ASYMPTOTIC MINIMAX CHARACTER OF THE SAMPLE DISTRIBUTION FUNCTION AND OF THE CLASSICAL MULTINOMIAL ESTIMATOR

By A. Dvoretzky,[1] J. Kiefer,[1] and J. Wolfowitz[2]

Cornell University

0. Summary. This paper is devoted, in the main, to proving the asymptotic minimax character of the sample distribution function (d.f.) for estimating an unknown d.f. in $\mathfrak{F}$ or $\mathfrak{F}_c$ (defined in Section 1) for a wide variety of weight functions. Section 1 contains definitions and a discussion of measurability considerations. Lemma 2 of Section 2 is an essential tool in our proofs and seems to be of interest per se; for example, it implies the convergence of the moment generating function of G_n to that of G (definitions in (2.1)). In Section 3 the asymptotic minimax character is proved for a fundamental class of weight functions which are functions of the maximum deviation between estimating and true d.f. In Section 4 a device (of more general applicability in decision theory) is employed which yields the asymptotic minimax result for a wide class of weight functions of this character as a consequence of the results of Section 3 for weight functions of the fundamental class. In Section 5 the asymptotic minimax character is proved for a class of integrated weight functions. A more general class of weight functions for which the asymptotic minimax character holds is discussed in Section 6. This includes weight functions for which the risk function of the sample d.f. is not a constant over $\mathfrak{F}_c$. Most weight functions of practical interest are included in the considerations of Sections 3 to 6. Section 6 also includes a discussion of multinomial estimation problems for which the asymptotic minimax character of the classical estimator is contained in our results. Finally, Section 7 includes a general discussion of minimization of symmetric convex or monotone functionals of symmetric random elements, with special consideration of the "tied-down" Wiener process, and with a heuristic proof of the results of Sections 3, 4, 5, and much of Section 6.

1. Introduction and Preliminaries. Throughout this paper we shall denote by $\mathfrak{F}$ the class of all univariate d.f.'s and by $\mathfrak{F}_c$ the subclass of continuous members of $\mathfrak{F}$ (for the sake of definiteness, members of $\mathfrak{F}$ will be considered continuous on the right). Let R^n denote n-dimensional Euclidean space, and let G be any subspace of the space of all real-valued functions on R^1. For simplicity we assume $\mathfrak{F} \subset G$, although it is really only necessary that G contain the function S_n, defined below, for every $x^{(n)}$. Let B be the smallest Borel field on G such that every element of $\mathfrak{F}$ is an element of B and such that, for every positive integer k and all sets of real numbers $\{t_1, \cdots, t_k\}$ and $\{a_1, \cdots, a_k\}$ with $t_1 < t_2 < \cdots <$

Received May 31, 1955. Revised October 5, 1955.

[1] Research sponsored by the Office of Naval Research.

[2] The research of this author was supported in part by the United States Air Force under Contract No. AF18(600)-685 monitored by the Office of Scientific Research.

t_k, the set $\{g \mid g \in G; g(t_1) < a_1, \cdots, g(t_k) < a_k\}$ is in B. (Thus, we might have $G = \mathfrak{F}$ and B the Borel field generated by open sets in the common metric topology.) Let D_n be the class of all real-valued functions ϕ_n on $B \times R^n$ with the following properties: for each $x^{(n)} \in R^n$, $\phi_n(\cdot; x^{(n)})$ is a probability measure (B) on G; and for each $\Delta \in B$, $\phi_n(\Delta; \cdot)$ is Borel-measurable on R^n.

The problem which confronts the statistician may now be described. Let $X_1, \cdots, X_n$ be independently and identically distributed according to some d.f. F about which it is known only that $F \in \mathfrak{F}_c$ (or even $F \in \mathfrak{F}$). The statistician is to estimate F. Write $X^{(n)} = (X_1, \cdots, X_n)$. Having observed $X^{(n)} = x^{(n)} = (x_1, \cdots, x_n)$, the statistician uses the decision function ϕ_n as follows: a function $g \in G$ is selected by means of a randomization according to the probability measure $\phi_n(\cdot; x^{(n)})$ on G; the function g so selected (which need not be a member of $\mathfrak{F}$) is then the statistician's estimate of the unknown F. It is desirable to select a procedure ϕ_n which may be expected to yield a g which will lie close to the true F, whatever the latter may be; the term "close" will be made precise in succeeding sections. We note that the decision procedure ϕ_n^* which for each $x^{(n)}$ assigns probability one to the "sample d.f." S_n defined by

$$S_n(x) = (\text{number of } x_i \leqq x)/n$$

is a member of D_n.

We now turn (in this and the four succeeding paragraphs) to measure-theoretic considerations which are relevant to this paper. Our point of view is to waste as little space as possible on these considerations, since our results hold under any measurability assumptions which imply the meaningfulness of certain probabilities and integrals involving elements ϕ of D_n, and, in fact, our results hold even if these are interpreted as inner measures and integrals (which will be proper ones when $\phi = \phi_n^*$), as we shall now see.

In Sections 3, 4, and 6 we shall be concerned, for a given n, $\phi \in D_n$, $r > 0$, and $F \in \mathfrak{F}$, with the probability that, when the procedure ϕ_n is used and the X_i have d.f. F, the selected estimate g of F will satisfy the inequality

$$\sup_x | g(x) - F(x) | > r.$$

We shall denote this probability by

$$P_{F,\phi}\{\sup_x | g(x) - F(x) | > r\}. \tag{1.1}$$

It is clear when $\phi = \phi_n^*$ that this probability is well defined. This probability will also be meaningful if G is sufficiently regular; for example, if G consists of functions continuous on the right, the supremum in the displayed expression is unchanged if it is taken over rational x, and the probability in question is well defined. For our considerations it is not even necessary to restrict G in this way; we need not concern ourselves with questions of measurability of

$$\sup_x | g(x) - F(x) |,$$

since the optimal properties proved for ϕ_n^* hold if the supremum is taken only over the rationals (this last supremum is never greater than the supremum over all x and is equal to the latter when $g = S_n$). Thus, for arbitrary G and ϕ, the "probability" expression displayed above may be interpreted with the supremum taken over the rationals (or, alternately, as an inner measure, or as the infimum over all positive integers k and sets of real numbers $t_1, \cdots, t_k$ of

$$P_{F,\phi}\{\max_{1 \leqq i \leqq k} | g(t_i) - F(t_i) | > r\}).$$

In Sections 4 and 6 expressions such as

$$\int W(r)\, d_r P_{F,\phi} \left\{\sup_x | g(x) - F(x) | \leqq r\right\} \tag{1.2}$$

appear, the integral being taken over the nonnegative reals with $W \geqq 0$ and nondecreasing. The probability appearing here is to be interpreted as unity minus the probability previously displayed in (1.1), but the integral is to be interpreted as including a term $\gamma \lim_{r\to\infty} W(r)$ if $\gamma > 0$, where

$$\gamma = \lim_{r\to\infty} P_{F,\phi} \left\{\sup_x | g(x) - F(x) | > r\right\}.$$

In Sections 5 and 6 we will encounter such expressions as

$$r(F, \phi) = E_{F,\phi} \int W(g(t) - F(t), F(t))\, dF(t), \tag{1.3}$$

or such an expression with the first two symbols (operations) interchanged. Here $W(x, t)$ is defined for x real and $0 \leqq t \leqq 1$, is measurable (in the Borel sense on R^2), is nonnegative, and for each t is even in x and nondecreasing in x for $x \geqq 0$. $E_{F,\phi}$ is the operation of expectation when the procedure ϕ is used and the X_i have d.f. F. If $\phi = \phi_n^*$, $r(F, \phi)$ is clearly well defined. For other ϕ, any of a number of general assumptions on W and G will suffice to make the integral meaningful; for example, if W is continuous, $F \varepsilon \mathfrak{F}_c$, and G consists of functions continuous on the right, then the integral is determined by the values of g on the rationals, and $r(F, \phi)$ is meaningful. Weaker assumptions may be made, and, in fact, one could treat $r(F, \phi)$ as an inner integral (which is a proper integral when $\phi = \phi_n^*$) and still obtain the optimum properties for ϕ_n^* which are derived in this paper.

Finally, in Sections 3, 4, 5, and 6, the method of proof used involves integration of expressions such as (1.1), (1.2), and (1.3) with respect to probability measures ξ_{kn} on $\mathfrak{F}_c$. These ξ_{kn} will always be measures (B) and, in fact, will be of a very simple form. Sometimes the order of integration will be interchanged in these sections. If $\phi = \phi_n^*$, the above operations are all easily justified. For other ϕ these operations may be justified, as in the previous three paragraphs, by suitable regularity assumptions on G and W; or, again, the integrals in question may be considered as inner integrals.

2. Two Lemmas. In this section we shall state two lemmas (and a corollary to the second) which will be used to prove the results of Sections 3 and 4, respectively. Lemma 1 is due to Anderson [8], while Lemma 2 is derived from results of Smirnoff [9].

For any set S in R^n and any n-vector ρ, we write $S + \rho = \{x \mid x - \rho \varepsilon S\}$. Denote m-dimensional Lebesgue measure by μ_m. The case of Anderson's result which will be of use to us is the following:

LEMMA 1. *Let P be a (possibly degenerate)*[3] *normal probability measure on R^n with means zero, and let T be any convex body in R^n which is symmetric about the origin. Then $P(T) \geqq P(T + \rho)$ for all ρ.*

We shall also use (in Section 5) the trivial fact that the result of Lemma 1 holds for $n = 1$ when P is a normal probability measure truncated at $(-\beta, \beta)$ for $\beta > 0$. In Section 7 we shall mention briefly an application of the more general form of Lemma 1 given in [8].

Before stating Lemma 2, we shall introduce some notation. Let U denote the uniform d.f. (i.e., the d.f. whose density with respect to μ_1 is unity) on [0, 1], and write, for $r \geqq 0$,

$$G_n(r) = P_U\left\{\sup_{0 \leqq x \leqq 1} | S_n(x) - x | \leqq r/\sqrt{n}\right\},$$

$$G_{k,n}(r) = P_U\left\{\max_{1 \leqq i \leqq k} | S_n(i/(k+1)) - i/(k+1) | \leqq r/\sqrt{n}\right\},$$

$$G(r) = 1 - 2\sum_{m=1}^{\infty} (-1)^{m+1} e^{-2m^2r^2}, \tag{2.1}$$

$$H_n(r) = P_U\left\{\sup_{0 \leqq x \leqq 1} [S_n(x) - x] \leqq r/\sqrt{n}\right\},$$

$$H_{k,n}(r) = P_U\left\{\max_{1 \leqq i \leqq k} [S_n(i/(k+1)) - i/(k+1)] \leqq r/\sqrt{n}\right\},$$

$$H(r) = 1 - e^{-2r^2}.$$

Then

$$G_{k,n}(r) \geqq G_n(r) \quad \text{and} \quad H_{k,n}(r) \geqq H_n(r) \tag{2.2}$$

for all k, n, r. Moreover,

$$\lim_{k \to \infty} G_{k,n}(r) = G_n(r), \quad \lim_{k \to \infty} H_{k,n}(r) = H_n(r), \tag{2.3}$$

and ([1], [2], [3])

$$\lim_{k \to \infty}\lim_{n \to \infty} G_{k,n}(r) = \lim_{n \to \infty} G_n(r) = G(r), \quad \lim_{k \to \infty}\lim_{n \to \infty} H_{k,n}(r) = \lim_{n \to \infty} H_n(r) = H(r). \tag{2.4}$$

[3] The fact that the measure need not be n-dimensional necessitates only trivial modifications of the argument in [8].

