The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1943–1948

A NOTE ON DISTRIBUTION-FREE STATISTICAL INFERENCE WITH UPPER AND LOWER PROBABILITIES[1]

By R. J. Beran
University of California, Berkeley

Upper and lower probabilities needed for distribution-free inference about the parameters are found for a class of statistical models. The results improve upon those of a previous paper in that continuity of the underlying distribution function is not assumed.

1. Introduction. An experiment is performed, resulting in observations $\mathbf{x} = (x_1, x_2, \cdots, x_N)$. These observations are generated from parameters θ and realized random variables $\mathbf{e} = (e_1, e_2, \cdots, e_N)$ by the mapping

$$\mathbf{x} = T_\theta^{-1}\mathbf{e}. \tag{1.1}$$

Here $\{T_\theta : \theta \in \Omega\}$ is a known family of nonsingular transformations mapping R^N into R^N, Ω is the parameter space, and the components of $\mathbf{e}$ are realized values of N independent identically distributed random variables with common distribution function F on the real line. The θ, the F, and the realized $\mathbf{e}$ producing the observed $\mathbf{x}$ are all unknown.

In a previous paper (1971), the author treated the problem of inference about θ within the general framework of Dempster's (1966) upper and lower probabilities and risks. The analysis was performed under two alternative sets of assumptions on F: first that F is continuous, and secondly that F is both continuous and symmetric about the origin.

The aim of this note is to carry out a similar analysis for θ without, however, assuming continuity of F. While the results are more complicated than those of Beran (1971), removing the continuity assumption makes the model more realistic and permits theoretical treatment of ties among the observations.

2. No assumptions on F. It is supposed only that $F \in \mathscr{F}$, the family of all distribution functions on the real line. Let $F^{-1}(z) = \inf\{t : F(t) \geqq z\}$, let $U^N = \{\mathbf{u} \in R^N : 0 \leqq u_i \leqq 1\}$, and let $\mathbf{F}^{-1}$ be the function that maps $\mathbf{u} \in U^N$ into $(F^{-1}(u_1), F^{-1}(u_2), \cdots, F^{-1}(u_N))$. Equation (1.1) may be rewritten as

$$\mathbf{x} = T_\theta^{-1}\{\mathbf{F}^{-1}(\mathbf{u})\}, \tag{2.1}$$

where $\mathbf{u}$ is a realization of a random variable distributed uniformly over U^N. Performing the experiment described in the Introduction is viewed as the realization of a $\mathbf{u} \in U^N$, the selection in some unspecified fashion of a $\theta \in \Omega$ and a $F \in \mathscr{F}$, and the observation of a $\mathbf{x}$ related to $\mathbf{u}$, θ, F through (2.1).

Received November 10, 1970.

[1] This research was partially supported by National Science Foundation Grant GP–15283.

Let

$$S_x(\mathbf{u}) = \{(\theta, F) \in \Omega \times \mathscr{F} : \mathbf{x} = T_\theta^{-1}\{\mathbf{F}^{-1}(\mathbf{u})\}\}, \tag{2.2}$$

let

$$U_x = \{\mathbf{u} \in U^N : S_x(\mathbf{u}) \neq \varnothing\}, \tag{2.3}$$

and for any Lebesgue measurable subset $B \subset U_x$, let

$$P(B \mid U_x) = P(B)/P(U_x), \tag{2.4}$$

where P is the uniform probability measure on U^N. It is shown later in (2.13) that $P(U_x)$ exists and is nonzero. The upper and lower probabilities of a set $D \subset \Omega \times \mathscr{F}$ are defined as

$$\begin{aligned} P^*(D) &= P(\mathbf{u} : S_x(\mathbf{u}) \cap D \neq \varnothing \mid U_x) \\ P_*(D) &= P(\mathbf{u} : S_x(\mathbf{u}) \subset D, S_x(\mathbf{u}) \neq \varnothing \mid U_x) \end{aligned} \tag{2.5}$$

respectively, whenever the arguments on the right are Lebesgue measurable.

Suppose $\mathscr{D}$ is a space of decisions and $l : \Omega \times \mathscr{F} \times \mathscr{D} \to R^+$ is a loss function. The upper and lower risks incurred by a decision $d \in \mathscr{D}$ under the loss function l are defined as

$$\begin{aligned} R^*(l, d) &= \textstyle\int_0^\infty P^*(l(\theta, F, d) > z)\, dz \\ R_*(l, d) &= \textstyle\int_0^\infty P_*(l(\theta, F, d) > z)\, dz \end{aligned} \tag{2.6}$$

respectively, provided the integrals exist.

