ASYMPTOTIC MINIMAX CHARACTER OF THE SAMPLE DISTRIBUTION FUNCTION FOR VECTOR CHANCE VARIABLES

By J. Kiefer[1] and J. Wolfowitz[2]

Cornell University

Summary. The purpose of this paper is to prove Theorem 1 stated in Section 1 below and Theorem 2 of Section 6 and the results of Section 7. These theorems are the generalizations to vector chance variables of Theorems 4 and 5 and Section 6 of [1], and state that the sample distribution function (d.f.) is asymptotically minimax for the large class of weight functions of the type described below. The main difficulties are embodied in the proof of Theorem 1 (Sections 2 to 5), where the loss function is a function of the maximum difference between estimated and true d.f. The proof utilizes the results of [2] and is not a straightforward extension of the result of [1], because the sample d.f. is no longer "distribution free" (even in the limit), and hence it is necessary to prove the uniformity of approach, to its limit, of the d.f. of the normalized maximum deviation between sample and population d.f.'s (for a certain class of d.f.'s). The latter fact enables us essentially to infer the existence of a uniformly (with the sample number) approximately least favorable (to the statistician) d.f., by means of which the proof of the theorem is achieved. Theorem 2 (Section 6) considers loss functions of integral type, and more general loss functions are treated in Section 7.

1. Introduction and preliminaries. The problem of finding a reasonable estimator of an unknown distribution function (d.f.) F in one or more dimensions is an old one. In the one-dimensional case the first extensive optimality results were obtained in [1]. It was shown there that, although a minimax procedure for sample size n may depend on the weight function as well as on n, the sample d.f. ϕ_n^* is asymptotically minimax as $n \to \infty$ for a very large class of weight functions which includes almost any weight function of practical interest. Also, an exact minimax procedure is extremely tedious to calculate in most practical cases, and is less convenient to use in practice than is ϕ_n^*. Moreover, one can obtain from [1] a bound on the relative difference between the maximum losses which can be encountered from using ϕ_n^* or the actual minimax procedure, and for many common weight functions this bound indicates that ϕ_n^* is very close to being minimax for fairly small values of n.

For dimension $m > 1$ the minimax problem presents difficulties which are not present when $m = 1$. (An outline of the main ideas and difficulties encountered in the proofs when $m = 1$ or when $m > 1$ will be given in Section 4; the

Received April 21, 1958.

[1] The research of this author was under contract with the Office of Naval Research.

[2] The research of this author was supported by the U. S. Air Force under Contract No. AF 18(600)-685, monitored by the Office of Scientific Research.

proof there is completed in Section 5; additional considerations for various weight functions are outlined in Sections 6 and 7.) These difficulties stem from the fact that neither ϕ_n^* nor any other known procedure which seems a reasonable candidate for optimality, has the distribution-free property possessed by ϕ_n^* when $m = 1$. This fact has led investigators of the problem when $m > 1$ to try (unsuccessfully) to find reasonable distribution-free procedures. Such investigations now seem to have been aimed in the wrong direction; for the main result of the present paper is that *ϕ_n^* is still asymptotically minimax for a large class of weight functions, even though it is no longer distribution free.*

The proof of the result just stated presents new difficulties far greater than those encountered when $m = 1$. In order to describe these difficulties briefly, let us suppose for the moment that the risk function is the expected value (under the true F) of $n^{1/2}$ times the maximum absolute deviation between estimated and true d.f. The computation of this risk function or its limit as $n \to \infty$ for the sequence of procedures ϕ_n^* (or any other reasonable sequence procedures) is known to present formidable difficulties, even for very simple continuous F (e.g., the uniform distribution on the unit square when $m = 2$). Our method of proof circumvents such a computation by showing that, when n is suitably large, the risk function of ϕ_n^* is changed arbitrarily little from what it would be if the maximum deviation were taken over a large but *finite* set of points instead of over all of m-space (this uses a result of [2]). Thus, the problem is reduced to a multinomial problem, similar to the reduction of [1] when $m = 1$, and we can circumvent the explicit computation of the risk there in a manner like that used in the multinomial case in [1], and which will be described in Section 3 below. But there remains another difficulty: in order to use a Bayes technique like that of [1] to prove the asymptotic minimax character of ϕ_n^*, we must show that there is a d.f. F_δ at which the risk function of ϕ_n^* is almost a maximum for all sufficiently large n; i.e., that the location of some approximate maximum does not "wander around" too much with n. Because of the distribution-free nature of the chance loss (for many common loss functions) under ϕ_n^* when $m = 1$, the existence of such an F_δ was automatic there (any continuous d.f. could be used); for $m > 1$, our proof requires the result of Lemma 1 of Section 2 below to obtain the existence of such an F_δ, at least when F is restricted to belong to a class of d.f.'s which in Section 5 is seen to be dense enough in an appropriate sense to yield the desired result. Once such an F_δ is known to exist, a sequence of approximately least favorable a priori distributions can be constructed for the approximating multinomial problem in the manner of [1]; this will be described in Section 4.

Aside from the difficulties described in the previous paragraph, the proofs of minimax results when $m > 1$ are very similar to those when $m = 1$. Therefore, rather than to repeat all of the details of [1], in each of Sections 4, 6, and 7 we will first describe the idea of the proof and then will indicate the modifications needed in the proof of the corresponding section of [1] to make it apply when $m > 1$.

We now give the notation used in this paper. m will denote any positive integer, fixed throughout the sequel. $\mathfrak{F}$ denotes the class of all d.f.'s on Euclidean m-space R^m, and $\mathfrak{F}^c$ denotes the subclass of continuous members of $\mathfrak{F}$. Let D be any subclass of the space of real functions on R^m. For simplicity we assume $\mathfrak{F} \subset D$, although it is really only necessary that D contains every possible function of the form S_n (defined below), for all n and $z^{(n)}$. Let B be the smallest Borel field on D such that every element of $\mathfrak{F}$ is an element of B and such that, for every positive integer k, real numbers $a_1, \cdots, a_k$, and m-vectors $t_1, \cdots, t_k$, the set $\{g \mid g \in D; g(t_1) < a_1, \cdots, g(t_k) < a_k\}$ is in B. (For example, we might have $D = \mathfrak{F}$ and B the Borel sets of the usual metric topology.) Let $\mathfrak{D}_n$ be the class of all real functions ϕ_n on $B \times R^{mn}$ such that $\phi_n(\cdot\,; z)$ is a probability measure (B) on D for each z in R^{mn} and such that $\phi_n(\Delta\,;\cdot)$ is a Borel-measurable function on R^{mn} for each Δ in B.

We now describe the statistical problem. Let $Z_1, \cdots, Z_n$ be independently and identically distributed m-vectors, each distributed according to some d.f. F about which it is known only that $F \in \mathfrak{F}$ (or $\mathfrak{F}^c$ or some other suitably dense subclass of $\mathfrak{F}$). The statistician wants to estimate F. Write $Z^{(n)} = (Z_1, \cdots, Z_n)$ and $z^{(n)} = (z_1, \cdots, z_n)$, where $z_i \in R^m$. Having observed $Z^{(n)} = z^{(n)}$, the statistician uses some decision function ϕ_n (a member of $\mathfrak{D}_n$) as follows: a function $g \in D$ is selected by means of a randomization according to the probability measure $\phi_n(\cdot\,; z^{(n)})$ on D; the function g so selected (which need not even 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 it may be; the precise meaning of "close" will be reflected by a weight function $W_n(F, g)$ which measures the loss when F is the true distribution function and g is the estimate of it. The probability of making a decision in Δ when ϕ_n is used and F is the true d.f. is

$$\mu_{F,\phi_n}(\Delta) = \int \phi_n(\Delta, z^{(n)}) F\,(dz^{(n)}), \tag{1.1}$$

which, as a function of Δ, will be a probability measure on D (see the next paragraph). Denoting expectation of a function on D with respect to this measure by E_{F,ϕ_n} (the symbol P_{F,ϕ_n} is used analogously, and the subscript ϕ_n will be omitted when it is not relevant), the risk function of the procedure ϕ_n is defined by

$$r_n(F, \phi_n) = E_{F,\phi_n} W_n(F, g); \tag{1.2}$$

i.e., it is the expected loss when F is true and ϕ_n is used. A sequence $\{\phi_n'\}$ of procedures is said to be *asymptotically minimax* relative to a sequence W_n of weight functions and a subclass $\mathfrak{F}'$ of $\mathfrak{F}$ if

$$\lim_{n\to\infty} \frac{\sup_{F \in \mathfrak{F}'} r_n(F, \phi_n')}{\inf_{\phi_n \in \mathfrak{D}_n} \sup_{F \in \mathfrak{F}'} r_n(F, \phi_n)} = 1. \tag{1.3}$$

(We note that this is a stronger property than that obtained by suppressing the supreme operation in the numerator and asking that the upper limit as

$n \to \infty$ be $\leqq 1$ for each F; this latter asymptotic property is much easier to verify than (1.3).) A nonrandomized decision function is one which for each $z^{(n)}$ assigns probability one to a single element (depending on $z^{(n)}$) of D. By ϕ_n^* we denote the nonrandomized procedure which chooses as decision the "sample d.f." S_n defined by

$$S_n(z) = n^{-1} \text{ (number of } Z_i \leqq z, 1 \leqq i \leqq n),$$

where as usual $Z \leqq z$ means that each component of Z is $\leqq$ the corresponding component of z. We shall not explicitly display the dependence of the chance function S_n on $Z^{(n)}$.