We shall now prove the following:

LEMMA 2. *There exists a finite positive constant c such that*

$$1 - H_n(r) < ce^{-2r^2} \tag{2.5}$$

and

$$1 - G_n(r) < ce^{-2r^2} \tag{2.6}$$

hold for all $r \geqq 0$ *and all positive integers n.*

An immediate consequence is

COROLLARY 2. *If* $W(r)$ *is any nondecreasing nonnegative function defined for* $r > 0$, *then*

$$\lim_{n\to\infty} \int_0^\infty W(r)\, dH_n(r) = \int_0^\infty W(r)\, dH(r) \tag{2.7}$$

and

$$\lim_{n\to\infty} \int_0^\infty W(r)\, dG_n(r) = \int_0^\infty W(r)\, dG(r). \tag{2.8}$$

Indeed, the lim inf of the integral on the left side of (2.7) or (2.8) is always $\geqq$ the respective integral on the right side. Now, if $\int_0^\infty W(r)re^{-2r^2}\, dr = \infty$, then by (2.1), the integrals on the right side of (2.7) and (2.8) are both infinite and thus (2.7) and (2.8) hold in this case. If, on the other hand,

$$\int_0^\infty W(r)re^{-2r^2}\, dr < \infty,$$

then Corollary 2 follows from (2.4), (2.5), and (2.6), and in this case both sides of (2.7) and (2.8) are finite.

PROOF OF LEMMA 2. Since $1 - G_n(r) \leqq 2\,(1 - H_n(r))$, it suffices to prove (2.5). We shall deduce (2.5) from the explicit expression for $1 - H_n(r)$ given by Smirnoff [9]. Obviously, $1 - H_n(r) = 0$ for $r \geqq \sqrt{n}$, while for $0 < r < \sqrt{n}$, equation (50) of [9] asserts

$$1 - H_n(r) = (1 - r/\sqrt{n})^n + r\sqrt{n} \sum_{j=[r\sqrt{n}]+1}^{n-1} Q_n(j, r), \tag{2.9}$$

where $[x]$ denotes the greatest integer $\leqq x$ and

$$Q_n(j, r) = \binom{n}{j} (j - r\sqrt{n})^j (n - j + r\sqrt{n})^{n-j-1} n^{-n}. \tag{2.10}$$

In what follows we may, and do, restrict ourselves to $0 < r < \sqrt{n}$.

Taking logarithms and differentiating, it is seen that the maximum of $(1 - r/\sqrt{n})^n e^{2r^2}$ occurs at $r = 0$; hence,

$$\left(1 - \frac{r}{\sqrt{n}}\right)^n e^{2r^2} < 1. \tag{2.11}$$

A simple computation yields for all j with $r\sqrt{n} < j < n$,

$$\frac{d}{dr}\log Q_n(j, r) = \frac{-rn^2}{(j - r\sqrt{n})(n - j + r\sqrt{n})} - \frac{\sqrt{n}}{n - j + r\sqrt{n}}$$

$$< \frac{-4r}{1 - \frac{4}{n^2}\left(\frac{n}{2} - j + r\sqrt{n}\right)^2}$$

$$< -4r - \frac{16r}{n^2}\left(\frac{n}{2} - j + r\sqrt{n}\right)^2,$$

which on integrating gives

$$(2.12)\quad Q_n(j, r) < Q_n(j, 0)\exp\left[-2r^2 - \frac{8r^2}{n^2}\left(\frac{n}{2} - j + \frac{2r\sqrt{n}}{3}\right)^2 - \frac{4r^4}{9n}\right],$$

as well as

$$(2.13)\quad Q_n(j, r) < c_1 Q_n(j, 1)\exp\left[-2r^2 - \frac{8r^2}{n^2}\left(\frac{n}{2} - j + \frac{2r\sqrt{n}}{3}\right)^2 - \frac{4r^4}{9n}\right]$$

for $r \geqq 1$; here c_1 denotes a universal finite constant (and similarly, c_2, c_3, c_4, c_5 in the sequel).

We divide the sum of (2.9) into two parts: $\sum'$ will denote summation over those j for which

$$(2.14)\qquad \left|j - \frac{n}{2}\right| \leqq \frac{n}{4}$$

and $\sum''$ will denote summation over the remaining values. It follows immediately from Stirling's formula that

$$Q_n(j, 0) < c_2 n^{-3/2}$$

for j satisfying (2.14). Hence we have from (2.12),

$$\sum{}' Q_n(j, r) < \frac{c_2}{n^{3/2}} e^{-2r^2} \sum{}' \exp\left[-8r^2\left(\frac{1}{2} + \frac{2r}{3\sqrt{n}} - \frac{j}{n}\right)^2\right]$$

$$< \frac{2c_2}{n^{3/2}} e^{-2r^2} \sum_{j=0}^{\infty} e^{-8r^2 j^2/n^2}$$

$$< \frac{2c_2}{\sqrt{n}} e^{-2r^2}\left(\frac{1}{n} + \int_0^\infty e^{-8r^2t^2}\, dt\right)$$

$$< \frac{c_3}{r\sqrt{n}} e^{-2r^2}.$$

Hence,

$$(2.15)\qquad r\sqrt{n}\sum{}' Q_n(j, r) < c_3 e^{-2r^2}.$$

Let us now deal with the j occurring in $\sum''$, i.e., those for which (2.14) does not hold. If $2r\sqrt{n}/3 \leqq n/8$, then the second term in the exponent in (2.13) is $\leqq -(r^2/8)$ while otherwise $r > 3\sqrt{n}/16$ and the last term in the exponent in (2.13) is $< -(4/9)(3/16)^2r^2$. Thus, in both cases we have for $r > 1$,

$$Q_n(j, r) < c_1 Q_n(j, 1)e^{-2r^2}e^{-c_4r^2} < \frac{c_5}{r} Q_n(j, 1)e^{-2r^2}.$$

Hence we have from (2.9),

$$r\sqrt{n}\sum'' Q_n(j, r) < c_5 e^{-2r^2}\sqrt{n}\sum'' Q_n(j, 1) < c_5 e^{-2r^2}. \tag{2.16}$$

(2.11), (2.15), and (2.16) imply (2.5) for $1 < r < \sqrt{n}$ and thus obviously for all r.

3. Asymptotic minimax character of ϕ_n^* for a fundamental class of weight functions. In this section we shall prove the asymptotic minimax character of ϕ_n^* (as $n \to \infty$) in a sense which is fundamental in that the minimax character relative to all reasonable weight functions of a certain type will follow (in Section 4) from the results of the present section. We shall now prove the following strong property of ϕ_n^*:

THEOREM 3. *For every value $r > 0$,*

$$\lim_{n\to\infty} \frac{\sup_{F \varepsilon \mathfrak{F}_c} P_F\{\sup_x |S_n(x) - F(x)| > r/\sqrt{n}\}}{\inf_{\phi \varepsilon D_n} \sup_{F \varepsilon \mathfrak{F}_c} P_{F,\phi}\{\sup_x |g(x) - F(x)| > r\sqrt{n}\}} = 1. \tag{3.1}$$

In fact, the probability in the numerator of (3.1) is independent of F for $F \varepsilon \mathfrak{F}_c$ and is no greater for any $F \varepsilon \mathfrak{F} - \mathfrak{F}_c$ than for $F \varepsilon \mathfrak{F}_c$ (see [1]); as an immediate consequence of Theorem 3, we thus have

COROLLARY 3. *The result of Theorem 3 holds if $\mathfrak{F}_c$ is replaced by $\mathfrak{F}$ in its statement.*

We also remark that (3.9) and (3.20) below may be used to give an explicit bound on the departure of ϕ_n^* from minimax character; the integer N of (3.9) may be computed explicitly by merely keeping track of the constants which go into various error orders in the proof which follows; an explicit estimate of departure for $n \leqq N$ could be given similarly. With slightly more difficulty such a bound could also be computed in the cases treated in Sections 4, 5, and 6.

In order to prove (3.1), we shall exhibit a sequence $\{\xi_{kn}\}$ of a priori probability measures on $\mathfrak{F}_c$ such that, letting A_k (k a positive integer) denote the set consisting of the k points $i/(k+1)$ (for $1 \leqq i \leqq k$), we have

$$\begin{aligned}\lim_{k\to\infty}\lim_{n\to\infty}\inf_{\phi\varepsilon D_n} &\int P_{F,\phi}\{\sup_{a\varepsilon A_k}|g(a) - F(a)| > r/\sqrt{n}\}\, d\xi_{kn} \\ &= \lim_{k\to\infty}\lim_{n\to\infty}\int P_F\{\sup_{a\varepsilon A_k}|S_n(a) - F(a)| > r/\sqrt{n}\}\, d\xi_{kn} \\ &= \lim_{k\to\infty}\lim_{n\to\infty} P_U\{\sup_{a\varepsilon A_k}|S_n(a) - a| > r/\sqrt{n}\},\end{aligned} \tag{3.2}$$

where U is the uniform distribution on [0, 1]. Now, the expression under the limit operations on the left side of (3.2) is, for each n and k, obviously no greater than the denominator of (3.1) for the same n. On the other hand, the right side of (3.2) is equal to the (positive) limit as $n \to \infty$ of the numerator of (3.1), by (2.4). Hence, (3.2) implies (3.1).

In order to prove (3.2), we shall for each k limit ourselves to measures ξ_{kn} which assign probability one to distribution functions in $\mathfrak{F}_c$ of the form

$$F_k(x) = \sum_{i=1}^{k+1} p_i\, U_{ik}(x), \qquad p_i > 0, \sum p_i = 1, \tag{3.3}$$

where $U_{ik}(x)$ is the uniform probability distribution on the interval $[(i-1)/(k+1), i/(k+1)]$. For fixed k and n, it is easily seen that a sufficient statistic for the vector $\{p_i\}$ (and thus, for the family of F_k's of the form (3.3)) is given by the vector $T_k^{(n)} = \{T_{k1}^{(n)}, T_{k2}^{(n)}, \cdots, T_{k,k+1}^{(n)}\}$, where $T_{ki}^{(n)}$ is equal to the number of components of $X^{(n)}$ which lie in the interval $[(i-1)/(k+1), i/(k+1)]$. Hence, the validity of (3.2) will be implied by the following stronger result: Let B_k be the family of vectors $\pi = \{p_i, 1 \leqq i \leqq k+1\}$ satisfying $p_i \geqq 0$, $\sum p_i = 1$; $T_k^{(n)}$ has the multinomial distribution arising from n observations on $k+1$ types of objects, according to some $\pi \in B_k$, i.e., for integers $x_i \geqq 0$ with $\sum_1^{k+1} x_i = n$,

$$P_\pi\{T_{ki}^{(n)} = x_i, 1 \leqq i \leqq k+1\} = \frac{n!}{x_1! \cdots x_{k+1}!} p_1^{x_1} \cdots p_{k+1}^{x_{k+1}}; \tag{3.4}$$

$\mathcal{E}_n$ is the class of all (possibly randomized) vector estimators

$$\psi_n = \{\psi_{n1}, \cdots, \psi_{n,k+1}\}$$

of $\pi = \{p_i\}$ based on $T_k^{(n)}$ (ψ_n need not take on values in B_k); the ξ_{kn} are probability measures on B_k, which will be chosen so that

$$\begin{aligned}
&\lim_{n\to\infty} \inf_{\psi_n \in \mathcal{E}_n} \int P_\pi\left\{\sup_i \left| \sum_{j=1}^{i} (\psi_{nj} - p_j) \right| > r/\sqrt{n}\right\} d\xi_{kn} \\
&\quad = \lim_{n\to\infty} \int P_\pi\left\{\sup_i \left| \sum_{j=1}^{i} (T_{kj}^{(n)}/n - p_j) \right| > r/\sqrt{n}\right\} d\xi_{kn} \\
&\quad = \lim_{n\to\infty} P_{V_k}\left\{\sup_i \left| \sum_{j=1}^{i} (T_{kj}^{(n)}/n - 1/(k+1)) \right| > r/\sqrt{n}\right\},
\end{aligned} \tag{3.5}$$

where $V_k = \{1/(k+1), \cdots, 1/(k+1)\} \in B_k$. Taking limits as $k \to \infty$ (we have seen that this limit exists for the last expression of (3.5)), we see that the demonstration of (3.5) will imply that of (3.2). If we prove (3.5) with $\mathcal{E}_n$ replaced by the class of *non*randomized ψ_n, then (3.5) will a fortiori be true in the form stated above. Hence, in what follows, all ψ_n will be nonrandomized.