A frequency-based statistical rationale for considering these upper and lower probabilities and risks has been described in Beran (1970) and (1971). In the context of this rationale, there is a natural optimality property to be desired of a decision d: a decision $d \in \mathscr{D}$ is said to be *minimax* under loss function l if $R^*(l, d) \leqq R^*(l, d')$ for every $d' \in \mathscr{D}$. Thus, to make a minimax decision concerning θ alone, under a loss function which does not depend on F, it is enough to know $P^*(D)$, $P_*(D)$ for all sets of the form $D = A \times \mathscr{F}$, where $A \subset \Omega$.

For such sets D, the expressions (2.5) for $P^*(D)$ and $P_*(D)$ may be simplified. Let

$$\begin{aligned} W_x(\mathbf{u}) &= \mathrm{proj}_\Omega(S_x(\mathbf{u})) \\ &= \{\theta \in \Omega : \mathbf{x} = T_\theta^{-1}\{\mathbf{F}^{-1}(\mathbf{u})\} \quad \text{for at least one} \quad F \in \mathscr{F}\}. \end{aligned} \tag{2.7}$$

Evidently

$$U_x = \{\mathbf{u} \in U^N : W_x(\mathbf{u}) \neq \varnothing\} \tag{2.8}$$

and

$$\begin{aligned} P^*(A \times \mathscr{F}) &= P(\mathbf{u} : W_x(\mathbf{u}) \cap A \neq \varnothing \mid U_x) \\ P_*(A \times \mathscr{F}) &= P(\mathbf{u} : W_x(\mathbf{u}) \subset A, W_x(\mathbf{u}) \neq \varnothing \mid U_x). \end{aligned} \tag{2.9}$$

The following definitions and notations are helpful in simplifying (2.9).

DEFINITION 2.1. *If* $\mathbf{z} = (z_1, z_2, \cdots, z_N) \in R^N$, $\text{rank}(z_i)$ *is defined to be the number of components of* $\mathbf{z}$ *that are not larger than* z_i. *The notation* $\text{rank}(\mathbf{z})$ *denotes the vector* $(\text{rank}(z_1), \cdots, \text{rank}(z_N))$.

A rank vector is said to be proper if no two of its components are equal.

DEFINITION 2.2 *A rank vector* $r' = (r_1', r_2', \cdots, r_N')$ *is said to be associated with a proper rank vector* $\mathbf{r} = (r_1, r_2, \cdots, r_N)$ *if* $r_i' \geqq r_i$ *for* $1 \leqq i \leqq N$. (*Notation*: $\mathbf{r}' \sim \mathbf{r}$.) Thus, for example, the rank vectors (1, 2, 3), (2, 2, 3), (1, 3, 3), (3, 3, 3) are each associated with the proper rank vector (1, 2, 3). However, (2, 3, 3) is not associated with (1, 2, 3) because (2, 3, 3) is not a rank vector.

For any rank vector $\mathbf{r} \in R^N$, define $\Omega(\mathbf{r})$ as

$$\Omega(\mathbf{r}) = \{\theta \in \Omega: \text{rank}(T_\theta \mathbf{x}) = \mathbf{r}\}. \tag{2.10}$$

For any *proper* rank vector $\mathbf{r} \in R^N$, define $\omega(\mathbf{r})$ as

$$\omega(\mathbf{r}) = \textstyle\bigcup_{\mathbf{r}' \sim \mathbf{r}} \Omega(\mathbf{r}'). \tag{2.11}$$

Then, if $\text{rank}(\mathbf{u})$ is a proper rank vector,

$$W_x(\mathbf{u}) = \omega(\text{rank}(\mathbf{u})). \tag{2.12}$$

Indeed, if $\theta \in W_x(\mathbf{u})$, there exists a $F \in \mathscr{F}$ such that $T_\theta \mathbf{x} = \mathbf{F}^{-1}(\mathbf{u})$; hence, since F^{-1} is monotone increasing and $\text{rank}(\mathbf{u})$ is assumed to be proper, $\text{rank}(T_\theta \mathbf{x}) \sim \text{rank}(\mathbf{u})$. Since, trivially, $\theta \in \Omega(\text{rank}(T_\theta \mathbf{x}))$, it follows from (2.11) that $\theta \in \omega(\text{rank}(\mathbf{u}))$. Conversely, if $\theta \in \omega(\text{rank}(\mathbf{u}))$, there exists a rank vector $\mathbf{r} \sim \text{rank}(\mathbf{u})$ such that $\theta \in \Omega(\mathbf{r})$. Consequently, there exists a $F \in \mathscr{F}$ for which $T_\theta \mathbf{x} = \mathbf{F}^{-1}(\mathbf{u})$; hence $\theta \in W_x(\mathbf{u})$.