Obvious measurability considerations arise in connection with (1.1), (1.2), etc. These are handled exactly as in Section 1 of [1].

We can now state the main result of this paper, whose proof will occupy the next four sections (modifications and extensions are considered in Sections 6 and 7).

THEOREM 1. *Suppose* $W_n(F, g) = W(n^{1/2} \sup_z | g(z) - F(z)|)$, *where, for* $r \geqq 0$, $W(r)$ *is continuous, nonnegative, monotonically nondecreasing, not identically zero, and satisfies*

$$\int_0^\infty rW(r)e^{-c'_m r^2} \, dr < \infty \tag{1.4}$$

where c'_m *is given by* (1.8). *Then* $\{\phi_n^*\}$ *is asymptotically minimax relative to* $\{W_n\}$ *and* $\mathfrak{F}$.

Before listing the results of [2] which will be used in the present paper, we introduce some additional notation. When Z_1 has d.f. F, define

$$D_n = \sup_{x \varepsilon R^m} |S_n(x) - F(x)|$$

and

$$G_n(r; F) = P_F\{D_n < r/n^{1/2}\}.$$

For k a positive integer, write A_k^m for the subset of $(k + 1)^m$ points in the m-dimensional unit cube $I^m = \{x \mid 0 \leqq x \leqq 1, x \varepsilon R^m\}$ for which each coordinate is an integral multiple of $1/k$. Write

$$D_{n,k} = \sup_{x \varepsilon A_k^m} |S_n(x) - F(x)|$$

and

$$G_{n,k}(r; F) = P_F'\left\{D_{n,k} < \frac{r}{n^{1/2}}\right\}$$

We also write

$$G_{\infty,k}(r; F) = \lim_{n\to\infty} G_{n,k}(r; F);$$

the existence of this limit follows from the multivariate central limit theorem. Finally, let $\mathfrak{F}^*$ be the class of d.f.'s F which are in $\mathfrak{F}^c$ and for which each one-dimensional marginal d.f. of F is uniform on I^1. Clearly, if Z_1 has d.f. F in $\mathfrak{F}^c$, we can perform continuous transformations on the components of Z_1, so as to make the result have a d.f. F^* in $\mathfrak{F}^*$, without changing G_n. This fact will be used in the sequel.

The results of [2] which will be used in the present paper are the following (some of these results hold with little or no modification for F in $\mathfrak{F}$, but we need them here only for F in $\mathfrak{F}^*$):

A. (Theorem 2 of [2].) For F in $\mathfrak{F}^*$, there is a d.f. $G(\cdot\,;F)$ such that

$$\lim_{n\to\infty} G_n(r;F) = G(r;F) \tag{1.5}$$

at every continuity point of the latter. Moreover, for F in $\mathfrak{F}^*$,

$$\lim_{k\to\infty} G_{\infty,k}(r;F) = G(r;F) \tag{1.6}$$

and (obviously)

$$\lim_{k\to\infty} G_{n,k}(r;F) = G_n(r;F). \tag{1.7}$$

B. (Theorem 1 of [2].) There are positive constants c_m^* and c_m' (independent of n, F, and r) such that, for F in $\mathfrak{F}^*$, all n, and all $r \geqq 0$,

$$1 - G_n(r;F) < c_m^* e^{-c_m' r^2}. \tag{1.8}$$

Further remarks on possible values of c_m' are contained in [1] and [2].

C. For each F in $\mathfrak{F}$ there is an F_1 in $\mathfrak{F}^*$ such that, for all n and r,

$$G_n(r;F_1) \leqq G_n(r;F). \tag{1.9}$$

(This is fairly obvious; see [2] for further discussion.)

Of course, (1.8) and (1.9) also hold in the limit; i.e., with the subscript n deleted.

D. (A consequence of (3.11) of [2].) For all F in $\mathfrak{F}^*$, and for each $d > 0$,

$$G_{n,k}(r;F) - G_n(r+d;F) < \frac{c_1 k}{n} + c_2 k \exp\{-c_3 d^2 k^{1/2}\} + \frac{c_4}{d^4}\left(\frac{1}{k} + \frac{1}{n}\right), \tag{1.10}$$

where the c_i are positive constants depending only on m.

A further result of [2] will be given in Lemma 2 of Section 2, after some additional notation has been introduced.

In most of the arguments of this paper we will be dealing with F's which are in $\mathfrak{F}^c$. To simplify the discussion in such cases, we shall always assume that, for every real number t and integer j, at most one Z_i has its jth coordinate equal to t. The probability that this be not so is zero.

2. Uniformity of approach of G_n to G in the subclass $\mathfrak{F}_\epsilon$. The purpose of this section is to prove Lemma 1 (stated below), which will be used in Section 4 to prove the existence of an F_δ with the properties described in Section 1, when F is restricted to belong to a suitable subclass $\mathfrak{F}_\epsilon'$ of $\mathfrak{F}$. This and the multinomial result of Section 3 will then be used in Section 4 to demonstrate Theorem 1 with $\mathfrak{F}$ replaced by $\mathfrak{F}_\epsilon'$. The proof of Theorem 1 is then completed in Section 5 by showing that $\mathfrak{F}_\epsilon'$ is suitably dense in $\mathfrak{F}$ as $\epsilon \to 0$. Thus, although by far the greatest amount of new effort needed to prove Theorem 1 when $m > 1$ over what is needed when $m = 1$, is contained in the arguments of the present section, the reader who is interested mainly in the ideas of the statistical proof may read the statement of Lemma 1 and then go on to Section 3.

We first introduce some notation which will be used in this and subsequent sections. Let ϵ be a small positive number and let r be a positive number, both of which will be fixed in the present section. Other ϵ's with subscripts will be used in this paper to denote positive variables which will approach zero. The symbol $o(1 \mid \epsilon_i)$ is to denote a quantity which, as ϵ_i approaches zero, approaches zero uniformly in all other relevant quantities. Sometimes the latter will be explicitly indicated. Thus $o(1 \mid \epsilon_i \mid n, F)$ denotes a quantity which approaches zero, uniformly for all n (sometimes for all large n) and for all F (either in $\mathfrak{F}$ or in some indicated subclass), as $\epsilon_i \to 0$. The symbol $o(1 \mid \epsilon_i, n \mid F)$ denotes a quantity which approaches zero as $\epsilon_i \to 0$, $n \to \infty$, uniformly in F (either in $\mathfrak{F}$ or some indicated subclass). The symbol $o(1 \mid n \mid F)$ denotes a quantity which approaches zero as $n \to \infty$, uniformly in F (either in $\mathfrak{F}$ or some indicated subclass). The symbol $o(1 \mid d, N(d) \mid \cdot, \cdot)$ is to mean a quantity which approaches zero as $d \to 0$ while n stays larger than a suitable function $N(d)$ of d (which may change in various appearances of the symbol, although we shall sometimes use N, N', etc., to denote several such symbols which arise in the proof of the same lemma), and the approach of this quantity to zero is uniform in all other relevant quantities, which may be indicated where the dots are. The symbols $o(1 \mid \epsilon_i, N(\epsilon_i) \mid \cdot, \cdot)$ and $o(1 \mid k, N(k) \mid \cdot, \cdot)$ (with $k \to \infty$) will be used similarly. Finally the symbol θ will always denote a generic quantity <1 in absolute value; two θ's in different places need not be the same. The quantity d will always be >0.