Some intuitive remarks are in order regarding the choice of ξ_{kn} (and the m_{kn} defining it) in the next paragraph. For simplicity, let us consider the case $k = 1$. We are then faced with a binomial estimation problem. The classical estimator of the parameter p_1 is asymptotically normal with maximum variance at $p_1 = \frac{1}{2}$ (this is V_1; in general, the corresponding phenomenon which concerns us occurs at $\pi = V_k$). In order to obtain our asymptotic Bayes result (3.5), we want ξ_{1n} to approximate a uniform measure on an interval of p_1 which has the following properties: on the one hand, the width e_n of this interval, when multiplied by $\sqrt{n}$, must tend to infinity with n; on the other hand, the width itself must tend to zero. In terms of the parameter $\sqrt{n}(p_1 - \frac{1}{2})$ and random variable $(T_{11}^{(n)} - n/2)/\sqrt{n}$, we will then be faced, asymptotically, with the problem of estimating the mean of a normal distribution (where, asymptotically, all real values are possible for the mean, with a uniform a priori distribution over a region whose width $\sqrt{n}e_n$ tends to ∞) with *almost constant variance*. The classical estimator will then be asymptotically Bayes for our weight function. Since a uniform a priori distribution would be slightly less simple to use (in keeping track of limits), we use instead one of the form (3.6) below; but the choice of the parameter m_{kn} therein is motivated by the remarks above.

Let $m = m_{k,n} =$ (greatest integer $\leqq n^{1/4}/k^2$), let $\epsilon = \epsilon_{k,n} = m/n$, and let $\xi_{k,n}$ be the probability measure on B_k which is given rise to by the probability density function

$$h_{k,n}(p_1, \cdots, p_k) = C_{k,n}\left[\left(1 - \sum_1^k p_i\right)\prod_{i=1}^k p_i\right]^m \tag{3.6}$$

with respect to Lebesgue measure on the k-simplex $\{0 \leqq \sum_1^k p_i \leqq 1, p_i \geqq 0 \ (1 \leqq i \leqq k)\}$ and is zero elsewhere. Here

$$C_{k,n} = \Gamma([m + 1][k + 1])/[\Gamma(m + 1)]^{k+1}.$$

Let $Y_{ki}^{(n)} = T_{ki}^{(n)}/n$. Let $\bar{\delta}_i = p_i - 1/(k + 1)$. The a posteriori density of $\bar{\delta}_1, \cdots, \bar{\delta}_k$, given that $Y_{ki}^{(n)} = y_i$ $(1 \leqq i \leqq k)$ (for possible values of the set $\{y_i\}$) when $\xi_{k,n}$ is the a priori probability measure on B_k is (the domain being obvious)

$$f_{k,n}(\delta_1, \cdots, \delta_k \mid y_1, \cdots, y_k) = \left[C_1 \prod_{i=1}^{k+1}\left(\delta_i + \frac{1}{k+1}\right)^{y_i+\epsilon}\right]^n, \tag{3.7}$$

where we have written $\delta_{k+1} = 1 - \sum_1^k \delta_i$ and $y_{k+1} = 1 - \sum_1^k y_i$ for typographical simplicity; here $(C_1)^n = \Gamma([m + 1][k + 1] + n)/\prod_1^{k+1}\Gamma(m + 1 + ny_i)$. Let $\bar{\eta}_i = \bar{\delta}_i - Y_{ki}^{(n)} + 1/(k + 1)$. Then the a posteriori density of $\bar{\eta}_1, \cdots, \bar{\eta}_k$ under the same conditions is (the domain again being obvious)

$$\begin{aligned} f_{k,n}^*(\eta_1, \cdots, \eta_k \mid y_1, \cdots, y_k) &= [g_{k,n}(\eta_1, \cdots, \eta_k \mid y_1, \cdots, y_k)]^n \\ &= \left[C_1 \prod_{i=1}^{k+1} (y_i + \eta_i)^{y_i+\epsilon}\right]^n, \end{aligned} \tag{3.8}$$

where $\eta_{k+1} = -\sum_1^k \eta_i$.

We shall now prove that, for each k and each r^* with $0 < r^* < \infty$, we have for $n > N(k, r^*)$ (the latter will be defined below)

$$(3.9)\qquad \begin{aligned} &E_t P_a^*\left\{\sup_i \left|\sum_{j=1}^{i} (p_j - Y_{kj}^{(n)})\right| \leqq \frac{r}{\sqrt{n}}\right\} \\ &\qquad \geqq E_t P_a^*\left\{\sup_i \left|\sum_{j=1}^{i} (p_j - \psi_{nj})\right| \leqq \frac{r}{\sqrt{n}}\right\} - n^{-1/9} \end{aligned}$$

for all r with $0 \leqq r \leqq r^*$ and all ψ_{ni} (not necessarily positive or summing to unity); here P_a^* denotes a posteriori probability of π (i.e., of $\{p_j\}$) when (3.6) is the a priori distribution, while E_t denotes expectation with respect to the measure on $B_k \times R^{k+1}$ given by (3.6) and (3.4). Noting that the second integral in (3.5) is unity minus the left side of (3.9) and that for each k the left side of (3.9) tends to a limit as $n \to \infty$ (this will follow from (3.20) below), we see that (3.9) actually implies that the first and second expressions of (3.5) are equal for each k. On the other hand, the limiting joint distribution function of the set of random variables $\{\sqrt{n}[Y_{ki}^{(n)} - 1/(k+1)], 1 \leqq i \leqq k\}$ under V_k is well known to be that whose density is given in (3.20), below, if we set all $y_i = 1/(k+1)$ and let $n \to \infty$ in the latter; since (3.20), which is the asymptotic a posteriori joint density of the $(p_i - T_{ki}^{(n)}/n)$, is continuous in the y_i, and since the $Y_{ki}^{(n)}$ tend in probability (according to (3.6) and (3.4)) to $1/(k+1)$ as $n \to \infty$, it follows that the second and third expressions of (3.5) are equal. (This last follows also from the continuity in π of $\lim_{n\to\infty} P_\pi\{\ \}$ in the second expression of (3.5) and the fact that $\lim_{n\to\infty} \xi_{kn}(J) = 1$ for any neighborhood J of V_k.) Thus, our theorem will be proved if we prove (3.9), and we now turn to this proof.

In this demonstration our calculations will be performed under the conditions

$$(3.10)\qquad \begin{aligned} &|y_i - 1/(k+1)| < 1/2(k+1) &&(1 \leqq i \leqq k+1), \\ &|\eta_i| < n^{-3/8}/4k(k+1) &&(1 \leqq i \leqq k+1), \\ &n > k^{40}. \end{aligned}$$

All orders $O(\cdot)$ will be uniform in the variables not appearing in the arguments. By (3.8),

$$(3.11)\qquad \log g_{k,n} = \log C_1 + \sum_1^{k+1} (y_i + \epsilon) \log y_i + \sum_1^{k+1} (y_i + \epsilon) \log\left(1 + \frac{\eta_i}{y_i}\right).$$

From (3.10), we have

$$(3.12)\qquad \left|\frac{\eta_i}{y_i}\right| < \frac{1}{2kn^{3/8}} \leqq \frac{1}{2},$$

and hence

$$\log\left(1 + \frac{\eta_i}{y_i}\right) = \frac{\eta_i}{y_i} - \frac{\eta_i^2}{2y_i^2} + \theta_i \frac{\eta_i^3}{y_i^3},$$

with

$$(3.13)\qquad |\theta_i| < 1, \qquad (1 \leqq i \leqq k+1).$$

Now, writing

$$(y_i + \epsilon) \log \left(1 + \frac{\eta_i}{y_i}\right) = \eta_i - \frac{\eta_i^2}{2y_i} + \theta_i \frac{\eta_i^3}{y_i^2} + \epsilon \frac{\eta_i}{y_i}\left(1 - \frac{\eta_i}{2y_i} + \theta_i \frac{\eta_i^2}{y_i^2}\right)$$

and remarking that $\sum_1^{k+1} \eta_i = 0$, that by (3.10) and (3.13)

$$\left|\sum_1^{k+1} \theta_i \frac{\eta_i^3}{y_i^2}\right| < (k+1) \frac{4(k+1)^2}{64k^3(k+1)^3 n^{9/8}} < \frac{1}{n^{9/8}},$$

and that by (3.10), (3.12), and the definition of ϵ

$$\epsilon \left| \sum_1^{k+1} \frac{\eta_i}{y_i}\left(1 - \frac{\eta_i}{2y_i} + \theta_i \frac{\eta_i^2}{y_i^2}\right)\right| < 2\epsilon \sum_1^{k+1} \left|\frac{\eta_i}{y_i}\right| < \frac{2}{k^2 n^{3/4}} \cdot \frac{k+1}{2kn^{3/8}} \leqq \frac{2}{n^{9/8}},$$

we obtain

$$\sum_1^{k+1} (y_i + \epsilon) \log \left(1 + \frac{\eta_i}{y_i}\right) = -\frac{1}{2} \sum_1^{k+1} \frac{\eta_i^2}{y_i} + \frac{3\theta}{n^{9/8}}, \tag{3.14}$$

with $|\theta| < 1$. Combining (3.14) and (3.11), we have

$$\log g_{k,n} = \log C_1 + \sum_1^{k+1} (y_i + \epsilon) \log y_i - \frac{1}{2} \sum_1^{k+1} \frac{\eta_i^2}{y_i} + O(n^{-9/8}). \tag{3.15}$$

Next, we note that

$$(C_1)^n \prod_1^{k+1} y_i^{ny_i+m} = p_{n+(k+1)m}^{(k)}(ny_1 + m, \cdots, ny_{k+1} + m; y_1, \cdots, y_{k+1}), \tag{3.16}$$

where $p_N^{(k)}(w_1, \cdots, w_{k+1}; q_1, \cdots, q_{k+1})$ is the (multinomial) probability that among N independent, identically distributed random variables taking on the value i with probability $q_i(\sum_1^{k+1} q_i = 1, q_i \geqq 0)$, there will be w_i taking on the value $i(\sum w_i = N)$. Using the familiar representation of this probability in terms of binomial probabilities, the definition of m, the inequalities (3.10), and the estimate for binomial probabilities

$$\begin{aligned} &p_N^{(1)}(Np + t\sqrt{Np(1-p)}, N(1-p) - t\sqrt{Np(1-p)}; p, 1-p) \\ &\qquad = [2\pi Np(1-p)]^{-1/2} e^{-t^2/2}[1 + O(N^{-1/2})] \end{aligned} \tag{3.17}$$

for $|t| < C_6$ and $|p - \frac{1}{2}| < C_7 < \frac{1}{2}$ (given in [5], p. 135), we obtain (with a conservative estimate of error)

$$(C_1)^n \prod_1^{k+1} y_i^{ny_i+m} = (1 + O(n^{-1/8}))(2\pi n)^{-k/2} \prod_1^{k+1} y_i^{-1/2}. \tag{3.18}$$

