)})(x_{(i+1)}-x_{(i)}) < M(I, J)(y_{(J)}-y_{(I)})$$

where $M(I, J) = \max_{I \leqq i < J}(x_{(i+1)}-x_{(i)})$.

Now since $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are the order statistics of a random sample from the uniform distribution over (0, 1), $x_{(i+1)}-x_{(i)}$ has the beta $(1, n)$ distribution for $i = 1, 2, \cdots, n-1$. Thus for all $\delta > 0$,

$$(4.18) \qquad P(x_{(i+1)}-x_{(i)} \geqq \delta) = (1-\delta)^n \quad \text{for} \quad i = 1, 2 \cdots n-1,$$

and $M(I, J)$ converges in probability to zero so long as $I \neq 0$ and $J \neq n+1$. It follows, therefore, from (2.9), (2.10) and (4.17) that for given $I \neq 0$ and $J \neq n+1$ both $L(I, J)$ and $U(I, J)$ converge in probability to $(1-\varepsilon_1-\varepsilon_2)\mu(\varepsilon_1, \varepsilon_2)$.

Now if $\Delta(I, J; \lambda)$ denotes that subset of $S(I, J)$ for which $L(I, J) < (1-\varepsilon_1-\varepsilon_2)\lambda < U(I, J)$, it follows from (2.11) and (2.12) that,

$$(4.19) \qquad \bar{S}(I, J\colon \Sigma) = \underline{S}(I, J\colon \Sigma)\cup(\Delta(I, J; \lambda_1)\cup\Delta(I, J; \lambda_2)).$$

For fixed $I \neq 0$ and $J \neq n+1$ notice that $(x_{(1)}, \cdots, x_{(n)}) \in \Delta(I, J; \lambda)$ implies that $L(I, J) < (1-\varepsilon_1-\varepsilon_2)\lambda < U(I, J)$ which, were it to hold in the limit, would imply that both $L(I, U)$ and $U(I, J)$ converge in probability to $(1-\varepsilon_1-\varepsilon_2)\lambda$. This leads to a contradiction. Thus

$$(4.20) \qquad \lim_{n\to\infty} P(\Delta(I, J; \lambda) \mid S(I, J)) = 0$$

for all $\lambda \neq \mu(\varepsilon_1, \varepsilon_2)$ and for all $I \neq 0, J \neq n+1$.

Finally from (2.13) and (2.14),

$$\bar{P}_n(\Sigma) - \underline{P}_n(\Sigma) = \sum_{I=0}^{n} \sum_{J=I+1}^{n+1} [P(\bar{S}(I, J : \Sigma) \mid S(I, J)) - P(\underline{S}(I, J : \Sigma) \mid S(I, J)] P(S(I, J))$$

$$\leq \max_{I \neq 0, J \neq n+1} \{P(\bar{S}(I, J : \Sigma) \mid S(I, J)) - P(\underline{S}(I, J ; \Sigma) \mid S(I, J)\} P(I \neq 0, J \neq n+1) + P(\{I = 0\} \cup \{J = n+1\}). \tag{4.21}$$

The fact that

$$\lim_{n \to \infty} \max_{I \neq 0, J \neq n+1} [P(\bar{S}(I, J : \Sigma) \mid S(I, J)) - P(\underline{S}(I, J : \Sigma) \mid S(I, J))] = 0 \tag{4.22}$$

for all $\Sigma \varepsilon \mathscr{E}$ follows immediately from (4.19) and (4.20). But

$$P(\{I = 0\} \cup \{J = n+1\}) = (1-\varepsilon_1)^n + (1-\varepsilon_2)^n - (1-\varepsilon_1-\varepsilon_2)^n \to 0 \tag{4.23}$$

as $n \to \infty$, and the proof is complete.

The prospect of finding explicitly a common limiting form for (2.13) and (2.14), or even the more modest prospect of proving that sequence $\bar{P}_n(\Sigma)$ (and thus $\underline{P}_n(\Sigma)$) converges, does not appear too encouraging because of the complicated way in which $A_n(I, J; P_n, \lambda)$ depends on the order statistics.

However, the large sample approximations, while less than completely satisfying, seem to give rather good results for moderately large samples. Table 4.1 gives both exact and approximate upper and lower probability distributions of $\mu(\varepsilon_1, \varepsilon_2)$ where $\varepsilon_1 = \varepsilon_2 = .025$ and the data consists of a random sample of 30 standard normal variates. The sample is drawn from the $N(0, 1)$ distribution for convenience only, since no distribution assumptions other than continuity are required. Because the upper and lower probability distributions are posterior distributions, the probabilities given in Table 4.1 will vary from sample to sample. Thus the only purpose of the table is to illustrate how well the approximations work.