Let $\mathfrak{F}_\epsilon$ be the subclass of those d.f.'s F in $\mathfrak{F}^*$ which have a Lebesgue density f_F in the subset of all points in I^m where at least one coordinate is $\geqq 1 - \epsilon$, and such that $\frac{1}{2} \leqq f_F \leqq 2$ almost everywhere in this region. The proofs of this section actually hold when $\mathfrak{F}_\epsilon$ is replaced by a somewhat larger class; but this is of little importance, the main use of Lemma 1 being to prove Theorem 1. (The relationship of $\mathfrak{F}_\epsilon'$ to $\mathfrak{F}_\epsilon$ will be stated in Section 4.)

LEMMA 1. *We have, for each fixed* m,

$$|G_n(r; F) - G(r; F)| = o(1 \mid n \mid F \,\varepsilon\, \mathfrak{F}_\epsilon). \tag{2.1}$$

The proof of Lemma 1 will require several supplementary lemmas. The proofs for all $m > 1$ are essentially the same, but the proof is most easily written out and followed in the case $m = 2$. *Hence, throughout the remainder of this section we shall carry out all proofs in the case* $m = 2$. The modifications in the statements and proofs which are necessary when $m > 2$ will usually be completely obvious;

and we shall explicitly mention, at appropriate points in the argument, those modifications which are not completely obvious.

Thus, we can write in coordinates $Z_n = (X_n, Y_n)$ and $z = (x, y)$, throughout the remainder of the section. (In most of the corresponding arguments for the case of m components, x will stand for the first $m - 1$ components of z, and y will stand for the last component of z.)

The idea of the proof of Lemma 1 is that (1.10) should somehow be used to prove Lemma 5, which, by a suitable uniformity result (Lemma 7) on the approach of the multinomial distribution to its limit, will yield (2.1). What is needed to obtain Lemma 5 from (1.10) is Lemma 4, the idea of which is that if $n^{1/2}|S_n(z) - F(z)|$ attains the value r somewhere, then it is very likely to attain the value $r + d$ somewhere, if d is small; it is the structure of $\mathfrak{F}_\epsilon$ which is used, in (2.18), to prove this.

For $0 < \epsilon_1 < 1$ we define the events

$$L_1(\epsilon_1) = \{\sup_{\substack{0 \leqq x \leqq \epsilon_1 \\ 0 \leqq y \leqq 1}} |S_n(z) - F(z)| \geqq r/n^{1/2}\}, \tag{2.2}$$

and

$$L_2(\epsilon_1) = \{\sup_{\substack{1-\epsilon_1 \leqq x \leqq 1 \\ 1-\epsilon_1 \leqq y \leqq 1}} |S_n(z) - F(z)| \geqq r/n^{1/2}\}. \tag{2.3}$$

(For the case of vectors with m components, the supremum in (2.2) is taken over the set where at least one of the $m - 1$ components of x is $\leqq \epsilon_1$; in (2.3), it is taken over the set where all m components are $\geqq 1 - \epsilon_1$.)

The next two lemmas lead up to Lemma 4.

LEMMA 2. *We have*

$$P_F\{L_1(\epsilon_1)\} = o(1 \mid \epsilon_1, n \mid F \varepsilon \mathfrak{F}^*), \tag{2.4}$$

and

$$P_F\{L_2(\epsilon_1)\} = o(1 \mid \epsilon_1, n \mid F \varepsilon \mathfrak{F}^*). \tag{2.5}$$

PROOF. An upper bound on the probability of $L_1(\epsilon_1)$ can be obtained from equations (3.6), (3.9), and (3.10) of [2], if, in the latter, we set $h = 0, j = 1, k = 1/\epsilon_1$ (the relevant argument of [2] is valid even if k is not an integer), $d = r$. We obtain

$$\begin{aligned} P_F\{L_1(\epsilon_1)\} &< \frac{1}{n} + c_o \exp\{-cr^2/2\epsilon_1^{1/2}\} \\ &\quad + \frac{16}{r^4}\left(3\epsilon_1^2 + \frac{\epsilon_1}{n}\right) = o(1 \mid \epsilon_1, n \mid F \varepsilon \mathfrak{F}^*). \end{aligned} \tag{2.6}$$

We shall now use an argument like that by which (2.22) of [2] was proved, in order to prove (2.5). The event $L_2(\epsilon_1)$ implies the occurrence of at least one of the following events:

$$L_2^1 = \Big\{ \sup_{\substack{1-\epsilon_1 \leqq x \leqq 1 \\ 1-\epsilon_1 \leqq y \leqq 1}} \mid (\text{number of } Z_1, \cdots, Z_n \text{ which satisfy } 0 \leqq X_i \leqq x, y \leqq Y_i \leqq 1) - \text{expected number} \mid \geqq \frac{rn^{1/2}}{3} \Big\},$$

$$(2.7) \quad L_2^2 = \Big\{ \sup_{\substack{1-\epsilon_1 \leqq x \leqq 1 \\ 1-\epsilon_1 \leqq y \leqq 1}} \mid (\text{number of } Z_1, \cdots, Z_n \text{ which satisfy } x \leqq X_i \leqq 1, y \leqq Y_i \leqq 1) - \text{expected number} \mid \geqq \frac{rn^{1/2}}{3} \Big\},$$

$$L_2^3 = \Big\{ \sup_{\substack{1-\epsilon_1 \leqq x \leqq 1 \\ 1-\epsilon_1 \leqq y \leqq 1}} \mid (\text{number of } Z_1, \cdots, Z_n \text{ which satisfy } x \leqq X_i \leqq 1, 0 \leqq Y_i \leqq y) - \text{expected number} \mid \geqq \frac{rn^{1/2}}{3} \Big\}.$$

The random variables in the original sequence $\{Z_j\}$ all have the same distribution as $Z_1 = (X_1, Y_1)$. Apply the argument by which (2.4) was obtained for sequences all of whose members have the same distribution as each of the following, in order: $(1 - Y_1, X_1)$, $(1 - X_1, 1 - Y_1)$, and $(1 - X_1, Y_1)$. We obtain that

$$(2.8) \qquad P_F\{L_2^i\} = o(1 \mid \epsilon_1, n \mid F \in \mathfrak{F}^*), \qquad i = 1, 2, 3.$$

Hence (2.5) is verified.

Define the events

$$(2.9) \qquad L_3(\epsilon_1) = \Big\{ \sup_{0 \leqq x \leqq 1} |S_n(x, 1) - x| < \frac{1}{\epsilon_1 n^{1/2}} \Big\}$$

and

$$(2.10) \qquad L_4(\epsilon_1) = \Big\{ \sup_{\substack{\epsilon_1 \leqq x \leqq 1 \\ 0 \leqq y \leqq 1-\epsilon_1}} |S_n(z) - F(z)| \geqq \frac{r}{n^{1/2}} \Big\}.$$

From (1.8) we obtain that

$$(2.11) \qquad P_F\{L_3(\epsilon_1)\} = 1 - o(1 \mid \epsilon_1 \mid n, F \in \mathfrak{F}^*).$$

Write $L(\epsilon_1) = L_3(\epsilon_1) \cap L_4(\epsilon_1)$. Whenever $L(\epsilon_1)$ occurs we can define chance variables H and T as follows: $H = h$, $\epsilon_1 \leqq h \leqq 1$, and $T = t$, $0 < t \leqq 1 - \epsilon_1$, if

$$(2.12) \qquad |S_n(h, t) - F(h, t)| \geqq r/n^{1/2}$$

and

$$(2.13) \qquad |S_n(h', t') - F(h', t')| < r/n^{1/2}$$

for $\epsilon_1 \leqq h' \leqq 1$, $0 \leqq t' < t$, as well as for $\epsilon_1 \leqq h' < h$, $t' = t$. (In the m-component case, h' has all $m - 1$ components $\geqq \epsilon_1$ and h can be specified by any rule which does not depend on y for $y > t$, and such that (2.12) holds.) Thus, if a horizontal line $y = t'$ is swept upward starting at $t' = 0$, the line $y = t$ is the first for which (2.12) can hold, and h is a well-defined value such that it does.