Let $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_M$ denote all proper rank vectors in R^N such that $\omega(\mathbf{r}_i) \neq \varnothing$. For $1 \leqq i \leqq M$, set $\omega_i = \omega(\mathbf{r}_i)$. The family of sets $\{\omega_i: 1 \leqq i \leqq M\}$ is a covering of Ω. From (2.8) and (2.12) follows

$$\begin{aligned} P(U_x) &= P(\mathbf{u}: \text{rank}(\mathbf{u}) \in \textstyle\bigcup_{i=1}^M \{\mathbf{r}_i\}) \\ &= M/N!. \end{aligned} \tag{2.13}$$

For $1 \leqq i \leqq M$ and $A \subset \Omega$, define

$$\begin{aligned} \delta^*(i, A) &= 1 \quad \text{if} \quad A \cap \omega_i \neq \varnothing \\ &= 0 \quad \text{otherwise} \\ \delta_*(i, A) &= 1 \quad \text{if} \quad \omega_i \subset A \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.14}$$

From (2.12)

$$\begin{aligned} P(\mathbf{u}: W_x(\mathbf{u}) \cap A \neq \varnothing) &= P(\mathbf{u}: \text{rank}(\mathbf{u}) \in \textstyle\bigcup_{i:\delta^*(i,A) \neq 0} \{\mathbf{r}_i\}) \\ &= (\textstyle\sum_{i=1}^M \delta^*(i, A))/N! \end{aligned} \tag{2.15}$$

and

(2.16) $$P(\mathbf{u}: W_x(\mathbf{u}) \subset A, W_x(\mathbf{u}) \neq \varnothing) = P(\mathbf{u}: \text{rank}(\mathbf{u}) \in \bigcup_{i:\delta_*(i,A)\neq 0} \{\mathbf{r}_i\}) = (\sum_{i=1}^{M} \delta_*(i,A))/N!$$

The last two expressions, together with (2.13) and (2.9), establish

PROPOSITION 2.1. *Let A be an arbitrary subset of Ω. Then under model* (2.1),

$$P^*(A \times \mathscr{F}) = M^{-1} \sum_{i=1}^{M} \delta^*(i,A)$$

$$P_*(A \times \mathscr{F}) = M^{-1} \sum_{i=1}^{M} \delta_*(i,A).$$

Let $\mathscr{F}_s$ and $\mathscr{F}_c$ denote, respectively, the class of all step distribution functions and the class of all continuous distribution functions on the real line. If in formulating the model associated with (2.1) the assumption $F \in \mathscr{F}$ is replaced by $F \in \mathscr{F}_s$, then Proposition 2.1 still holds, provided $A \times \mathscr{F}_s$ replaces $A \times \mathscr{F}$. On the other hand, if the assumption $F \in \mathscr{F}$ is replaced by $F \in \mathscr{F}_c$, Proposition 2.1 must be replaced by the corresponding result in Beran (1971). There are notable differences in application between Proposition 2.1 and its analogue in Beran (1971). (Compare the example in Section 4 with that in Section 3.1 of Beran (1971).)

3. Symmetry assumption on F. It is supposed now that F is symmetric about the origin. Let F_+ denote the distribution function of the random variable $|e_1|$, let F_+^{-1} denote its inverse, let $V^N = \{\mathbf{v} \in R^N: -1 \leqq v_i \leqq 1\}$, and let $\mathbf{G}^{-1}$ be the function that maps $\mathbf{v} \in V^N$ into $(\text{sign}(v_1)F_+^{-1}(|v_1|), \text{sign}(v_2)F_+^{-1}(|v_2|), \cdots, \text{sign}(v_N)F_+^{-1}(|v_N|))$. Here sign (z) is -1, 0, or 1 according to whether z is negative, zero, or positive. Since F is symmetric, (1.1) may be rewritten as

(3.1) $$\mathbf{x} = T_\theta^{-1}\{\mathbf{G}^{-1}(\mathbf{v})\},$$

where $\mathbf{v}$ is a realization of a random variable distributed uniformly over V^N and where $F_+ \in \mathscr{F}_+$, the family of all distribution functions with carrier set $[0, \infty]$.