Hence, in the region (3.10) we obtain from (3.15) and (3.18), writing again $\eta_{k+1} = -\sum_1^k \eta_i$ and $y_{k+1} = 1 - \sum_1^k y_i$,

$$\begin{aligned} &f_{k,n}^*(\eta_1, \cdots, \eta_k \mid y_1, \cdots, y_k) \\ &\qquad = (1 + O(n^{-1/8}))(2\pi n)^{-k/2} \left(\prod_1^{k+1} y_i\right)^{-1/2} \exp\left(-\frac{n}{2} \sum_1^{k+1} \frac{\eta_i^2}{y_i}\right). \end{aligned} \tag{3.19}$$

For the corresponding a posteriori joint density of $\bar{\gamma}_i = \sqrt{n}\bar{\eta}_i$, $i = 1, \cdots, k$, in the region (3.10), we thus obtain (writing $\gamma_{k+1} = -\sum_1^k \gamma_i$)

$$(3.20) \qquad (1 + O(n^{-1/8}))(2\pi)^{-k/2}\left(\prod_1^{k+1} y_i\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_1^{k+1}\frac{\gamma_i^2}{y_i}\right).$$

Except for the first factor, this is a k-dimensional normal distribution centered at the origin. Note also that the probability assigned by this density to the complement of the region $|\bar{\eta}_i| < n^{-3/8}/4k(k+1)$ of (3.10) (for a single i) is (by Chebychev's inequality) $\leqq [1 + O(n^{-1/8})]O(k^4 n^{-1/4})$, so that the probability of the above inequality on the $\bar{\eta}_i$ for all i according to (3.20) (using $k < n^{1/40}$) is at least $1 - O(n^{-1/8})$. Also, the p_i or (3.6) have means $1/(k+1)$ and variances $O(m^{-1}k^{-2}) = O(n^{-1/4})$, while $Y_{ki}^{(n)}$ (given the p_i) has mean p_i and variance $O(n^{-1})$, whatever the p_i may be. Hence, for a single i, the probability (according to (3.6) and (3.4)) that $|Y_{ki}^{(n)} - 1/(k+1)| < \frac{1}{2}(k+1)$ is

$$\geqq P\{|p_i - 1/(k+1)| < \tfrac{1}{4}(k+1)\} \times P\{|Y_{ki}^{(n)} - p_i| < \tfrac{1}{4}(k+1) \,|p_i\}$$
$$\geqq 1 - k^2\{O(n^{-1/4}) + O(n^{-1})\}.$$

The probability that $|Y_{ki}^{(n)} - 1/(k+1)| < 1/2(k+1)$ for all i is thus

$$\geqq 1 - k^3 O(n^{-1/4}) \geqq 1 - O(n^{-1/8}).$$

We conclude, then, that the region of $Y_{ki}^{(n)}$, $\bar{\eta}_i$ $(1 \leqq i \leqq k+1)$ specified in (3.10) (putting $Y_{ki}^{(n)}$ for y_i and $\bar{\eta}_i$ for η_i), and hence where (3.20) holds, has probability $1 - O(n^{-1/8})$ according to (3.6) and (3.4).

Now, for fixed $r^* > 0$, let $N_1(k, r^*)$ be such that if $n > N_1(k, r^*)$, then

$$8r^* n^{-1/2} < n^{-3/8}/4k(k+1)$$

and the probability under (3.20) that all $|\bar{\eta}_i|$ are $< n^{-3/8}/16k(k+1)$ is $\geqq 1 - n^{-1/9}/2$; clearly, such a number $N_1(k, r^*)$ exists. For $0 < r \leqq r^*$, let T_r be the region where $|\sum_{j=1}^i \gamma_j| \leqq r$, $i = 1, \cdots, k+1$. Note that T_r is contained in the region where $|\gamma_j| \leqq 2r^*$ for all j. If ρ is any vector all of whose $(k+1)$ components are $\leqq n^{1/8}/8k(k+1)$ and if $n > N_1(k, r^*)$, then T_r and $T_r + \rho$ both lie entirely in the region of (3.10) (where (3.20) holds), whose probability according to (3.6) and (3.4) is $1 - O(n^{-1/8})$. Write C_r and D_r for the events in brackets on the left and right sides of (3.9), and define $L = L(X_1, \cdots, X_n, \psi_n)$ to be 1 or 0 according to whether or not

$$\max_j \sqrt{n}\,|Y_{kj}^{(n)} - \psi_{nj}| \leqq n^{1/8}/8k(k+1).$$

From the previous remarks of this paragraph and Lemma 1 we conclude that

$$(3.21) \qquad E_t[L \cdot P_a^*\{C_r\}] \geqq E_t[L \cdot P_a^*\{D_r\}] - n^{-1/9}/3$$

for $0 < r \leqq r^*$, $n > N(k, r^*)$, and all ψ_n, where $N(k, r^*)$ is chosen (as it clearly may be because $n^{-1/8} = o(n^{-1/9})$) to be enough larger than $N_1(k, r^*)$ to give the term $n^{-1/9}/3$ in (3.21). On the other hand, if any component of ρ has magnitude $> n^{1/8}/8k(k+1)$, then with probability $1 - O(n^{-1/8})$ according to

(3.4) and (3.6), $T_r + \rho$ has a posteriori probability $<n^{-1/9}/2$. Hence, the $N(k, r^*)$ above may clearly also be chosen so large that

$$E_t[(1 - L)P_a^*\{D_r\}] - 2n^{-1/9}/3 \leqq 0 \tag{3.22}$$

for $0 < r \leqq r^*$, $n > N(k, r^*)$, and all ψ_n. Equation (3.9) follows from (3.21) and (3.22), completing the proof of Theorem 3.

We remark that ϕ_n^* will not be minimax in the sense of Theorem 3 for all r and fixed finite n. The first nontrivial case is that of $n = 3$. A tiresome but straightforward computation in this case shows that, among the procedures ϕ_c which for a given number c ($0 \leqq c \leqq \frac{1}{2}$) assign probability one to

$$g_c(x) = \begin{cases} 0 & \text{if } x > Z_1, \\ c & \text{if } Z_1 \leqq x < Z_2, \\ 1 - c & \text{if } Z_2 \leqq x < Z_3, \\ 1 & \text{if } Z_3 \leqq x, \end{cases}$$

where the Z_i are the ordered $X_i^{(3)}$, the expression $P_U\{\sup_x| g_c(x) - x | \leqq z\}$ is maximized for $\frac{1}{6} \leqq z \leqq \frac{1}{3}$ at $c = \frac{1}{3}$ (i.e., by ϕ_3^*), for $\frac{1}{3} \leqq z \leqq \frac{1}{2}$ by $c = z$, and for $\frac{1}{2} < z \leqq 1$ by any $c \geqq 1 - z$ (for $z \leqq \frac{1}{6}$, all values of c give probability zero). Similar remarks apply to the problems considered in the next three sections. For example, $E_U\{\sup_x| g_c(x) - x |\}$ in the above example is minimized by $c = [33 - 3(17)^{1/2}]/52 = 0.397$. Similar calculations are more easily made in the case studied in Section 4 (where the distribution of the maximum deviation need not be calculated), and such calculations may be found in the reference cited at the end of that section.

4. Other loss functions which are functions of distance. In this section we show that the asymptotic minimax character of ϕ_n^* proved in Section 3 may be extended to a broad class of weight functions. It turns out that it is unnecessary to start anew in order to prove this; the class of weight functions considered in Section 3 (see below) is the basic class in the sense that the minimax character relative to many other weight functions may be concluded from the results of Section 3 and the integrability result given in Corollary 2. It is clear that the method of attack used here, i.e., of carrying out the detailed proof of the minimax character for the basic class of weight functions and then extending to other weight functions, can be stated as a general theorem to apply to other statistical problems; we shall not bother to state this obvious extension in a general setting.

Throughout this section W will represent any nonnegative function defined on the nonnegative reals which is nondecreasing in its argument, not identically zero (the case $W \equiv 0$ is trivial), and which satisfies

$$\int_0^\infty W(r)re^{-2r^2}\,dr < \tag{4.1}$$

The main result of this section is the following:

THEOREM 4. *Under the above assumptions on* W,

$$\lim_{n\to\infty} \frac{\sup_{F\varepsilon\mathfrak{F}_c} \int W(r)\, d_r P_F\{\sup_x | S_n(x) - F(x) | < r/\sqrt{n}\}}{\inf_{\phi\varepsilon D_n} \sup_{F\varepsilon\mathfrak{F}_c} \int W(r)\, d_r P_{F,\phi}\, \{\sup_x | g(x) - F(x) | < r/\sqrt{n}\}} = 1. \tag{4.2}$$

As in Section 3 (and for the same reason), an immediate corollary is

COROLLARY 4. *The result of Theorem 4 holds if* $\mathfrak{F}_c$ *is replaced by* $\mathfrak{F}$ *in its statement.*

PROOF OF THEOREM 4. By a reduction like that of Section 3, it is seen that (4.2) will be proved if, for the sequence $\{\xi_{kn}\}$ of Section 3, we can prove the following three statements, (4.3), (4.4), and (4.5):

$$\begin{aligned} &\lim_{n\to\infty} \inf_{\phi\varepsilon D_n} \int W(r)\, d_r P_{F,\phi}\, \{\max_{a\varepsilon A_k} | g(a) - F(a) | < r/\sqrt{n}\}\, d\xi_{kn} \\ &\quad = \lim_{n\to\infty} \int W(r)\, d_r P_F\, \{\max_{a\varepsilon A_k} | S_n(a) - F(a) | < r/\sqrt{n}\}\, d\xi_{kn} \end{aligned} \tag{4.3}$$

for each positive integer k;

$$\begin{aligned} &\lim_{n\to\infty} \int W(r)\, d_r P_F\, \{\max_{a\varepsilon A_k} | S_n(a) - F(a) | < r/\sqrt{n}\}\, d\xi_{kn} \\ &\quad = \lim_{n\to\infty} \int W(r)\, d_r P_U\, \{\max_{a\varepsilon A_k} | S_n(a) - a | < r/\sqrt{n}\} \end{aligned} \tag{4.4}$$

for each positive integer k;

$$\begin{aligned} 0 &< \lim_{k\to\infty} \lim_{n\to\infty} \int W(r)\, d_r P_U\, \{\sup_{a\varepsilon A_k} | S_n(a) - a | < r/\sqrt{n}\} \\ &= \lim_{n\to\infty} \int W(r)\, d_r P_U\, \{\sup_{0\leqq x\leqq 1} | S_n(x) - x | < r/\sqrt{n}\} < \infty. \end{aligned} \tag{4.5}$$

(This includes, of course, proving the existence of the indicated limits.)