LEMMA 3. *We have, for some* $N(d)$ and $\epsilon_1 = d^{1/4}$,

$$P_F\left\{ \sup_{T<y\leqq 1} |S_n(H, y) - F(H, y)| \geqq \frac{r+d}{n^{1/2}} \,\middle|\, L(\epsilon_1)\right\} = 1 - o(1 \mid d, N(d) \mid F \,\varepsilon\, \mathfrak{F}_\epsilon) . \tag{2.14}$$

PROOF. We suppose that

$$S_n(H, T) - F(H, T) = r/n^{1/2} \tag{2.15}$$

and we will prove that, conditional on $L(\epsilon_1)$ occurring, the probability that

$$S_n(H, y) - F(H, y) \geqq (r + d)/n^{1/2} \tag{2.16}$$

for some y, $T < y \leqq 1$, is $1 - o(1 \mid d, N(d) \mid F \,\varepsilon\, \mathfrak{F}_\epsilon)$. This will be enough to prove (2.14), for (a) if the left member of (2.15) is greater than $r/n^{1/2}$ the result we want to prove is a fortiori true, and (b) if the left member of (2.15) is $\leqq -r/n^{1/2}$, it is proved, in the same way as below, that the probability (conditional on $L(\epsilon_1)$ occurring) that the left member of (2.16) be $\leqq -(r + d)/n^{1/2}$ for some y, $T < y \leqq 1$, is $1 - o(1 \mid d, N(d) \mid F \,\varepsilon\, \mathfrak{F}_\epsilon)$.

Define

$$\begin{aligned} n_1 &= n(S_n(H, 1) - S_n(H, T)), \\ \bar{y} &= \frac{F(H, y) - F(H, T)}{H - F(H, T)}, \\ n_2(\bar{y}) &= n(S_n(H, y) - S_n(H, T)). \end{aligned} \tag{2.17}$$

From (2.9), (2.10), (2.15), and the definition of $\mathfrak{F}_\epsilon$, we have, in $L(\epsilon_1)$, if $\epsilon_1 < \epsilon$,

$$n_1 = n(H - F(H, T)) + \frac{\theta n^{1/2}}{\epsilon_1} - rn^{1/2} > n\epsilon_1^2/2 - n^{1/2}\left(\frac{1}{\epsilon_1} + r\right), \tag{2.18}$$

which goes to ∞ as $n \to \infty$ (uniformly in H and $T \leqq 1 - \epsilon_1$), and is thus arbitrarily large for $n >$ some $N'(\epsilon_1)$. Using (2.15) we find that (2.16) is equivalent to

$$\frac{n_2(\bar{y})}{n_1} - \frac{n\bar{y}(H - F(H, T))}{n_1} \geqq \frac{dn^{1/2}}{n_1}. \tag{2.19}$$

From (2.18) we obtain that the probability that (2.19) occur for some $\bar{y}$, $0 \leqq \bar{y} \leqq 1$, is $\geqq$ the probability that, for some $\bar{y}$,

$$\frac{n_2(\bar{y})}{n_1} - \bar{y} \geqq \frac{2d\,\epsilon_1^{-1}}{n_1^{1/2}} + \frac{4\epsilon_1^{-2}\bar{y}}{n_1^{1/2}}, \tag{2.20}$$

provided that ϵ_1 is small enough and $n >$ a suitable $N_1(\epsilon_1)$.

Now set

$$\epsilon_1 = d^{1/4} \tag{2.21}$$

and suppose that d is small enough that (2.20) holds when ϵ_1 is given by (2.21).

Let $\{W(t), 0 \leqq t < \infty\}$ be the separable (Wiener) process with independent, normally distributed increments, $W(o) = 0$, $E(W(t)) = 0$, $\text{Var}(W(t)) = t$. Given H, T, and n_1, the left member of (2.20) clearly is distributed as the difference between a sample d.f. and the uniform d.f. on the one-dimensional interval $0 \leqq \bar{y} \leqq 1$, when the sample d.f. is that of n_1 independent, uniformly distributed random variables. It follows from [3] and [4] and the fact that $n_1 \to \infty$ as $n \to \infty$ that, under (2.21), the conditional probability (given that $L(\epsilon_1)$ occurs) that (2.20) hold for some $\bar{y}$ approaches, uniformly in $\mathfrak{F}_\epsilon$, as $n \to \infty$,

$$P\{W(t) \geqq 2d^{3/4} + (2d^{3/4} + 4d^{-1/2})t \text{ for some } t > 0\}. \tag{2.22}$$

The latter is, by [4], equation (4.2),

$$\exp\{-2(2d^{3/4})(2d^{3/4} + 4d^{-1/2})\} \tag{2.23}$$

which approaches one as $d \to 0$. Hence, for d sufficiently small and $n >$ some $N(d)$, the conditional probability (given that $L(\epsilon_1)$ occurs) that (2.16) holds for some $y > T$ is arbitrarily close to 1, uniformly in $\mathfrak{F}_\epsilon$. This proves (2.14).

LEMMA 4. *We have, for some* $N(d)$,

$$G_n(r + d; F) - G_n(r; F) = o(1 \mid d, N(d) \mid F \varepsilon \mathfrak{F}_\epsilon). \tag{2.24}$$

PROOF. Substituting (2.21) into (2.2), (2.4), (2.9), (2.10) and (2.11) (none of which previously depended on d in any way), and using Lemma 3, we have

$$P_F\{r \leqq \sup_{\substack{0 \leqq x \leqq 1 \\ 0 \leqq y \leqq 1 - d^{1/4}}} \sqrt{n}|S_n(z) - F(z)| \leqq r + d\} = o(1 \mid d, N(d) \mid F \varepsilon \mathfrak{F}_\epsilon), \tag{2.25}$$

where of course the $N(d)$ may differ from that of Lemma 3. Now, the definition of $\mathfrak{F}_\epsilon$ is such that, by interchanging the roles of x and y, we obtain, in the same way that (2.25) was obtained,

$$P_F\{r \leqq \sup_{\substack{0 \leqq x \leqq 1 - d^{1/4} \\ 0 \leqq y \leqq 1}} \sqrt{n}|S_n(z) - F(z)| \leqq r + d\} = o(1 \mid d, N(d) \mid F \varepsilon \mathfrak{F}_\epsilon). \tag{2.26}$$

(In the case of vectors with m components, there are $m - 2$ additional analogues of (2.26).) Finally, substituting (2.21) into (2.5) and combining the result with (2.25) and (2.26), we obtain (2.24).

LEMMA 5. *We have*

$$0 \leqq G_{n,k}(r; F) - G_n(r; F) = o(1 \mid k, N'(k) \mid F \varepsilon \mathfrak{F}_\epsilon). \tag{2.27}$$

PROOF. The left side of (2.27) is trivial. Adding (1.10) and (2.24), we have

$$\begin{aligned} G_{n,k}(r; F) - G_n(r; F) &\leqq o(1 \mid d, N(d) \mid F \varepsilon \mathfrak{F}_\epsilon) \\ &\quad + c_2 k \exp\{-c_3 d^2 k^{1/2}\} + c_4/d^4 k + c_1 k/n + c_4/d^4 n. \end{aligned} \tag{2.28}$$

Let $\epsilon' > 0$ be given arbitrarily. Let $d_1 > 0$ be such that the first term on the right side of (2.28) is $< \epsilon'/3$ if $d = d_1$ and $n > N(d_1)$. Let k_1 be such that the sum of the next two terms on the right side of (2.28) is $<\epsilon'/3$ when $d = d_1$ and $k > k_1$. For $k > k_1$, let $N'(k)$ be $>N(d_1)$ and be such that the sum of the last two terms of (2.28) is $<\epsilon'/3$ when $d = d_1$ and $n > N'(k)$. Then, putting $d = d_1$, we have that the right side of (2.28) is $<\epsilon'$ when $k > k_1$ and $n > N'(k)$. Thus, Lemma 5 is proved.

The discussion which immediately follows, as well as Lemma 6, leads up to the proof of Lemma 7.

There are k^2 cells into which I^2 is divided by the lines $x = i/k$, $y = j/k$, $i, j = 0, 1, \cdots, k$. (There are, of course, k^m cells in the case of vectors with m components.) Number the cells as follows: The cell bounded by $x = (i-1)/k$, $x = i/k$, $y = (j-1)/k$, $y = j/k$, is to be called the (i, j) cell. Write

$$\pi_{Fij} = P_F\{Z_1 \varepsilon \text{ cell } (i, j)\}.$$

Write $(i', j') \leqq (i, j)$ if $i' \leqq i, j' \leqq j$, and write $(i', j') < (i, j)$ if $(i', j') \leqq (i, j)$ and either $i' < j$ or $j' < j$. Let $\bar{H}$ be any collection of cells. For any fixed (i_0, j_0) not in $\bar{H}$, there clearly exist integers c_{ij} (depending only on $\bar{H}$ and (i_0, j_0)) such that we can write

$$(2.29)\quad F(i_0/k, j_0/k) = \sum_{\substack{(i,j) \leqq (i_0,j_0) \\ (i,j) \varepsilon H}} c_{ij} F(i/k, j/k) + \sum_{\substack{(i,j) \leqq (i_0,j_0) \\ (i,j) \notin H}} c_{ij}\, \pi_{Fij},$$

identically in F (i.e., in the π_{Fij}).