An analogue of Proposition 2.1 can be stated for the model just described. Details of the derivation are omitted. The following definitions and notations are needed.

If $\mathbf{z} \in R^N$, let $|\mathbf{z}| = (|z_1|, |z_2|, \cdots, |z_N|)$ and $\text{sign}(\mathbf{z}) = (\text{sign}(z_1), \text{sign}(z_2), \cdots, \text{sign}(z_N))$. A sign vector is said to be proper if none of its components equal zero. The phrase *rank and sign vector pair* refers to any of the possible pairs $(\text{rank}(|\mathbf{z}|), \text{sign}(\mathbf{z}))$, where $\mathbf{z} \in R^N$.

DEFINITION 3.1. *A rank and sign vector pair* $(\mathbf{r}', \mathbf{s}')$ *is said to be associated with a proper rank and proper sign vector pair* $(\mathbf{r}, \mathbf{s})$ *if* $r_i' \geqq r_i$ *for* $1 \leqq i \leqq N$ *and if* $s_i' = s_i$ *whenever* $s_i' \neq 0$.

(*Notation*: $(\mathbf{r}', \mathbf{s}') \sim (\mathbf{r}, \mathbf{s})$.)

For any rank and sign vector pair $(\mathbf{r}, \mathbf{s}) \in R^N \times R^N$, define $\Omega_+(\mathbf{r}, \mathbf{s})$ as

(3.2) $$\Omega_+(\mathbf{r}, \mathbf{s}) = \{\theta \in \Omega: \text{rank}(|T_\theta \mathbf{x}|) = \mathbf{r}, \text{sign}(T_\theta \mathbf{x}) = \mathbf{s}\}.$$

For any *proper* rank and *proper* sign vector pair $(\mathbf{r}, \mathbf{s}) \in R^N \times R^N$, define $\omega_+(\mathbf{r}, \mathbf{s})$ as

$$\omega_+(\mathbf{r}, \mathbf{s}) = \bigcup_{(\mathbf{r}', \mathbf{s}') \sim (\mathbf{r}, \mathbf{s})} \Omega_+(\mathbf{r}', \mathbf{s}'). \tag{3.3}$$

Let $(\mathbf{r}_1, \mathbf{s}_1), (\mathbf{r}_2, \mathbf{s}_2), \cdots, (\mathbf{r}_{M_+}, \mathbf{s}_{M_+})$ denote all proper rank and proper sign vector pairs in $R^N \times R^N$ such that $\omega_+(\mathbf{r}_i, \mathbf{s}_i) \neq \varnothing$. For $1 \leqq i \leqq M_+$, set $\omega_i^{\ +} = \omega_+(\mathbf{r}_i, \mathbf{s}_i)$. Define $\delta^*(i, A)_+$ and $\delta_*(i, A)_+$ as in (2.14), replacing the $\{\omega_i : 1 \leqq i \leqq M\}$ by the $\{\omega_i^{\ +} : 1 \leqq i \leqq M_+\}$.

PROPOSITION 3.1. *Let A be an arbitrary subset of Ω. Then under model* (3.1),

$$P^*(A \times \mathscr{F}_+) = M_+^{\ -1} \textstyle\sum_{i=1}^{M_+} \delta^*(i, A)_+$$
$$P_*(A \times \mathscr{F}_+) = M_+^{\ -1} \textstyle\sum_{i=1}^{M_+} \delta_*(i, A)_+.$$

The remarks following Proposition 2.1 apply with obvious modifications.

4. Example: the two-sample location shift model. In this model, $\mathbf{x} = (x_1, \cdots, x_m, y_1, \cdots, y_n)$, $N = m+n$, $\theta = \mu$, $F \in \mathscr{F}$, and $T_\mu \mathbf{x} = (x_1, \cdots, x_m, y_1 - \mu, \cdots, y_n - \mu)$.

Let $a_1 < a_2 < \cdots < a_L$, where $L \leqq mn$, denote the distinct values among $\{d_{ij} = y_j - x_i, 1 \leqq i \leqq m, 1 \leqq j \leqq n\}$. A simple geometric argument shows that for this location-shift model, the nonvoid sets $\Omega(\mathbf{r})$ are, in fact, the $2L+1$ sets $(-\infty, a_1), (a_1, a_2), \cdots, (a_L, \infty), \{a_1\}, \{a_2\}, \cdots, \{a_L\}$.