Firstly, (4.5) is an immediate consequence of (4.1), (2.4), (2.2), the continuity of G and of the d.f. $\lim_{n\to\infty} G_{k,n}$, and of Corollary 2.

In order to prove (4.4), we note first that, for fixed k and any $F \varepsilon \mathfrak{F}_c$, we have (similarly to (2.2)) the inequality $P_F\{\max_{a\varepsilon A_k} | S_n(a) - F(a) | \leqq r/\sqrt{n}\} \geqq G_n(r)$. Hence, by Corollary 2, the integral with respect to r on the left side of (4.4) is bounded uniformly in n and F. On the other hand, given any $\epsilon > 0$, there exists an integer N_0 such that, for $n > N_0$, ξ_{kn} assigns probability at least $1 - \epsilon$ to a set of F for which the expressions $P_F\{\quad\}$ and $P_U\{\quad\}$ of (4.4) differ by less than ϵ for all r (this rests on the continuity in π, for π in a neighborhood of V_k, of the normal approximation (for large n) to the joint distribution of the random variables $\sqrt{n}(Y_{ki}^{(n)} - p_i)$, $1 \leqq i \leqq k$). Since $P_U\{\quad\}$ is continuous in r, (4.4) follows.

Finally, we must prove (4.3). Consider any fixed k. Write $P_n^*(r; x^{(n)}, \phi)$ for the probability, calculated according to the a posteriori probability distribution of π (given that $X^{(n)} = x^{(n)}$ and when ξ_{kn} is the a priori probability measure on B_k) and the probability measure $\phi(\cdot\,; x^{(n)})$ on G (where $\phi \,\varepsilon\, D_n$ and perhaps $\phi = \phi_n^*$) of the set of (g, π) in $G \times B_k$ for which $\max_{a \varepsilon A_k}|g(a) - F(a)| < r/\sqrt{n}$. If (4.3) is false, there exists a value $\epsilon > 0$ such that, for every positive N, there is an $n > N$ and a $\phi_n \,\varepsilon\, D_n$ for which (the operation E_t being as defined in Section 3)

$$(4.6)\quad E_t \int W(r)\, d_r P_n^*(r; X^{(n)}, \phi_n) < E_t \int W(r)\, d_r P_n^*(r; X^{(n)}, \phi_n^*) - 2\epsilon.$$

It is clear from the preceding paragraphs that there is a real number $q > 0$ such that $W(q) > 0$ and

$$(4.7)\qquad E_t \int_q^\infty W(r)\, d_r P_n^*(r; X^{(n)}, \phi_n^*) < \epsilon$$

for all n. Write $W_q(r) = \min(W(r), W(q))$. Then (4.6) and (4.7) imply

$$(4.8)\qquad E_t \int_0^\infty W_q(r)\, d_r\, \{P_n^*(r; X^{(n)}, \phi_n) - P_n^*(r; X^{(n)}, \phi_n^*)\} < -\epsilon.$$

Since $W_q(r) \leqq W(q)$, the integral on the left side of (4.8) is $\geqq -W(q)$. Hence, (4.8) implies that, with probability at least $\epsilon/W(q)$ (under (3.6) and (3.4)), $X^{(n)}$ will be such that

$$(4.9)\qquad \int_0^\infty W_q(r)\, d_r\, \{P_n^*(r; X^{(n)}, \phi_n) - P_n^*(r; X^{(n)}, \phi_n^*)\} < -\epsilon.$$

Let $\epsilon' = \epsilon/2W(q)$. The discussion of the previous paragraph shows that we can find an R^* and M such that, for $n > M$, the probability (under (3.6) and (3.4)) will be $>1 - \epsilon'$ that $X^{(n)}$ will be such that

$$(4.10)\qquad P_n^*(R^*; X^{(n)}; \phi_n^*) > 1 - \epsilon'.$$

Let $\gamma_n = \sup_r\{P_n^*(r; X^{(n)}, \phi_n) - P_n^*(r; X^{(n)}, \phi_n^*)\}$. We shall show below that (4.9) implies

$$(4.11)\qquad \gamma_n > \epsilon'.$$

Then (4.10) and (4.11) (the latter of which is an event of probability at least $2\epsilon'$ according to (3.6) and (3.4)) will imply that for each $N > M$ there is an $n > N$ and a $\phi_n \,\varepsilon\, D_n$ for which, with probability $>\epsilon'$ according to (3.6) and (3.4), $X^{(n)}$ will be such that

$$(4.12)\qquad \{P_n^*(r; X^{(n)}, \phi_n) - P_n^*(r; X^{(n)}, \phi_n^*)\} > \epsilon'$$

for some r with $0 \leqq r < R^*$ (here r depends on n, $\phi^{(n)}$, $X^{(n)}$). This contradicts the fact that, with probability $1 - O(n^{-1/8})$ according to (3.6) and (3.4), the region T_r of the last paragraph of Section 3 was seen to maximize with respect to ρ (uniformly in $0 \leqq r \leqq R^*$), to within an (added) error of $O(n^{-1/9})$, the a posteriori probability of $T_r + \rho$. Thus, it remains only to prove that (4.9) implies (4.11). For fixed n, ϕ_n, $X^{(n)}$, abbreviate the bracketed expression in (4.9) as $B(r) - C(r)$. Let

$$B^*(r) = \begin{cases} 0 & \text{if } r \leqq 0. \\ \min (C(r) + \gamma_n, 1) & \text{if } r > 0. \end{cases} \tag{4.13}$$

Clearly, $B(r) \leqq B^*(r)$. Hence, since $W_q(r)$ is nondecreasing in r, we have

$$\int W_q(r)\, dB(r) \geqq \int W_q(r)\, dB^*(r). \tag{4.14}$$

Let α be the infimum of values r for which $B^*(r) = 1$. From (4.9), (4.14), and the fact that $B^*(r) - C(r)$ is constant for $0 < r < \alpha$, we obtain

$$\begin{aligned} \epsilon &< \int_{0-}^{\infty} W_q(r)\, d(C(r) - B(r)) \\ &\leqq \int_{0-}^{\infty} W_q(r)\, d(C(r) - B^*(r)) \\ &\leqq \int_{0+}^{\alpha-} W_q(r)\, d(C(r) - B^*(r)) + W(0)\gamma_n + W(q)\gamma_n \\ &\leqq 2W(q)\gamma_n, \end{aligned} \tag{4.15}$$

which proves (4.11) and thus completes the proof of Theorem 4.

5. Integral weight functions. In this section we consider weight functions W_n^* arising from integration of a function W in the following manner:

$$W_n^*(F, g) = \int_{-\infty}^{\infty} W(\sqrt{n}[g(x) - F(x)], F(x))\, dF(x). \tag{5.1}$$

Here $W(y, z)$, which is defined for y real and $0 \leqq z \leqq 1$, is nonnegative and is symmetric in y and nondecreasing in y for $y \geqq 0$; it may be thought of as a measure of the contribution to W_n^* arising from a deviation of $y\sqrt{n}$ of the estimator g from the true F at an argument x for which $F(x) = z$. Typical W's which might be of interest are $W(y, z) = |y|^p$ or 0 according to whether or not $a \leqq z \leqq b$ (here $p > 0$ and $0 \leqq a < b \leqq 1$), $W(y, z) = 0$ or 1 according to whether $|y| \leqq a$ or $|y| > a$ where a is a suitably chosen constant, $W(y, z) = y^2/z(1 - z)$, etc.

We now turn to considerations of the asymptotic minimax character of ϕ_n^* with respect to a sequence of risk functions $r_n(F, \phi) = E_{F,\phi}W_n^*(F, g)$, where $\phi \in D_n$. (The remainder of the present paragraph will be somewhat heuristic in order to compare the present problem with those of Sections 3 and 4; the statement and proof of Theorem 5 begin in the next paragraph.) These considerations are much easier than those of the previous two sections, since in obtaining a

Bayes solution with respect to the a priori probability measure ξ_{kn} of Section 3 it will suffice (as will be seen below) to minimize with respect to ϕ, *for each fixed* x (more precisely, for each irrational x),

$$(5.2)\quad r_{kn}(x, \phi, t_k^{(n)}) = \int_{B_k} E_\phi\, W(\sqrt{n}[g(x) - F(x, \pi)], F(x, \pi))\, d_\pi\, \xi_{kn}^*(\pi; x, t_k^{(n)});$$

here B_k is as in Section 3, $F(x, \pi)$ denotes the distribution function of (3.3) for a given value of $\pi = (p_1, \cdots, p_{k+1})$, and for any measurable subset B of B_k we set

$$(5.3)\qquad \xi_{kn}^*(B, x, t_k^{(n)}) = \frac{\int_B f(x, \pi) P_\pi\, \{t_k^{(n)}\}\, d\xi_{kn}(\pi)}{\int_{B_k} f(x, \pi) P_\pi\, \{t_k^{(n)}\}\, d\xi_{kn}(\pi)},$$

where ξ_{kn} is given by (3.6) of Section 3, $f(x, \pi) = dF(x, \pi) / dx$ (this derivative exists for x irrational), and $P_\pi\{t_k^{(n)}\} = P_\pi\{T_{ki}^{(n)} = t_{ki}^{(n)}, 1 \leqq i \leqq k + 1\}$ is the probability function defined in (3.4). (Of course, ϕ in (5.2) may randomize over many g, which accounts for the presence of the E_ϕ operation.) Thus, present considerations will involve only the obtaining of a (univariate) normal approximation to the a posteriori distribution (more precisely, to a slight modification (5.3) of it) of $F(x, \pi)$ *for fixed irrational* x, which is much easier than the multivariate approximation (3.20) which it was necessary to obtain in Section 3. (We shall actually use (3.20), which implies easily the needed univariate approximation; however, the latter could have been obtained more easily directly.) The above remarks will be made precise in what follows. We hereafter denote the infimum of $r_{kn}(x, \phi, t_k^{(n)})$ over all ϕ in D_n by $r_{kn}^*(x, t_k^{(n)})$. The set of reals

$$\{z \mid 0 < z < \epsilon \text{ or } 1 - \epsilon < z < 1\}$$

will be denoted by I_ϵ for $0 < \epsilon < \frac{1}{2}$.

We now state Theorem 5. Our statement of this theorem is not the most general possible. (The set I_ϵ may be replaced by other sets where $W(y, z)$ is large, the continuity conditions on W may be weakened by considering continuous approximations to (a more general) measurable W, the integrability condition may be weakened, and W may be replaced by a distribution (rather than a density) in z so as to obtain results, e.g., on the estimation of F at a finite number of quantiles.) Rather, it is stated in a form which allows W to be any of the functions which would usually be of interest in applications, e.g., any of those functions given at the end of the first paragraph of this section, etc. (It should be noted that if the assumptions of Theorem 5 below were altered by deleting (5.5) and putting $\epsilon = 0$ in (5.4), then such weight functions as $y^2/z(1 - z)$ would be excluded. The circumlocution of including the condition (5.5) could be avoided in such cases if one could obtain a sufficiently strong bound on

$$P_U\{\sqrt{n}[S_n(x) - x] > r\sqrt{x(1 - x)}\}$$

which is independent of x. The difficulty of obtaining such an approximation is discussed in [4], p. 285.)