Let $\epsilon_2 > 0$ be given. Call the cell (i, j) regular if $\pi_{Fij} \geqq \epsilon_2$ and $(i, j) \neq (k, k)$. Call the cell (i, j) singular if $\pi_{Fij} < \epsilon_2$ and $(i, j) \neq (k, k)$. Let $\bar{H}_F$ be the collection of regular cells, let (i_0, j_0) be singular under F, and let the c_{ij} be as in (2.29). Denote a summation over the region $(i, j) \leqq (i_0, j_0)$, $(i, j) \varepsilon \bar{H}_F$ by $\sum^{(F, i_0, j_0)}$. Then, clearly,

$$(2.30)\qquad |F(i_0/k, j_0/k) - \textstyle\sum^{(F, i_0, j_0)} c_{ij} F(i/k, j/k)| < h(k)\epsilon_2,$$

where h is a suitable positive function of k alone, which can be chosen so that (2.30) is valid for every ϵ_2, every F, and every (i_0, j_0) singular for such an F; here the $\bar{H}_F$ depends on the F and ϵ_2 being considered, but the c_{ij} depend on these quantities only through $\bar{H}_F$.

Define $Q_{F,n}(\epsilon_2)$ to be the probability that

$$(2.31)\qquad \begin{aligned} &|S_n(i/k, j/k) - F(i/k, j/k)| < r/n^{1/2} \text{ for all } (i, j) \text{ in } \bar{H}_F, \\ &|\textstyle\sum^{(F, i_0, j_0)} c_{ij}[S_n(i/k, j/k) - F(i/k, j/k)]| < r/n^{1/2} \\ &\qquad\qquad \text{for all } (i, j) \neq (k, k) \text{ and not in } \bar{H}_F. \end{aligned}$$

The proof of the next lemma is actually valid when $\mathfrak{F}^*$ is replaced by the class of all d.f.'s on I^2.

LEMMA 6. *We have*

$$(2.32)\qquad |Q_{F,n}(\epsilon_2) - G_{n,k}(r; F)| = o(1 \mid \epsilon_2, N(\epsilon_2) \mid F \varepsilon \mathfrak{F}^*).$$

PROOF. Define, for (i_0, j_0) singular,

$$U = S_n(i_0/k, j_0/k) \tag{2.33}$$

and

$$V = \Sigma^{(F, i_0, j_0)} c_{ij} S_n(i/k, j/k). \tag{2.34}$$

Let B be the event defined by

$$B = \{|(U - V) - E(U - V)| < \epsilon_2^{1/4}/n^{1/2}\}. \tag{2.35}$$

Now, $U - V$ is just the last sum of (2.29) with π_{Fij} replaced by n^{-1} (number of $Z_1, \cdots, Z_n$ falling in cell (i, j)). Hence, $U - V$ has variance $< h'(k)\epsilon_2/n$, where h' is a suitable positive function of k. Thus, by Chebyshev's inequality,

$$P_F\{B\} > 1 - h'(k)\epsilon_2^{1/2}. \tag{2.36}$$

Of course, the definition of B depends on F, was well as on (i_0, j_0); but, again, h' can be chosen so that (2.36) holds for all F. Consider the events

$$A_1 = \{|U - EU| \geqq r/n^{1/2}, |V - EV| < r/n^{1/2}\} \tag{2.37}$$

and

$$A_2 < \{|U - EU| < r/n^{1/2}, |V - EV| \geqq r/n^{1/2}\}. \tag{2.38}$$

The definition of these events also depends on F and (i_0, j_0). Define $P_{Ft} = P_F\{A_t\}$, $t = 1, 2$. We are first going to show that, for all F for which (i_0, j_0) is singular,

$$P_{Ft} = o(1 \mid \epsilon_2, N(\epsilon_2) \mid F \varepsilon \mathfrak{F}^*), \qquad \text{for } t = 1, 2. \tag{2.39}$$

In proving this, let W stand for U in the case $t = 1$ and for V in the case $t = 2$. Then W is n^{-1} times the sum of n independent, identically distributed random variables, each bounded in absolute value by some constant L (independent of F). Let σ^2 be the variance and β_3 the absolute third moment about its expected value of each summand (i.e., of $W - EW$ when $n = 1$). Now, if $\sigma^2 < \epsilon_2^{1/8}$, Chebyshev's inequality yields $P_{Ft} < r^{-2}\epsilon_2^{1/8}$, so that (2.39) is verified in that case. On the other hand, if $\sigma^2 \geqq \epsilon_2^{1/8}$, by (2.36) we have

$$\begin{aligned} P_{Ft} &\leqq P_F\{B \cap A_t\} + P_F\{\bar{B}\} < P_F\{B \cap A_t\} + h'(k)\epsilon_2^{1/2} \\ &\leqq P_F\{r \leqq n^{1/2}|W - EW| \leqq r + \epsilon_2^{1/4}\} + h'(k)\epsilon_2^{1/2} \\ &\leqq P_F\{r/\sigma \leqq n^{1/2}|W - EW| / \sigma \leqq r/\sigma + \epsilon_2^{3/16}\} + h'(k)\epsilon_2^{1/2}. \end{aligned} \tag{2.40}$$

By the Berry-Esseen estimate (see, e.g., [5]) and the fact that $\beta_3/\sigma^3 \leqq L/\sigma$, we have from (2.40) for all F for which $\sigma^2 \geqq \epsilon_2^{1/8}$,

$$P_{Ft} \leqq h'(k)\epsilon_2^{1/2} + \epsilon_2^{3/16} + c_5 L n^{-1/2}\epsilon_2^{-1/16}, \tag{2.41}$$

where c_5 is a positive constant. Thus, (2.39) is proved.

Lemma 6 follows at once from (2.39).

The proof of the next lemma is also valid when $\mathfrak{F}^*$ is replaced by the class of all d.f.'s on I^2.

LEMMA 7. *For any fixed positive integer k, we have*

$$|G_{\infty,k}(r; F) - G_{n,k}(r; F)| = o(1|n|F \varepsilon \mathfrak{F}^*). \tag{2.42}$$

PROOF. Let $\epsilon_3 > 0$ be given arbitrarily. Choose ϵ_2 so small and $N(\epsilon_2)$ so large that, for this value of ϵ_2, the left side of (2.32) is $\leqq \epsilon_3/4$ for all F when $n > N(\epsilon_2)$. We shall show below that, writing $Q_F(\epsilon_2) = \lim_{n\to\infty} Q_{F,n}(\epsilon_2)$, we have

$$|Q_F(\epsilon_2) - Q_{F,n}(\epsilon_2)| < \epsilon_3/2 \tag{2.43}$$

for n sufficiently large, uniformly in F. Hence, we shall have, for n sufficiently large, uniformly in F, that the left side of (2.42) is no greater than

$$\begin{aligned} &|G_{\infty,k}(r; F) - Q_F(\epsilon_2)| + |Q_F(\epsilon_2) - Q_{F,n}(\epsilon_2)| \\ &\qquad + |Q_{F,n}(\epsilon_2) - G_{n,k}(r; F)| < \epsilon_3/4 + \epsilon_3/2 + \epsilon_3/4 = \epsilon_3, \end{aligned} \tag{2.44}$$

and (2.42) will be proved.

We shall now fix $\bar{H}$ and prove that (2.43) holds, uniformly in all F for which $\bar{H}_F = \bar{H}$, for n sufficiently large. Since k is fixed, the number of possible choices of $\bar{H}$ is finite, so that Lemma 7 will be proved.

Consider the joint distribution of the $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$ for all regular (i, j), which, as $n \to \infty$, approaches a multivariate normal distribution. Since $\pi_{Fij} \geqq \epsilon_2$ for any regular point it follows that the determinant of the covariance matrix of the $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$ (for regular (i, j)) is bounded away from 0 (and, of course, from ∞ as well) by a function of ϵ_2, uniformly in all F for which $\bar{H}_F = \bar{H}$. It follows from [6], page 121, that the maximum of the absolute value of the difference between the joint d.f. of these $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$ and their limiting multivariate normal d.f. is less than $n^{-1/2}M(\epsilon_2)$, where M is a real function of ϵ_2 only. The maximum of the density of this limiting normal d.f. is a real function only of ϵ_2, say $M'(\epsilon_2)$. Thus, the statements in the last two sentences are uniform in all F for which $\bar{H}_F = \bar{H}$.