TABLE 4.1

Exact Probabilities			*Approximate Probabilities*		
λ	upper	lower	λ	upper	lower
−1.0	.527	.037	−1.0	.473	.000
0	.594	.088	0	.606	.093
.1	.693	.171	.1	.703	.178
.2	.804	.279	.2	.806	.279
.3	.897	.385	.3	.846	.385
1.0	.999	.532	1.0	1.000	.535

APPENDIX

The first two theorems in this appendix pertain to the distribution theory of Section 3. Theorem A.3 is applied in Section 4.

THEOREM A.1. *Let $x_1, x_2, \cdots, x_n$ be random variables which are uniformly distributed over the n-dimensional simplex*

$$S_n = \{(x_1, \cdots, x_n): 0 \leqq x_i \text{ for all } i, \quad \text{and} \quad \textstyle\sum_1^n x_i \leqq 1\}, \tag{A.1}$$

and let $c_1, c_2, \cdots, c_n$ and c be real constants such that

$$c_1 > c_2 > \cdots > c_r \geqq c > c_{r+1} > \cdots > c_n > 0. \tag{A.2}$$

Then

$$\begin{aligned} P(\textstyle\sum_1^n c_i x_i \leqq c) &= 1 - \textstyle\sum_{j=1}^r \{(c_j - c)^n / c_j \prod_{1, i \neq j}^n (c_j - c_i)\} \\ &= c^n / \textstyle\prod_1^n c_i - \sum_{j=r+1}^n \{(c - c_j)^n / c_j \prod_{1, i \neq j}^n (c_i - c_j)\}. \end{aligned} \tag{A.3}$$

For a proof of this theorem the reader is referred to Dempster and Kleyle (1968).

Before proceeding to Theorem A.2 it is convenient to define a few terms which will be used in the statement of the theorem. Suppose that $c, c_1, c_2, \cdots, c_n$ and $d, d_1, d_2, \cdots, d_n$ are real constants, and let

$$\begin{aligned} \phi(j) &= (d - d_j)/(c - c_j), \\ \phi_i(j) &= (d_i - d_j)/(c_i - c_j) \quad \text{for} \quad i \neq j, \text{ and} \\ \phi_j(j) &= d_j / c_j. \end{aligned} \tag{A.4}$$

When $\phi(j) > 0$, let $\Delta_i(j)$ denote the ith *largest* $\phi_1(j), \phi_2(j), \cdots, \phi_n(j)$, and let m_j denote the integer for which $\Delta_{m_j}(j) \geqq \phi(j) > \Delta_{m_j+1}(j)$. When $\phi(j) < 0$, $\Delta_i(j)$ denotes the ith *smallest* $\phi_1(j), \phi_2(j), \cdots, \phi_n(j)$, and m_j the integer for which $\Delta_{m_j}(j) \leqq \phi(j) < \Delta_{m_j+1}(j)$.

Let

$$f_{nj}(\mathbf{x}) = |x_j - x|^n / x_j \textstyle\prod_{1, i \neq j}^n |x_j - x_i|, \tag{A.5}$$

where $\mathbf{x} = (x, x_1, x_2, \cdots, x_n)$, and

$$\begin{aligned} g_n(p, \mathbf{x}) &= \textstyle\sum_1^p f_{nj}(\mathbf{x}) \\ &= 1 - |x|^n / \textstyle\prod_1^n |x_i| + (|x|/x) \sum_{j=p+1}^n f_{nj}(\mathbf{x}). \end{aligned} \tag{A.6}$$

THEOREM A.2. *Suppose the hypotheses of Theorem* A.1 *hold and that*

$$d_1 > d_2 > \cdots > d_s \geqq d > d_{s+1} > \cdots > d_n > 0. \tag{A.7}$$

Then

$$\begin{aligned} P(\textstyle\sum_1^n c_i x_i \leqq c, \sum_1^n d_i x_i > d) &= g_n(s, \mathbf{d}) \\ &\quad + \textstyle\sum_{j=1}^s (-1)^j f_{nj}(\mathbf{c}) \{1 - g_n(m_j, \mathbf{\Delta}(j))\} \\ &\quad + \textstyle\sum_{j=s+1}^r (-1)^i f_{nj}(\mathbf{c}) g_n(m_j, \mathbf{\Delta}(j)), \end{aligned} \tag{A.8}$$

where $\mathbf{\Delta}(j) = (\phi(j), \Delta_1(j), \Delta_2(j), \cdots, \Delta_n(j))$.

The joint probability given by the left-hand side of (A.8) can be written in several alternate forms, but the expression given by the right-hand side of (A.8) is the simplest and is the one used in Section 3.