THEOREM 5. *Let* $W(y, z) \geqq 0$ *be defined for* $0 \leqq y < \infty, 0 < z < 1$ *and assume that* $W(y, z)$ *is monotone nondecreasing in* y *and (to avoid trivialities) that* $W(y, z)$ *is not almost everywhere zero (in the two-dimensional Lebesgue sense). Suppose further that* (a) *to every* z', $0 < z' < 1$, *not belonging to an exceptional set of linear measure zero, and every* $\delta > 0$ *there corresponds* $\epsilon(\delta, z') > 0$ *with the property that the set of* y *for which* $W(y, z)$ *is discontinuous for at least one* z *satisfying* $|z - z'| < \epsilon(\delta, z')$ *has exterior (linear Lebesgue) measure smaller than* δ. *Suppose also that* (b) *for each* ϵ *with* $0 < \epsilon < \frac{1}{2}$ *there is a function* $V(y, \epsilon)$ *such that* $W(y, z) \leqq V(y, \epsilon)$ *for* $\epsilon < z < 1 - \epsilon$ *and* $0 \leqq y < \infty$ *and such that*

$$\int_0^\infty V(y, \varepsilon) y e^{-2y^2} \, dy < \infty. \tag{5.4}$$

Suppose, finally, that (c)

$$\lim_{\epsilon \to 0} \sup_n \int_{I_\epsilon} E_U \, W(\sqrt{n}[S_n(x) - x], x) \, dx = 0. \tag{5.5}$$

Then

$$\lim_{n \to \infty} \frac{\sup_{F \varepsilon \mathfrak{F}_c} r_n(F, \phi_n^*)}{\inf_{\phi \varepsilon D_n} \sup_{F \varepsilon \mathfrak{F}_c} r_n(F, \phi)} = 1. \tag{5.6}$$

PROOF. $r_n(F, \phi_n^*)$ is, of course, independent of F for F in $\mathfrak{F}_c$. Because of Corollary 2 and the assumptions of Theorem 5, the numerator of (5.6) approaches a finite positive limit, say L, as $n \to \infty$. For any δ with $0 < \delta < L$ we may choose ϵ so small that $r_{n,\epsilon}(F, \phi_n^*)$ tends to a limit $> L - \delta$ when $n \to \infty$, where $r_{n,\epsilon}$ is the risk function corresponding to loss function $W_\epsilon(y, z)$ defined by

$$W_\epsilon(y, z) = \begin{cases} W(y, z) & \text{if } z \not\varepsilon I_\epsilon, \\ 0 & \text{if } z \varepsilon I_\epsilon. \end{cases} \tag{5.7}$$

It clearly suffices to prove (5.6) with r_n replaced by $r_{n,\epsilon}$. We hereafter drop the subscript ϵ on W_ϵ and $r_{n,\epsilon}$ and (because of (5.7)) may restate what is to be proved as (5.6) under the continuity assumption (a) on W and (replacing (5.4) and (5.5)) the assumption that $W(y, z) \leqq V(y)$ for $0 \leqq y < \infty$ and $0 < z < 1$, where

$$\int_0^\infty V(y) y e^{-2y^2} \, dy < \infty. \tag{5.8}$$

In what follows we denote (for fixed k, n, irrational x) by $P_x^*\{A\}$ the probability of any event A which is expressed in terms of $T_k^{(n)}$ when the probability function of $T_k^{(n)}$ is given by

$$
\begin{aligned}
&P\{T_{ki}^{(n)} = t_i, 1 \leqq i \leqq k+1\} \\
(5.9) \qquad &= \frac{1}{d(k, n, x)} \int_{B_k} f(x, \pi) P_\pi\{T_{ki}^{(n)} = t_i, 1 \leqq i \leqq k+1\}\, d\xi_{kn}(\pi),
\end{aligned}
$$

where P_π is given by (3.4) and $d(k, n, x)$ is the sum over all $(t_1, \cdots, t_{k+1})$ of the integral on the right side of (5.9). Expectation with respect to the probability function (5.9) will be denoted by E_x^*. We now have

$$
\begin{aligned}
\int r_n(F, \phi)\, d\xi_{kn} &= \int_0^1 \int_{B_k} E_{\pi,\phi} W(\sqrt{n}[g(x) - F(x, \pi)], F(x, \pi))\, d\xi_{kn}(\pi)\, dx \\
(5.10) \qquad &= \int_0^1 E_x^* r_{kn}(x, \phi, T_k^{(n)})\, d(k, n, x)\, dx,
\end{aligned}
$$

where the last integration (and each integration which follows) is over irrational x. Hence, in order to prove (5.6), it suffices to show that (5.8) and our continuity assumption on W imply that

$$
\begin{aligned}
&\lim_{k\to\infty} \lim_{n\to\infty} \int_0^1 E_x^* r_{kn}^*(x, T_k^{(n)})\, d(k, n, x)\, dx \\
(5.11) \qquad &\qquad = \lim_{n\to\infty} \int_0^1 E_U W(\sqrt{n}[S_n(x) - x], x)\, dx,
\end{aligned}
$$

since the right side of (5.11) is the limit of the finite positive numerator of (5.6).

Let x be an irrational number, $0 < x < 1$, which is a *non*exceptional z' of our continuity assumption (a). For fixed k with $1/(k+1) < \min(x, 1-x)$, we may write $x = (i_0 + t)/(k+1)$ with $1 \leqq i_0 \leqq k-1$ and $0 \leqq t < 1$. Write $q(r, \sigma^2) = (2\pi\sigma)^{-1/2} \exp(-r^2/2\sigma^2)$. We shall show that, given any x and k as above and any $\epsilon' > 0$, there is an integer $N = N(\epsilon', x, k)$ such that for $n > N$ we have $|d(k, n, x) - 1| < \epsilon'$ and such that, for $n > N$, P_x^* assigns probability at least $1 - \epsilon'$ to a set of $T_k^{(n)}$ values for which

$$
(5.12) \qquad r_{kn}^*(x, T_k^{(n)}) + \epsilon' > \int_{-\infty}^{\infty} W(y, x) q(y, x(1-x) + h)\, dy,
$$

where $h = (t^2 - t)/(k+1)$. But, for fixed irrational and nonexceptional x, the right side of (5.12) tends, as $k \to \infty$ (and thus, $h \to 0$), to the limit as $n \to \infty$ of the integrand in the right-hand member of (5.11). The integral of this limit is, by (5.8), the same as the right-hand member of (5.11). Thus, using (5.12) and applying Fatou's lemma to the left side of (5.11), we conclude that (5.11) will be proved if we demonstrate the statement of the sentence containing (5.12).

For fixed x and k as above and for any $\epsilon > 0$, ξ_{kn} assigns to the set of π for which $|f(x, \pi) - 1| < \epsilon$ a probability which tends to unity as $n \to \infty$. It follows that $d(k, n, x) \to 1$ as $n \to \infty$ and that (noting the relationship between ξ_{kn}^* and the f_{kn}^* of Section 3), for any $\epsilon > 0$ and for n sufficiently large, P_x^* assigns probability at least $1 - \epsilon$ to a set of values $t_k^{(n)}$ of $T_k^{(n)}$ for which, writing $y_i = t_{ki}^{(n)}/n$, the joint density function of the $\bar{\gamma}_i = \sqrt{n}(p_i - y_i)$ $(1 \leqq i \leqq k)$ according to $\xi_{kn}^*(\cdot, x, t_k^{(n)})$ in a spherical region centered at 0 in the space of the $\bar{\gamma}_i$ and of probability at least $1 - \epsilon$ according to ξ_{kn}^* is at least

$$(1 - \epsilon)(2\pi)^{-k/2}\left(\prod_1^{k+1} y_i\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_1^{k+1} \gamma_i^2/y_i\right). \tag{5.13}$$

Now, in the notation of Section 3, for $1 \leqq i \leqq k$,

$$F(i/(k+1), \pi) = p_1 + \cdots + p_i .$$

Hence, if $T_{ki}^{(n)} = ny_i$ $(1 \leqq i \leqq k)$, we have (because of the form of (3.3))

$$F(x, \pi) = p_1 + \cdots + p_{i_0} + tp_{i_0+1} = (y_1 + \cdots + y_{i_0} + ty_{i_0+1}) + (\bar{\gamma}_1 + \cdots + \bar{\gamma}_{i_0} + t\bar{\gamma}_{i_0+1})/\sqrt{n}.$$

Now, $T_{ki}^{(n)}/n$ tends in probability (according to P_x^*) to $1/(k+1)$, and expression (5.13) with $\epsilon = 0$ is continuous in the y_i (in the region where all $y_i > 0$). Moreever, if we had $\epsilon = 0$ in (5.13) and assumed the validity of this expression for all values of the γ_i and put all $y_i = 1/(k+1)$, then $(\bar{\gamma}_1 + \cdots + \bar{\gamma}_{i_0})$ and $\bar{\gamma}_{i_0+1}$ would, according to (5.13), have a bivariate normal density function with means zero and covariance matrix

$$\frac{1}{(k+1)^2}\begin{pmatrix} i_0(k+1-i_0) & -i_0 \\ -i_0 & k \end{pmatrix}. \tag{5.14}$$

The corresponding density function of $\bar{\gamma}_i + \cdots + \bar{\gamma}_{i_0} + t\bar{\gamma}_{i_0+1}$ would then be normal with mean zero and variance

$$[i_0(k+1-i_0) - 2ti_0 + t^2k]/(k+1)^2 \doteq x(1-x) + h. \tag{5.15}$$

Hence, if we carry through this last argument with the actual form of (5.13) and its region of validity, we conclude that, for any $\epsilon'' > 0$ and for n sufficiently large, P_x^* assigns probability at least $1 - \epsilon''$ to a set of values $t_k^{(n)}$ of $T_k^{(n)}$ for which, on a real interval centered at 0 and of probability at least $1 - \epsilon''$ according to $\xi_{kn}^*(\cdot, x, t_k^{(n)})$, this last measure induces a distribution function J for

$$\sqrt{n}[F(x, \pi) - (y_1 + \cdots + y_{i_0} + ty_{i_0+1})] = \Lambda_x \text{ (say)}$$

whose absolutely continuous component has a corresponding density (the derivative of J) whose magnitude is at least

$$(1 - \epsilon'')q(\lambda, x(1-x) + h) \tag{5.16}$$

almost everywhere on this interval of λ-values.