It follows from (1.8) that the probability of a sufficiently large cube C in the space of the $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$ (for all regular (i, j)) which is centered at the origin, is greater than $1 - \epsilon_3/12$ uniformly in F and n. Hence this is also true of the limiting multivariate normal d.f. of the $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$.

Consider the region R in the space of these $n^{1/2}(S_n(i/k, j/k) - F(i/k, j/k))$, which is defined by (2.31) and whose probability is $Q_{F,n}(\epsilon_2)$. The region $R \cap C$ is a bounded polyhedron and can be approximated from within by a finite union R_1 of "rectangles" with sides parallel to the coordinate planes, such that the volume of the region $R_2 = [(R \cap C) - R_1]$ is $< \epsilon_4$, where $\epsilon_4 > 0$ is such that $\epsilon_4 M'(\epsilon_2) < \epsilon_3/12$. The set R_2 can be covered by a finite union R_3 of rectangles with sides parallel to the coordinate planes whose total volume is $< 2\epsilon_4$. Let m_3 be the number of rectangles in R_3, and m_1 be the number of rectangles in R_1.

The probability of R_3 according to the limiting normal d.f. is less than

$$2\epsilon_4 M'(\epsilon_2) < \epsilon_3/6. \tag{2.45}$$

The probability $P_F\{R_2\}$ of the region R_2 according to F is $<P_F\{R_3\}$, which, by the aforementioned result of Bergstrom [6], differs from the probability of R_3 according to the limiting normal d.f. by less than $4m_3M(\epsilon_2)n^{-1/2}$. Hence,

$$P_F\{R_2\} < \epsilon_3/6 + 4m_3M(\epsilon_2)n^{-1/2}. \tag{2.46}$$

Also, by Bergstrom's result just cited, the probability of R_1 according to the limiting normal d.f. differs from $P_F\{R_1\}$ by less than $4m_1M(\epsilon_2)n^{-1/2}$. Since the sum of this and the second term in the right member of (2.46) can be made less than $\epsilon_3/6$ by making n sufficiently large, it follows from the present paragraph and the previous two paragraphs that (2.43) holds for n sufficiently large, uniformly in all F for which $\bar{H}_F = \bar{H}$. This completes the proof of Lemma 7.

Proof of Lemma 1. Let $\epsilon_5 > 0$ be chosen arbitrarily. Choose k' such that the right side of (2.27) is $<\epsilon_5/3$ for $k = k'$ and $n > N'(k')$. In particular, $0 \leqq G_{\infty,k'}(r; F) - G(r; F) \leqq \epsilon_5/3$. Choose N to be $>N'(k')$ and such that, for $k = k'$, the left member of (2.42) is $<\epsilon_5/3$ for $n > N'$. Then, for $n > N'$ and all F in $\mathfrak{F}_\epsilon$, we have

$$\begin{aligned} |G_n(r; F) - G(r; F)| &\leqq |G_n(r; F) - G_{n,k'}(r; F)| \\ &+ |G_{n,k'}(r; F) - G_{\infty,k'}(r; F)| + |G_{\infty,k'}(r; F) - G(r; F)| < \epsilon_5 . \end{aligned} \tag{2.47}$$

Since ϵ_5 was arbitrary, Lemma 1 is proved.

3. The multinomial result. We have mentioned in Section 1 that the main results of this paper are obtained by approximating the original problem by an appropriate multinomial problem. In the present section we summarize the needed multinomial results which were obtained in [1], and sketch the ideas of the proofs, unencumbered by the tedious details of [1]. Actually, we do not need the full strength of the results of [1], which are broader than those of Lemma 8 below in that, in the derivation of Section 3 of [1], the calculations were carried out in fine detail in order to obtain an error term which can be used to calculate an upper bound on the departure of ϕ_n^* from minimax character (in view of the lack of knowledge about the distribution of D_n, it seems more difficult to obtain a useful bound of this kind when $m > 1$). In fact, if one does not bother to obtain an error term, it is obvious how to shorten considerably the proof of the multinomial result in Section 3 of [1], and we shall see that this simple multinomial result without error term rests mainly on a result of v. Mises ([7], especially pages 84–86) which is almost forty years old.

We now introduce the needed notation. Let h be a positive integer and let B_h be the family of $(h + 1)$-vectors $\pi = \{p_i, 1 \leqq i \leqq h + 1\}$ with real components satisfying $p_i \geqq 0$, $\sum p_i = 1$. Let B_h' be a specified subset of $B_h \cdot B_h'$ can actually be fairly arbitrary in structure; to avoid trivial circumlocutions, we shall suppose in this section that B_h' is the closure of an h-dimensional open subset of B_h, although it will be obvious that Lemmas 8, 9, and 10 hold much more generally. Let $T^{(n)} = \{T_i^{(n)}, 1 \leqq i \leqq h + 1\}$, a vector of $h + 1$ chance variables, have a multinomial probability function arising from n observations with $h + 1$ possible outcomes, according to some π in B_h'; i.e., for integers $x_i \geqq 0$ with $\Sigma_1^{h+1} x_i = n$,

$$P_\pi\{T_i^{(n)} = x_i, 1 \leqq i \leqq h + 1\} = \frac{n!}{x_1! \cdots x_{h+1}!} p_1^{x_1} \cdots p_{h+1}^{x_{h+1}}. \tag{3.1}$$

Let L be a positive integer, let γ_i be an $(h + 1)$-vector, $1 \leqq i \leqq L$, and let $\rho_i = \gamma_i'\pi$ (scalar product) be corresponding linear functions of π, $1 \leqq i \leqq L$. To avoid trivialities, we assume at least one ρ_i is not constant on B_h'. Let $\mathcal{E}_n$ be the class of all (possibly randomized) vector estimators of $\rho = \{\rho_i, 1 \leqq i \leqq L\}$, the weight function (which depends on n) being the simple one for which the risk function of a procedure ψ_n in $\mathcal{E}_n$ is

$$1 - P_{\pi,\psi_n}\{|d_i - \rho_i| \leqq r/n^{1/2}, 1 \leqq i \leqq L\}, \tag{3.2}$$

where r is a positive value and we have written $d = \{d_i, 1 \leqq i \leqq L\}$ for the vector of decisions. Let ψ_n^* be the nonrandomized estimator whose ith component is $\gamma_i'T^{(n)}/n$ (the allowable decisions may be restricted to $\gamma'\pi$ for π in B_h' with only trivial modifications in what follows). Finally, a point π in B_h' is called an interior point if all its components p_i are positive, and if it has a neighborhood (in B_h) which is a subset of B_h'. The required multinomial result is:

LEMMA 8. *For any interior point π^* of B_h' there is a sequence $\{\xi_n\}$ of a priori distributions on B_h' converging in distribution to the distribution which gives probability one to π^* and such that $\{\psi_n^*\}$ is asymptotically Bayes relative to $\{\xi_n\}$ as $n \to \infty$, uniformly for $0 \leqq r \leqq R$ for any $R < \infty$; i.e., such that, uniformly in such r,*

$$\lim_{n\to\infty} \frac{\int P_\pi\{|\rho_i - \gamma_i'T^{(n)}/n| > r/n^{1/2}, 1 \leqq i \leqq L\}\, \xi_n(d\pi)}{\inf_{\psi_n \varepsilon \mathcal{E}_n} \int P_{\pi,\psi_n}\{|d_i - \rho_i| > r/n^{1/2}, 1 \leqq i \leqq L\}\, \xi_n(d\pi)} = 1. \tag{3.3}$$

Of course, continuity considerations show that the positive (since not all ρ_i are constant) limit of the numerator of (3.3) is obtained by putting $\pi = \pi^*$ instead of integrating with respect to ξ_n, and then using the multivariate central limit theorem to compute the limiting probability.