The corresponding family of sets $\{\omega_i : 1 \leqq i \leqq M\}$ may be described as follows. Let

$$\begin{aligned} n_{xj} &= \text{number of } x\text{'s of rank } j \text{ in } \mathbf{x}, \quad (1 \leqq j \leqq m) \\ n_{yj} &= \text{number of } y\text{'s of rank } j \text{ in } \mathbf{y}, \quad (1 \leqq j \leqq n) \\ n_{ij} &= \text{number of components of rank } j \text{ in } T_{a_i}\mathbf{x}, \\ &\qquad\qquad (1 \leqq i \leqq L, 1 \leqq j \leqq N). \end{aligned} \tag{4.1}$$

Furthermore, let

$$\begin{aligned} C &= [\textstyle\prod_{j=1}^m n_{xj}!][\prod_{k=1}^n n_{yk}!] \\ C_i &= \textstyle\prod_{j=1}^N n_{ij}! \qquad (1 \leqq i \leqq L). \end{aligned} \tag{4.2}$$

Then the sets $\{\omega_i : 1 \leqq i \leqq M\}$ for this model consist of $(-\infty, a_1], [a_1, a_2], \cdots, [a_L, \infty)$ each repeated C times and of $\{a_i\}$ repeated $C_i - 2C$ times for each i, $1 \leqq i \leqq L$. Consequently, $M = \sum_{i=1}^L C_i - (L-1)C$.

Evidently, C_i/C is an integer for each i, $1 \leqq i \leqq L$. Therefore, in computing upper and lower probabilities from Proposition 2.1, it is possible and convenient to use, instead of the family $\{\omega_i : 1 \leqq i \leqq M\}$, just the subfamily consisting of $(-\infty, a_1], [a_1, a_2], \cdots, [a_L, \infty)$ and of $\{a_i\}$ repeated $D_i = C_i/C - 2$ times for each i, $1 \leqq i \leqq L$. Correspondingly, M is replaced by $K = \sum_{i=1}^L D_i + L + 1$.

As an example of how the upper and lower probabilities so found can be applied to inference, suppose it is desired to estimate μ under the loss function

$$\begin{aligned} l(\mu, d) &= |\mu - d| \qquad \text{if } |\mu - d| \leqq b \\ &= b \qquad\qquad\ \ \text{if } |\mu - d| > b, \end{aligned} \tag{4.3}$$

where $b > a_L - a_1$. Let $b_i = (a_i + a_{i+1})/2$ for $1 \leqq i \leqq L-1$ and let $b_0 = a_L - b$, $b_L = a_1 + b$. For the loss function (4.3) the upper risk incurred in estimating μ by d is

$$R^*(l, d) = K^{-1}[\sum_{j \neq i} |a_j - d| + \sum_{j=1}^{L} D_j |a_j - d| + 2b], \tag{4.4}$$

provided $d \in (b_{i-1}, b_i)$ for some i, $1 \leqq i \leqq L$. Otherwise

$$\begin{aligned} R^*(l, d) &> R^*(l, b_0) \quad \text{if} \quad d < b_0 \\ R^*(l, d) &> R^*(l, b_0) \quad \text{if} \quad d > b_L. \end{aligned} \tag{4.5}$$

It is easily verified that the values of d which minimize $R^*(l, d)$ are precisely the medians of a set B, which consists of $b_1, b_2, \cdots, b_{L-1}$ together with a_i repeated D_i times for each i, $1 \leqq i \leqq L$. In other words, any median of B is a minimax estimate of μ under the loss function (4.3). If the observations in each sample are all distinct and if the differences $\{d_{ij}: 1 \leqq i \leqq m, 1 \leqq j \leqq n\}$ are also all distinct, then $D_i = 0$ for $1 \leqq i \leqq L$. In this special case, $B = \{b_1, b_2, \cdots, b_{L-1}\}$ and therefore median $\{d_{ij}\}$, the well-known Hodges-Lehmann estimate for μ is minimax.

REFERENCES

Beran, R. J. (1971). On distribution-free statistical inference with upper and lower probabilities. *Ann. Math. Statist.* **42** 157–168.

Beran, R. J. (1970). Upper and lower risks and minimax procedures. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* To appear.

Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* **37** 355–374.