Next, we note that

$$W(\sqrt{n}[g(x) - F(x, \pi)], F(x, \pi)) = W(\rho - \Lambda_x, x + \mu + \Lambda_X/\sqrt{n}), \tag{5.17}$$

where $\rho = \sqrt{n}[g(x) - (y_1 + \cdots + y_{i_0} + ty_{i_0+1})]$ and

$$\mu = -x + (y_1 + \cdots + y_{i_0} + ty_{i_0+1}).$$

For fixed x and k as above, denote by α the right side of (5.12). Let β be such that the right side of (5.12) is at least $\alpha - \epsilon'/4$ if the limits of integration are changed to $(-\beta, \beta)$. Let $c = W(\beta, x)$. Let the δ of our assumption (a) be

$$\epsilon'/8cq(0, x(1 - x) + h),$$

and let $z' = x$ where x is nonexceptional. The set $0 \leqq y \leqq \beta$, $|z - x| \leqq \epsilon(\delta, x)$ minus a suitable countable set of open intervals of total length $<\delta$ covering the points of discontinuity is closed and bounded. Hence, W is uniformly continuous on this set. Hence, there is a value $\epsilon_1 > 0$ such that $W(y, z) \geqq W(y, x) - \epsilon'/4$ for $|x - z| \leqq \epsilon_1$ and $0 \leqq y \leqq \beta$ but y not in the excluded set. If $0 \leqq y \leqq \beta$ and y is in the exceptional set, y is in a maximal subinterval of the exceptional set of either the form $a < y < b$ with $a > 0$ or else of the form $0 \leqq y < b$. Define $\tilde{W}(y, x) = W(a, x)$ in the former case and $\tilde{W}(y, x) = 0$ in the latter. If $0 \leqq y \leqq \beta$ but y is not exceptional, define $\tilde{W}(y, x) = W(y, x)$. If $y > \beta$, define $\tilde{W}(y, x) = \tilde{W}(\beta, x)$. Finally, set $\tilde{W}(-y, x) = \tilde{W}(y, x)$. The function $\tilde{W}$ so defined is symmetric in y, nondecreasing in y for $y \geqq 0$, and has the property that

$$W(y, z) \geqq \tilde{W}(y, x) - \epsilon'/4 \qquad \text{for } |x - z| \leqq \epsilon_1 \text{ and all } y, \tag{5.18}$$

and also that

$$\int_{-\beta}^{\beta} \tilde{W}(y, x)q(y, x(1 - x) + h)\, dy \geqq \alpha - \epsilon'/2. \tag{5.19}$$

Now, let $N = N(\epsilon', x, k)$ be such that, for $n > N$ and with $\epsilon'' = \epsilon'/4(\alpha + 1)$, the conclusion (5.16) holds with the λ-interval including the interval $(-\beta, \beta)$, and such that $|d(k, n, x) - 1| < \epsilon'$ for $n > N$. Write $\bar{\mu}$ for the random variable defined by putting $T_{ki}^{(n)}/n$ for y_i in the definition of μ. Since $\bar{\mu}$ tends to zero in probability (according to P_x^*) as $n \to \infty$, we may also suppose N to be such that, for $n > N$, $P_x^*\{|\bar{\mu}| + \beta/\sqrt{n} < \epsilon_1\} \geqq 1 - \epsilon''$. Next, we recall the statement made immediately following the statement of Lemma 1, that for $n = 1$ the conclusion of Lemma 1 holds if the normal probability density is replaced by one truncated at $(-\beta, \beta)$. We also note that the integral (with respect to λ) of this truncated density multiplied by $\tilde{W}(\rho - \lambda, x)$ is easily seen (by an argument like that used to deduce (4.11) from (4.9)) to be minimized at $\rho = 0$. We note, as in previous sections, that if (5.12) is true under the restriction to nonrandomized ϕ (in the definition of r^*), then (5.12) is a fortiori true without this restriction. Thus, from (5.2), (5.16), (5.17), (5.18), and (5.19), we have for $n > N$ that, with P_x^*-probability at least

$$1 - 2\epsilon'' > 1 - \epsilon', \tag{5.20}$$

$T_k^{(n)}$ will be such that

$$\begin{aligned} r_{kn}^*(x, T_k^{(n)}) & \\ &\geqq \inf_\rho \int_{-\beta}^{\beta} W(\rho - \lambda, x + \mu + \lambda/\sqrt{n})(1 - \epsilon'')q(\lambda, x(1 - x) + h)\, d\lambda \\ &\geqq (1 - \epsilon'') \inf_\rho \int_{-\beta}^{\beta} [\tilde{W}(\rho - \lambda, x) - \epsilon'/4]q(\lambda, x(1 - x) + h)\, d\lambda \\ &= (1 - \epsilon'') \int_{-\beta}^{\beta} [\tilde{W}(-\lambda, x) - \epsilon'/4]q(\lambda, x(1 - x) + h)\, d\lambda \\ &\geqq (1 - \epsilon'')(\alpha - 3\epsilon'/4) > \alpha - \epsilon'. \end{aligned} \tag{5.21}$$

This completes the proof of (5.12) and thus of Theorem 5.

We have not stated a corollary to Theorem 5 of the type given after Theorems 3 and 4. For $F \varepsilon \mathfrak{F} - \mathfrak{F}_c$, a weight function of the form (5.1) seems less meaningful because the loss contributed at a saltus x of F is measured by $W(y, z)$, where $z = F(x + 0)$. There are also certain technical difficulties in that the numerator of (5.6) need no longer be the same if $\mathfrak{F}_c$ is replaced by $\mathfrak{F}$. We shall not bother with the circumlocutions (e.g., additional restrictions on W) necessary to obtain a corollary from Theorem 5 in the same trivial manner as such corollaries were obtained from Theorems 3 and 4.

Theorem 5 implies certain much weaker results which, for special forms of W, may also be obtained from results obtained by Aggarwal [6]. He considers only the class C_n of procedures which with probability one set $g(x) = c_i^{(n)}$ for $Z_i^{(n)} \leqq x < Z_{i+1}^{(n)}$, where the $\{Z_i^{(n)}\}$ are the ordered values of the $\{X_i^{(n)}\}$. (Such procedures have constant risk for $F \varepsilon \mathfrak{F}_c$ and W_n^* of the form (5.1).) For the special functions $W(y, z) = |y|^r$ and $W(y, z) = |y|^r/z(1 - z)$ (r a positive integer), he obtains the best $c_i^{(n)}$ explicitly in a few cases and in the other cases characterizes them as the solutions of certain equations. In the former cases ϕ_n^* may be seen to be asymptotically best in C_n. This result is an immediate consequence of Theorem 5, where the result is proved for the class D_n of all procedures, of which the class C_n is a small subclass.

6. Other loss functions; multinomial estimation problems. The results obtained in the previous three sections may be extended to a more general class of loss functions to which the same methods of proof may be seen to apply. Thus, for example, in Sections 3 and 4 we could consider the maximum deviation over a set of x values for which $F(x)$ is in a specified *sub*set of the unit interval (this will involve techniques like those used in Section 5); the formulation of Theorem 5 already includes weight functions which may (e.g.) vanish for certain values of $F(x)$, and other modifications (e.g., to consider a finite set of points) are mentioned in the paragraph preceding Theorem 5. We may also consider (in Section 4) loss functions such as $W_1(r_1) + W_2(r_2)$ where r_1 and r_2 are the maximum deviations over two (not necessarily disjoint) sets of the type mentioned above and W_1 and W_2 are functions of the type considered in Section 4. Linear combinations of loss functions of this last type and the type considered in Section 5 may similarly be treated. In all of the above we may replace $\sup_x |g(x) - F(x)|$ by $\sup_x [|g(x) - F(x)|h(F(x))]$, where h is any nonnegative function (suitably regular), without any difficulty; this includes as a special case maximization over a subset as described above.

Thus, it appears that our results hold for a very general class of weight functions. It would of course be of interest to subsume all cases under one unified criterion and one method of proof. In the portion of Section 7 which is devoted to heuristic remarks, such a criterion (symmetry and convexity of a certain functional) is indicated; unfortunately, it does not include all cases treated above (e.g., the result of Section 3, which is apparently somewhat deeper), some of which will be seen in Section 7 to be slightly more difficult to handle than the symmetric convex functionals. A more general class Ω of monotone functionals for which (perhaps under slight regularity conditions) our results would seem likely to hold, and which includes the weight functions of Sections 3, 4, and 5 as well as those of the previous paragraph, is also indicated in Section 7. In the present context, this class consists of nonnegative functionals W of the function $|\delta|$ defined by $\delta(y) = g(F^{-1}(y)) - y$, $0 \leqq y \leqq 1$ (where we suppose for simplicity that the possible c.d.f.'s F under consideration are members of $\mathfrak{F}_c$ which are for each F strictly increasing for $\sup F^{-1}(0) \leqq y \leqq \inf F^{-1}(1)$) for which $W(|\delta_1(y)|) \leqq W(|\delta_2(y)|)$ whenever $|\delta_1(y)| \leqq |\delta_2(y)|$ for $0 \leqq y \leqq 1$. However, at this writing it is not evident how to give a rigorous *unified* proof (as distinguished from the heuristic one of Section 7) even for the class of weight functions which are convex symmetric functionals (of δ, in the present context), let alone to give one for the class Ω.

Another modification is to consider $\sup_x |g(x) - F(x)|h(x)$ above instead of $\sup_x |g(x) - F(x)|h(F(x))$. In this case ϕ_n^* will not have constant risk over $\mathfrak{F}_c$. However, this case is easily treated as follows: suppose for simplicity that h is continuous and bounded (the unbounded case is trivial and may be treated by a similar argument). Let J be an interval in which h is entirely within a prescribed $\epsilon > 0$ of $\sup_x h(x)$. We may for simplicity suppose J to be the unit interval. Then the risk function of ϕ_n^* will attain a value close to its maximum for $F = U$. The argument of Sections 3 and 4 may now be applied. In a similar manner we may consider in Section 5 loss functions for which the risk function of ϕ_n^* is not a constant; for example, (5.1) could be replaced by

$$W_n^*(F, g) = \int_{-\infty}^{\infty} W(\sqrt{n}[g(x) - F(x)], x)\, d\mu(x) \tag{6.1}$$

for a specified function W and measure μ satisfying certain regularity conditions.

An interesting question is whether or not our results can be extended to yield a *sequential* asymptotic minimax character, e.g., in the sense of Wald [7]. This is too large a topic to be discussed thoroughly in this paragraph, but a few indicative comments are in order. An essential idea present in the form of the ξ_{kn} of Sections 3, 4, and 5 is that, when k is large, a certain multinomial estimation problem is *almost as difficult* as the problem of estimating F. This suggests that, when the weight function considered here is such that the corresponding multinomial problem has (perhaps only asymptotically) a fixed sample-size minimax estimator (among all sequential estimators), then we may conclude that the fixed sample-size procedure ϕ_n^* is asymptotically minimax among all *sequential* procedures. An examination of [7] shows that such an asymptotic sequential minimax property for the multinomial problem will often be easy to prove using methods like Wald's.

Finally, the methods of this paper (without the limit considerations as $k \to \infty$) may be used to prove certain asymptotic minimax results for the estimation of the parameter π of the multinomial distribution (3.4) as $n \to \infty$, for any fixed k. To see this, we note that, under fairly general conditions of monotonicity and symmetry of the weight function (similar to those of Sections 3 to 5), the limiting risk function of $T_k^{(n)}/n$ as $n \to \infty$ will be continuous in a neighborhood of the point of B_k at which its maximum is achieved. Hence, for any $\epsilon > 0$, there will exist an interior point V_k^* of B_k in a neighborhood of which the limiting risk function of $T_k^{(n)}/n$ is continuous and at which point the limiting risk of $T_k^{(n)}/n$ is within ϵ of its maximum. One can then find a sequence $\{\xi_{kn}\}$ of a priori distributions on B_k (similar to the sequence used in Sections 3, 4, and 5) which assigns to any neighborhood of V_k^* a probability approaching one as $n \to \infty$ and which "shrinks down" on V_k^* at a slow enough rate (see the remarks of the paragraph preceding that containing (3.6)) to make the a posteriori probability distribution of the $\sqrt{n}(p_j - T_{kj}^{(n)}/n)$ normal with mean 0 so that $T_k^{(n)}/n$ is asymptotically Bayes with respect to $\{\xi_{kn}\}$, with integrated risk approaching the limiting risk of $T_k^{(n)}/n$ at V_k^*. The asymptotic minimax character of $T_k^{(n)}/n$ follows. We need not detail the wide variety of weight functions for which this optimum asymptotic property of the classical multinomial estimator $T_k^{(n)}/n$ follows from the methods and the results of the three previous sections as well as of the present section. It is perhaps worth while to remark that, although the results in Sections 3, 4, and 5 are stated in terms of deviations of sums $\psi_{n1} + \psi_{n2} + \cdots + \psi_{nj}$ of components ψ_{ni} of the estimator ψ_n from $p_1 + p_2 + \cdots + p_j$ $(1 \leqq j \leqq k + 1)$, the given proofs apply with only trivial modifications to weight functions depending on differences $\psi_{nj} - p_j$. Thus, for example, for any set of numbers $c_j > 0$, the asymptotic minimax character of $T_k^{(n)}/n$ for estimating $\pi \in B_k$ for the risk function

$$r_n(\pi, \psi_n) = 1 - P_\pi\{|\psi_{nj} - p_j| \leqq c_j/\sqrt{n}, (1 \leqq j \leqq k + 1)\} \tag{6.2}$$

follows from the asymptotic normality of the a posteriori distribution, noted above, and from the convexity and symmetry about ψ_n of the set of π (in R^{k+1}, not B_k) satisfying the inequalities in brackets in (6.2). (It is clear from this example that V_k^* need not be the V_k of Section 3.) The result for other risk functions follows similarly, using the methods and results of Sections 4, 5, and 6.