The idea of the proof of Lemma 8 is very simple. Let Γ^* be the (nonsingular) covariance matrix of the limiting h-variate normal distribution of $n^{1/2}(n^{-1}T_i^{(n)} - p_i^*)$, $1 \leqq i \leqq h$, when $\pi = \pi^*$. Let ϵ be a small positive value and let ξ be the uniform *a priori* distribution in the (solid) sphere of radius ϵ about π^* in B_h'. (ϵ is small enough that this sphere consists entirely of interior points.) According to the result [7] of v. Mises, for any π'' in this sphere, with probability one when $T^{(n)}$ is distributed according to π'', the *a posteriori* density function of $n^{1/2}(p_i - T_i^{(n)}/n)$, $1 \leqq i \leqq h$ (calculated assuming ξ to be the *a priori* distribution) will tend to the h-variate normal density with means 0 and covarance matrix Γ'' (corresponding to π'') as $n \to \infty$. If the *a posteriori* density were really normal with the stated parameters, it would follow at once from a result [8] of Anderson that the *a posteriori* probability of the event

$$\{n^{1/2}|\rho_i - d_i| \leqq r, 1 \leqq i \leqq L\} \tag{3.4}$$

(this probability is unity minus the *a posteriori* risk) is a maximum for $d = \gamma' T^{(n)}/n$, since the region (3.4) is for each d a convex symmetric (about a point depending on d) subset in the space of the h variables $n^{1/2}(p_i - T_i^{(n)}/n)$ (considering the latter to be unrestricted in magnitude). Since the actual *a posteriori* density is almost normal (with high probability as $n \to \infty$), ψ_n^* will be asymptotically Bayes. Finally, let ξ_n be the ξ just described when $\epsilon = \epsilon_n$, where ϵ_n goes to zero slowly enough that the above result still holds for ψ_n^* as $n \to \infty$. (For example, $\epsilon_n = n^{-\alpha}$ with $0 < \alpha < \frac{1}{2}$. The crucial consideration is that the radius $n^{1/2}\epsilon_n$ of the set of possible values of $n^{1/2}(\pi - \pi^*)$ approach infinity with n, as will therefore the radius of the set of possible values of $n^{1/2}(\pi - T^{(n)}/n)$ w.p.1 under ξ_n. The asymptotic problem is thus approximately one of estimating the mean of a multivariate normal distribution with known constant covariance matrix, when the mean can take on any value in an appropriate Euclidean space).

The actual proof—the precise handling of the approximations mentioned above, the uniformity in r, etc.—may be handled as in [1] or by complementing with appropriate estimates the argument of [7], but the main idea is really the simple one of [7].

The reason for wanting Lemma 8 in its stated form with the sequence $\{\xi_n\}$ shrinking down on π^* has to do with the problem of multinomial minimax estimation for the risk function (3.2). Let π^0 be the value of π at which the positive limit b (as $n \to \infty$) of the continuous risk function of ψ_n^* is a maximum. Since the ρ_i are not all constant, for any $\delta > 0$ there will, by continuity, be an interior point π^* of B_h' at which the limit of the risk function of ψ_n^* is at most $(1 + \delta)b$. From Lemma 8 and the sentence following (3.3) we conclude:

LEMMA 9. $\{\psi_n^*\}$ *is asymptotically minimax relative to* B_h' *and the risk function* (3.2).

We next consider a generalization of this result to other weight functions which are nondecreasing functions of $\max_i |d_i - \rho_i|$. Of course, the risk function is defined in the usual way. (A Bayes result analogous to Lemma 8 can be proved in the course of the demonstration, but we shall not bother to state it.) Let C_0 and C be positive constants such that

$$P_\pi\{n^{1/2} \max_i \gamma_i' |T^{(n)}/n - \pi| \geqq r\} < C_0 e^{-Cr^2} \tag{3.5}$$

for all r, all n, and all π in B_h'. The existence of such positive constants (which depend on h and the structure of B_h') follows from well known results on the multinomial (or, in fact, the binomial) distribution; in Section 4 we shall actually refer to (1.8) for appropriate values of these constants.

LEMMA 10. *Let $W(r)$ be a nondecreasing real function of r for $r \geqq 0$, not identically zero, and satisfying*

$$\int_0^\infty W(r)re^{-Cr^2}\,dr < \infty. \tag{3.6}$$

Then $\{\psi_n^*\}$ is asymptotically minimax relative to B_h' and the weight functions

$$W_n(\pi, d) = W(n^{1/2} \max_i |\rho_i - d_i|). \tag{3.7}$$

The proof of Lemma 10 can be carried out, starting from scratch, along lines like those of Lemma 8. An easier proof, which was given in [1], rests upon the idea of reducing the proof essentially to that for the simple weight function already considered in Lemma 8. Specifically, if the *a posteriori* distribution of the variables $n^{1/2}(p_i - T_i^{(n)}/n)$ were actually normal with means 0 and the appropriate covarance matrix, then $d_i = \gamma_i' T^{(n)}/n$, $1 \leqq i \leqq L$, would minimize the *a posteriori* risk; for, if this choice of the d_i did not minimize the *a posteriori* risk and if H_1 and H_2 were respectively, the d.f.'s of $n^{1/2} \max_i |\rho_i - d_i|$ for the above choice of d_i and for a better choice, we would have

$$\int_0^\infty W(r)\, d[H_1(r) - H_2(r)] > 0,$$

which is easily seen to imply that $H_1(r') < H_2(r')$ for some r', contradicting Anderson's result cited previously (i.e., when the error terms are included, this contradicts the result of Lemma 8). The details of the proof are contained in Section 4 of [1].

We note that Lemma 10 exemplifies a principle which is of more general use in statistics: If one can verify suitable (asymptotic) Bayes results for an appropriate class of simple weight functions, the results will automatically hold for a general class of monotone weight functions.

We remark that Anderson's result can be used to prove the result of Lemma 10 for a larger class of weight functions, namely, every function of $n^{1/2}(d_i - \rho_i)$, $1 \leqq i \leqq L$, which is symmetric about the origin and which for each real value c has a convex (or empty) set for the domain where the function is $\leqq c$.

4. Proof of Theorem 1 when $\mathfrak{F}$ is replaced by $\mathfrak{F}_\epsilon'$. Define $\mathfrak{F}_\epsilon'$ to consist of every d.f. in $\mathfrak{F}^c$ which gives probability one to I^m and which can be realized as the d.f. of Z_1' (say) when Z_1 has a d.f. in $\mathfrak{F}_\epsilon$ and Z_1' is obtained from Z_1 by continuous monotonic transformations on the individual coordinate functions. Thus, $\mathfrak{F}_\epsilon' \supset \mathfrak{F}_\epsilon$, but $\mathfrak{F}_\epsilon'$ includes d.f.'s which are not in $\mathfrak{F}^*$. Clearly, for any F' in $\mathfrak{F}_\epsilon'$ there is an F in $\mathfrak{F}_\epsilon$ such that $G_n(r; F) = G_n(r; F')$ for all n and r.

In this section we use the results of Sections 2 and 3 to prove the following

LEMMA 11. *For m a positive integer, suppose that*

$$W_n(F, g) = W(n^{1/2} \sup_z |F(z) - g(z)|), \tag{4.1}$$

where $W(r)$ for $r \geqq 0$ is continuous, nonnegative, nondecreasing in r, not identically *zero, and satisfies* (1.4). *Then, for each* ϵ *with* $0 < \epsilon < 1$, $\{\phi_n^*\}$ *is asymptotically minimax relative to* $\{W_n\}$ *and* $\mathfrak{F}'_\epsilon$.

PROOF. We divide the proof into three paragraphs; ϵ is fixed in what follows.

1. By (1.4), (1.8), the last sentence of the first paragraph of this Section, and Lemma 1, the function $r_n(F, \phi_n^*)$ approaches a bounded limit as $n \to \infty$, *uniformly* for F in $\mathfrak{F}'_\epsilon$. (This limit is positive, by the known results in the case $m = 1$.) Hence, for any $\delta > 0$, there is a d.f. F_δ in $\mathfrak{F}_\epsilon$ and an integer N_δ such that

$$\sup_{F \varepsilon \mathfrak{F}'_\epsilon} r_n(F, \phi_n^*) < (1 + \delta) r_n(F_\delta, \phi_n^*) \tag{4.2}$$

for $n > N_\delta$. Define

$$r_{nk}(F, \phi_n) = E_{F,\phi_n} W(n^{1/2} \sup_{z \varepsilon A_k^m} |F(z) - g(z)|), \tag{4.3}$$

so that

$$r_{nk}(F, \phi_n^*) = \int_0^\infty W(r) d_r G_{n,k}(r; F). \tag{4.4}$$

Since $r_{nk} \leqq r_n$, it follows from (4.2) and the arbitrariness of δ that Lemma 11 will be proved if we show that

$$\liminf_{k \to \infty} \liminf_{n \to \infty} \inf_{\phi_n \varepsilon \mathfrak{D}_n} \sup_{F \varepsilon \mathfrak{F}'_\epsilon} r_{nk}(F, \phi_n) \geqq \lim_{n \to \infty} r_n(F_\delta, \phi_n^*). \tag{4.5}$$