The proof of Theorem A.2 proceeds along the following lines. The probability that $\sum_1^n c_i x_i \leqq c$ is given by the right-hand side of (A.3). Each of the terms appearing in (A.3) represents (except for a factor of $n!$) the volume of an n-dimensional simplex. Imposing the additional condition $\sum_1^n d_i x_i > d$ partitions each of these simplices into two parts (one of which may be empty). The geometrical argument used by Dempster and Kleyle to prove Theorem A.1 is then applied to each of these partitioned simplices to find the desired volumes. Details of this argument, which are tedious but straightforward, are given by Kleyle (1967).

Before proceeding to Theorem A.3 two preliminary lemmas will be proved.

LEMMA A.1. *For every positive integer n, let $z_{n1}, z_{n2}, \cdots, z_{nn}$ be a sequence of independent, identically distributed random variables having the exponential density $f(x) = e^{-x}, x > 0$. Let $c_{n1}, c_{n2}, \cdots, c_{nn}$ be a sequence of real constants having the property that*

$$\lim_{n\to\infty}(\max_{1\leqq j\leqq n}|c_{nj}|)/(\textstyle\sum_1^n c_{nj}^2)^{\frac{1}{2}} = 0. \tag{A.9}$$

Then if

$$U_n = (\textstyle\sum_1^n c_{nj}^2)^{-\frac{1}{2}}\sum_1^n c_{nj}(z_{nj}-1) \tag{A.10}$$

and

$$V_n = \textstyle\sum_1^n (z_{nj}-1)/n^{\frac{1}{2}}, \tag{A.11}$$

and if

$$\rho = \lim_{n\to\infty}\textstyle\sum_1^n c_{nj}/(n\sum_1^n c_{nj}^2)^{\frac{1}{2}} \tag{A.12}$$

exists, (U_n, V_n) converges in distribution to the bivariate normal with zero means and convariance matrix $\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$.

Lemma A.1 is a special case of a more general result due to Chernoff, Gastwirth, and Johns (1967). Condition (A.9), however, is essentially Lindeberg's condition (cf. Hájek and Šidák, page 153), and Lemma A.1 can be easily proved by an argument similar to that used to establish the sufficiency of the Lindeberg condition.

PROOF. By expanding the principal value of $\ln(1-iy)$ in a power series, the natural logarithm of the joint characteristic function of (U_n, V_n) can be written

$$\ln\phi_n(t_1, t_2) = -\tfrac{1}{2}\textstyle\sum_1^n y_{nj}^2(t_1, t_2) + R_n(t_1, t_2), \tag{A.13}$$

where

$$y_{nj}(t_1, t_2) = t_1 c_{nj}(\textstyle\sum_1^n c_{nj}^2)^{-\frac{1}{2}} + t_2/n^{\frac{1}{2}} \tag{A.14}$$

and

$$R_n(t_1, t_2) = \textstyle\sum_{j=1}^n\sum_{k=3}^\infty (iy_{nj}(t_1, t_2))^k/k. \tag{A.15}$$

Now for fixed $M > 0$ let M^* denote that region of the Euclidean plane for which $\max(|t_1|, |t_2|) \leqq M$. Thus for $(t_1, t_2) \in M^*$

(A.16) $$|y_{nj}(t_1, t_2)| \leqq M[|c_{nj}|(\sum_1^n c_{nj}^2)^{-\frac{1}{2}} + 1/n^{\frac{1}{2}}],$$

which implies that $y_{nj}(t_1, t_2) \to 0$ uniformly in $(t_1, t_2) \in M^*$. Thus there exists a positive integer N independent of (t_1, t_2) such that whenever $(t_1, t_2) \in M^*$ and $n \geqq N$, $|y_{nj}(t_1, t_2)| < \frac{1}{2}$ for all $j \leqq n$, and

(A.17) $$|R_n(t_1, t_2)| \leqq \sum_{j=1}^n \sum_{k=3}^\infty |2y_{nj}(t_1, t_2)|^k/k2^k$$
$$< \sum_{j=1}^n |2y_{nj}(t_1, t_2)|^3 \sum_{k=3}^\infty \tfrac{1}{2}^k < \sum_{j=1}^n |2y_{nj}(t_1, t_2)|^3.$$

But from (A.16)

(A.18) $$\sum_{j=1}^n |2y_{nj}(t_1, t_2)|^3 \leqq (2M)^3 \sum_{j=1}^n [|c_{nj}|/(\sum_1^n c_{nj}^2)^{\frac{1}{2}} + 1/n^{\frac{1}{2}}]^3$$
$$\leqq (2M)^3 \max_{1 \leqq j \leqq n} |c_{nj}|^3/(\sum_1^n c_{nj}^2)^{\frac{3}{2}} + O(1/n^{\frac{1}{2}}).$$