These asymptotic results for the multinomial estimation problem do not seem to have appeared previously in the literature. As indicated two paragraphs above, some of these multinomial results may also be extended to sequential problems.

7. Convex functionals and monotone functionals of stochastic processes; heuristic considerations. The first part of this section will be devoted to some simple remarks concerning convex symmetric functionals of random elements; these remarks will then be applied to give a short heuristic argument for many of the results obtained in previous sections.

Let $B = \{b\}$ be a linear space (or system) and ζ a random element with range in B and having a symmetric distribution, i.e., such that whenever A is a measurable subset of B so is $-A$ and $P\{\zeta \varepsilon A\} = P\{\zeta \varepsilon -A\}$. Let ω be a measurable real-valued convex functional on B which is symmetric ($\omega(b) = \omega(-b)$ for $b \varepsilon B$) and convex ($\omega(\lambda b_1 + (1 - \lambda)b_2) \leqq \lambda\omega(b_1) + (1 - \lambda)\omega(b_2)$ for $0 < \lambda < 1$ and $b_1, b_2 \varepsilon B$). We now note that, since $\min_b \omega(b) = \omega(0) > -\infty$ so that the expected value $E\omega(\zeta)$ is always defined, we may conclude that

$$E\omega(\zeta) = \min_{b \varepsilon B} E\omega(\zeta + b) \tag{7.1}$$

from the equation (implied by symmetry of P)

$$E\omega(\zeta + b) = E\omega(-\zeta + b) = \tfrac{1}{2}E\{\omega(\zeta + b) + \omega(-\zeta + b)\} \tag{7.2}$$

and the equation (implied by symmetry and convexity of ω)

$$\omega(\zeta + b) + \omega(-\zeta + b) = \omega(\zeta + b) + \omega(\zeta - b) \geqq 2\omega(\zeta). \tag{7.3}$$

We shall now apply (7.1) to the "tied-down" Wiener process (see [2]). B is now the space of continuous functions $b(t)$ on the unit interval $0 \leqq t \leqq 1$. The probability measure P assigns probability one to the subset B_0 of elements b of B satisfying $b(0) = b(1) = 0$. The measurable sets are generated by all sets of the form $\{b | b(t_0) < a_0\}$ for $0 \leqq t_0 \leqq 1$ and a_0 real. The joint distribution of $\zeta(t_1), \cdots, \zeta(t_n)$ for any $0 \leqq t_1 \leqq \cdots \leqq t_n \leqq 1$ is normal with $E\zeta(s) = 0$ and $E\{\zeta(s)\zeta(t)\} = \min(s, t) - st$ for $0 \leqq s, t \leqq 1$. Note that the distribution of ζ is symmetric.

Let W be any symmetric real-valued convex function on R^1. Then

$$W(\max_t |b(t)|)$$

is a convex functional of b and (7.1) implies that

$$EW(\max_t |\zeta(t)|) \leqq EW(\max_t |\zeta(t) + \rho(t)|) \tag{7.4}$$

for all continuous functions ρ. Generalizations of this result to the case where $\max_t$ is replaced by $\max_t h(t)$ in the manner of Section 6, or where ρ is allowed to be of a more general class than the continuous functions, are easily achieved by adjoining additional functions to B. One may also note that, for every $r > 0$,

$$P\{\max_t |\zeta(t)| > r\} \leqq P\{\max_t |\zeta(t) + \rho(t)| > r\} \tag{7.5}$$

for all continuous (or more general, as noted above) functions ρ. However, this cannot be proved in the same manner as (7.4), since the characteristic function of the subset of B for which $\max_t |b(t)| > r$ is not a convex functional on B. The validity of (7.5) follows, however, from (2.4) and Lemma 1. This strong result of a domination of an entire distribution function in the sense of (7.5) is deeper than the result (7.4); for (7.5) requires (in the proof of Lemma 1 in [8]) not merely the symmetry of the probability distribution, but also the convexity for every $u > 0$ of the set where the joint density of $\zeta(t_1), \cdots, \zeta(t_k)$ (for any k and $t_1, \cdots, t_k$) is $\geqq u$. (Note, for example, that it is not necessarily true for a symmetrically distributed real-valued random variable X that $P\{|X| > r\} \leqq P\{|X + \rho| > r\}$ for all real ρ.) Similarly, the result (7.4) for real functions $W(z)$ on the nonnegative reals which are nondecreasing in z for $z \geqq 0$ (but not necessarily convex) is a consequence of (7.5) but cannot be proved directly in the manner of (7.4) for convex W. Thus, to summarize, (7.4) for convex W follows from the symmetry of the probability measure, while in proving (7.4) for nondecreasing W (and, in particular, (7.5)) we use the additional assumption on the probability measure which is used to prove Lemma 1. We note that it has not been necessary to assume any integrability condition here.

It is interesting to note that, for the special case of a linear function $\rho(t) = c + dt$, the right side of (7.5) is given by formula (4.3) of [2] with $a = r - c - d$, $b = r - c$, $\alpha = r + c + d$, $\beta = r + c$ (unless $a \geqq 0$, $b > 0$, $\alpha \geqq 0$, $\beta > 0$, the probability in question is unity; our $\zeta(t)$ is Doob's $X(t)$). It does not seem completely apparent from the form of (4.3) of [2] that this expression, with the above substitutions, is a minimum for $c = d = 0$. The same is true of expectations with respect to the d.f. (4.3) of [2] of functions W of the type considered above.

Next, let the real-valued function $W(y, z)$ be symmetric and convex in y for each z ($-\infty < y < \infty$, $0 \leqq z \leqq 1$) and satisfy obvious measurability conditions. Let μ be any measure on the unit interval. Then

$$\omega(b) = \int_0^1 W(b(t), t)\, d\mu(t) \tag{7.6}$$

is a convex functional on B and hence (7.1) holds. In this case the result for the case where $W(y, z)$ is symmetric in y and nondecreasing in y for $y \geqq 0$ (but not necessarily convex for each z) is not much more difficult, although it cannot be handled by using (7.1): we need only apply Lemma 1 for $n = 1$ for each fixed z in this case in order to obtain the desired result.

More general convex functionals (such as combinations of the two varieties as treated in Section 6) or nonconvex functionals with certain monotonicity properties may be handled, similarly, by using (7.1) or consequences of Lemma 1 similar to (7.5), respectively. It is possible that the conclusion (7.1) holds for the class Ω of (not necessarily convex) functionals ω which are nonnegative, for which $\omega(\zeta) = \omega(|\zeta|)$, and for which $\omega(|\zeta_1|) \leqq \omega(|\zeta_2|)$ whenever $|\zeta_1(t)| \leqq |\zeta_2(t)|$ for all t. Similarly, results on processes other than the tied-down Wiener process, whose distributions are symmetrical or also satisfy the property which (as mentioned above) is used in [8] in proving the more general form of Lemma 1, may be obtained by using (7.1) or the generalization of Lemma 1 in [8], respectively;

We now turn to a heuristic argument for the results obtained in previous sections (except for certain results of Section 6, as noted below). This discussion may also be thought of as an outline of one intuitive explanation of why these results hold, the epsilontics and use of Bayes solutions in the previous sections supplying the needed rigor. However, the discussion which follows does not use Bayes solutions, and it would certainly be worth while to obtain an independent argument which would show that in the limit one need only consider "limiting" decision procedures of the type considered below, and thus to conclude that the argument which follows can be made rigorous by means of only brief additions.

In the previous sections we were concerned with estimating (for various weight functions) an unknown element F of $\mathfrak{F}_c$. Denote by $g_n(x; X^{(n)})$ such an estimator of $F(x)$ based on $X^{(n)}$ (for notational simplicity, we have considered a nonrandomized $\phi_n \varepsilon D_n$). Suppose we could show that for our considerations, at least asymptotically, it is only necessary to consider functions g_n of the form

$$g_n(x; X^{(n)}) = \psi_n(S_n(x)),$$

i.e., procedures in the class C_n mentioned in the last paragraph of Section 5 (this is one of two crucial gaps in our heuristic argument, for it is not obvious how to give a short proof, which in no way depends on the results of Sections 3 to 5, of this supposition). Procedures in C_n will have constant risk for $F \varepsilon \mathfrak{F}_c$ and for any of the weight functions of Sections 3, 4, and 5 (and those of Section 6 for which ϕ_n^* was not remarked to have constant risk). Thus, we may consider the distribution of the random function $\psi_n(S_n(t)) - t$ for $0 \leqq t \leqq 1$ where $F = U$. Write $\psi_n(z) = z + \rho_n(z)/\sqrt{n}$. Then

$$\sqrt{n}[\psi_n(S_n(t)) - t] = \sqrt{n}[S_n(t) - t] + \rho_n(S_n(t)).$$

If we could now suppose (and this is the other crucial nonrigorous development) that there is a sequence $\{\psi_n\}$ of minimax procedures (for $n = 1, 2, \cdots$) such that the corresponding sequence $\{\rho_n(z)\}$ has a continuous limit $\rho(z)$ uniformly in z as $n \to \infty$, and note that $\rho_n(S_n(t))$ would then be bounded (for n sufficiently large) and would tend to $\rho(t)$ with probability one as $n \to \infty$, then by [2] and [3] our consideration of $\sqrt{n}[\psi_n(S_n(t)) - t]$ would be reduced, asymptotically, to that of $\zeta(t) + \rho(t)$. The earlier comments of this section would then yield the desired asymptotic minimax properties of ϕ_n^*.

REFERENCES

[1] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Inst. Ital. Atti. Giorn.*, Vol. 4 (1933), pp. 83–91.

[2] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[3] M. Donsker, "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[4] PAUL LÉVY, *Théorie de L'Addition des Variables Aléatoires*, Gauthier-Villars, Paris (1937).
[5] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., New York (1937).
[6] O. P. AGGARWAL, "Some minimax invariant procedures for estimating a c.d.f.," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 450–463.
[7] A. WALD, "Asymptotic minimax solutions of sequential point estimation problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press (1951), pp. 1–11.
[8] T. W. ANDERSON, "The integral of a symmetric unimodal function," *Proc. Amer. Math. Soc.*, Vol. 6 (1955), pp. 170–176.
[9] N. V. SMIRNOFF, "Approach of empiric distribution functions," *Uspyekhi Matem. Nauk.*, Vol. 10 (1944), pp. 179–206.