Thus,

(A.19) $$\lim_{n\to\infty} R_n(t_1, t_2) = 0,$$

and

(A.20) $$\lim_{n\to\infty} \ln \phi_n(t_1, t_2) = -\tfrac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2).$$

LEMMA A.2. *Let $(U_1, V_1), (U_2, V_2), \cdots$ be a sequence of random vectors such that $n^{\frac{1}{2}}(U_n, V_n)$ converges in distribution to the bivariate normal with zero means, unit variances and covariance ρ. Let $(a_1, b_1), (a_2, b_2), \cdots$ be a sequence of constant vectors having the property that as $n \to \infty$ $(a_n, b_n) \to (a, b)$ where $a/b = \rho$. Then if*

(A.21) $$W_n = (a_n + U_n)/(b_n + V_n) - a_n/b_n$$

$n^{\frac{1}{2}} W_n$ converges in distribution to $N(0, \rho^2(1-\rho^2)/a^2)$.

PROOF.

(A.22) $$W_n = \rho_n(U_n^* - V_n^*)/(1 + V_n^*),$$

where $U_n^* = U_n/a_n$, $V_n^* = V_n/b_n$, and $\rho_n = a_n/b_n$. The limiting distribution of $n^{\frac{1}{2}}\rho_n(U_n^* - V_n^*)$ is $N(0, \rho^2(1-\rho^2)/a^2)$, and since $V_n^* \to_p 0$, $n^{\frac{1}{2}} W_n$ has the same asymptotic distribution as $n^{\frac{1}{2}}\rho_n(U_n^* - V_n^*)$.

THEOREM A.3. *Suppose that for each positive integer n, $x_{n1}, x_{n2}, \cdots, x_{n\,n+1}$ denotes a sequence of random variables uniformly distributed over the n-dimensional simplex*

(A.23) $$S_n = \{(x_1, \cdots, x_{n+1}) : x_i \geqq 0 \text{ for all } i = 1, 2, \cdots, n+1 \text{ and } \textstyle\sum_1^{n+1} x_i = 1\},$$

and that $c_{n1}, c_{n2}, \cdots, c_{n\,n+1}$ *denotes a sequence of real constants for which condition* (A.9) *holds. Then the asymptotic distribution of*

$$W_n = (\sum_1^{n+1} c_{ni}^2)^{-\frac{1}{2}}[(n+1)\sum_1^{n+1} c_{ni}x_{ni} - \sum_1^{n+1} c_{ni}] \tag{A.24}$$

is $N(0, 1-\rho^2)$ *where* ρ *is defined by* (A.12).

PROOF. For each n,

$$x_{ni} = z_{ni}/\sum_1^{n+1} z_{ni}, \tag{A.25}$$

where $z_{ni}, z_{n2}, \cdots, z_{n\,n+1}$ are independent, exponentially distributed random variables. Moreover,

$$W_n = (n+1)^{\frac{1}{2}}[(a_{n+1} + U_{n+1})/(1+V_{n+1}) - a_{n+1}], \tag{A.26}$$

where U_{n+1} and V_{n+1} are given by (A.10) and (A.11), and

$$a_{n+1} = \sum_1^{n+1} c_{ni}[(n+1)\sum_1^{n+1} c_{ni}^2]^{-\frac{1}{2}}. \tag{A.27}$$

Thus, the proof is a direct consequence of Lemmas A.1 and A.2.

Acknowledgment. The author wishes to thank A. P. Dempster for his advice and guidance during the preparation of this paper. The author also wishes to thank P. J. Bickel for pointing out the connection between Lemma A.1 and the Chernoff, Gastwirth, and Johns paper, and the referee for several helpful comments and suggestions.

REFERENCES

[1] CHERNOFF, H., GASTWIRTH, J. L. and JOHNS, M. V. JR. (1967). Asymptotic distribution of linear combinations of order statistics with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

[2] DEMPSTER, A. P. (1966). New methods for reasoning toward posterior distributions based on sample data. *Ann Math. Statist.* **37** 355–374.

[3] DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–339.

[4] DEMPSTER, A. P. (1968). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.* **39** 957–966.

[5] DEMPSTER, A. P. and KLEYLE, R. M. (1968). Distributions determined by cutting a simplex with hyerplanes. *Ann. Math. Statist.* **39** 1473–1478.

[6] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[7] KLEYLE, R. M. (1967). An application of the direct probability method of inference. Ph.D. thesis, Department of Statistics, Harvard Univ.