2. Define

$$r_{\infty k}^* = \int_0^\infty W(r) d_r G_{\infty,k}(r; F_\delta) \tag{4.6}$$

and

$$r^* = \int_0^\infty W(r) d_r G(r; F_\delta). \tag{4.7}$$

Let $\mathfrak{F}'_{\epsilon,k}$ be the subset of $\mathfrak{F}'_\epsilon$ consisting of every absolutely continuous d.f. in $\mathfrak{F}'_\epsilon$ which has a density function which is a constant on each of the k^m open m-cubes of side $1/k$ in I^m whose corners are points of A_k^m. From equations (1.4) through (1.8) and the fact that $F_\delta \varepsilon \mathfrak{F}^*$, we have

$$\lim_{n \to \infty} r_{nk}(F_\delta, \phi_n^*) = r_{\infty k}^* \tag{4.8}$$

and

$$\lim_{n \to \infty} r_n(F_\delta, \phi_n^*) = r^* = \lim_{k \to \infty} r_{\infty k}^*. \tag{4.9}$$

Let $F_{\delta k}$ be that member of $\mathfrak{F}'_\epsilon$ for which $F_{\delta k}(z) = F_\delta(z)$ whenever $z \varepsilon A_k^m$. Clearly, for each k and n,

$$r_{nk}(F_{\delta k}, \phi_n^*) = r_{nk}(F_\delta, \phi_n^*). \tag{4.10}$$

From equations (4.6) through (4.10) and the fact that $\mathfrak{F}'_{\epsilon k} \subset \mathfrak{F}'_\epsilon$, we see that

(4.5) will be proved if we show that, for each fixed $k > 1$,

$$\liminf_{n\to\infty} \inf_{\phi_n \varepsilon \mathfrak{D}_n} \sup_{F \varepsilon \mathfrak{F}'_{\epsilon k}} r_{nk}(F, \phi_n) \geqq \lim_{n\to\infty} r_{nk}(F_{\delta k}, \phi_n^*). \tag{4.11}$$

Since a sufficient statistic for $\mathfrak{F}'_{\epsilon k}$ based on $Z^{(n)}$ is the collection $T^{(n,k)}$ of k^m real random variables which are equal to the number of components of $Z^{(n)}$ taking on values in each of the k^m cubes just described, we may replace $\mathfrak{D}_n$ in (4.11) by the class $\mathfrak{D}_{n,k}$ of decision functions depending only on $T^{(n,k)}$. But the definition of r_{nk} then shows that the left side of (4.11) may be viewed as the limiting minimax risk associated with the problem of estimating certain linear combinations of multinomial probabilities. If we put $h + 1 = k^m$ in Section 3 and think of the p_i as being assigned to the k^m cubes and think of the $L = (k + 1)^m$ quantities ρ_i as being the values of the unknown d.f. at the $(k + 1)^m$ points in A_k^m, then the left side of (4.11) without the limit in n may be identified with the minimax risk for a multinomial problem with the setup of Lemma 10. (We shall discuss B'_h and the C of (3.5) in the next paragraph.)

3. Fix $k > 1$. For any F in $\mathfrak{F}'_{\epsilon,k}$, let π_F be the associated multinomial probability vector whose components are the p_i described in the previous paragraph. Let B'_h be the set of all such π_F in $\mathfrak{F}'_{\epsilon,k}$. From the definitions of $\mathfrak{F}_\epsilon$ and $\mathfrak{F}'_\epsilon$ it is clear that B'_h is a closed convex h-dimensional subset of the h-dimensional set B_h, and thus satisfies the requirements of Lemma 10. For the ρ_i defined in the previous paragraph, we can clearly take the C and C_0 of (3.5) to be the c'_m and c_m^* of (1.8). Hence, from Lemma 10, for each k, we have for the multinomial problem of Section 3 where B'_h and the ρ_i are as described above and the function W is that given in the statement of Lemma 11,

$$\lim_{n\to\infty} \inf_{\psi_n \varepsilon \mathcal{E}_n} \sup_{\pi \varepsilon B'_h} r'(\pi, \psi_n) = \lim_{n\to\infty} \sup_{\pi \varepsilon B'_h} r'(\pi, \psi_n^*), \tag{4.12}$$

where we have written r' for the risk function in the multinomial problem. Since $r'(\pi_F, \psi_n^*) = r_{nk}(F, \phi_n^*)$ and since the left sides of (4.11) and (4.12) are equal because of the correspondence of $\mathcal{E}_n$ to $\mathfrak{D}_{n,k}$, of r' to r, and of B'_h to $\mathfrak{F}'_{\epsilon,k}$, we see that (4.11) follows from (4.12). Thus, Lemma 11 is proved.

5. Completion of the proof of Theorem 1; passage to the limit with ϵ. We now complete the proof of Theorem 1 by showing that $\mathfrak{F}_\epsilon$ is suitably dense in $\mathfrak{F}^*$ (and hence that $\mathfrak{F}'_\epsilon$ is suitably dense in $\mathfrak{F}^c$) as $\epsilon \to 0$. We require two lemmas to do this.

As in Section 2, the proof of the next two lemmas is very similar for all $m > 1$, but is most briefly written out when $m = 2$. For simplicity of presentation, we shall therefore again write out the details only in the case $m = 2$, and shall state explicitly all modifications for the case $m > 2$ which are not completely obvious.

Let $\mathfrak{F}^I$ denote the class of all d.f.'s on I^2 (in the general case, on I^m). For F in $\mathfrak{F}^I$ and $0 < \epsilon < 1$, define

$$\begin{aligned} \bar{F}(x, y) &= (1 - \epsilon)F(x, y / (1 - \epsilon)), \quad y \leqq 1 - \epsilon; \\ \bar{F}(x, y) &= (1 - \epsilon)F(x, 1) + x(y - 1 + \epsilon), \quad y > 1 - \epsilon. \end{aligned} \tag{5.1}$$

We shall not display the dependence on ϵ of the bar operation defined by (5.1). If $F \varepsilon \mathfrak{F}^*$ and we perform the bar operation of (5.1) on F to obtain $\bar{F}$ and then, interchanging the roles of x and y, perform the bar operation on $\bar{F}$ to obtain F^* (say), we clearly have $F^* \varepsilon \mathfrak{F}_\epsilon$. (In the case of chance vectors with m components, F^* is obtained after m such steps.) Let $\bar{Z}_1, \cdots, \bar{Z}_n$ be independent chance vectors with the common d.f. $\bar{F}$, let $\bar{S}_n$ be their sample (empiric) d.f., and define

$$\bar{D}_n = \sup_{z \varepsilon I} |\bar{S}_n(z) - \bar{F}(z)|.$$

Also, define $m = m(n, \epsilon)$ to be the greatest integer $\leqq n(1 - \epsilon)$.

We now prove the following lemma:

LEMMA 12. *We have*

$$P_{\bar{F}}\{\bar{D}_n < r/n^{1/2}\} \leqq P_F\{D_m < [r(1 + \epsilon) + 7\epsilon^{1/4}] / m^{1/2}\} + o(1 \mid \epsilon, N(\epsilon) \mid r, F \varepsilon \mathfrak{F}^I). \tag{5.2}$$

PROOF. Let C^* be the event

$$\{|\bar{S}_n(1, 1 - \epsilon) - (1 - \epsilon)| < \epsilon^{1/4} n^{-1/2}\}.$$

From Chebyshev's inequality we obtain

$$P_F\{C^*\} = 1 + o(1 \mid \epsilon \mid n, F \varepsilon \mathfrak{F}^I). \tag{5.3}$$

For small ϵ we have

$$\left|1 - \frac{n(1 - \epsilon)}{n(1 - \epsilon) + \theta n^{1/2}\epsilon^{1/4}}\right| < 4\epsilon^{1/4}/n^{1/2}. \tag{5.4}$$

Hence, when C^* occurs and ϵ is small,

$$\left|1 - \frac{n(1 - \epsilon)}{n\bar{S}_n(1, 1 - \epsilon)}\right| < 4\epsilon^{1/4}/n^{1/2}. \tag{5.5}$$

Since

$$\left|\frac{(1 - \epsilon)\bar{S}_n(z)}{\bar{S}_n(1, 1 - \epsilon)} - \bar{S}_n(z)\right| \leqq \left|\frac{(1 - \epsilon)}{\bar{S}_n(1, 1 - \epsilon)} - 1\right|, \tag{5.6}$$

we have

$$\bar{D}_n \geqq \sup_{\substack{0 \leqq x \leqq 1 \\ 0 \leqq y \leqq 1-\epsilon}} \left|\frac{(1 - \epsilon)\bar{S}_n(z)}{\bar{S}_n(1, 1 - \epsilon)} - \bar{F}(z)\right| - \left|1 - \frac{(1 - \epsilon)}{\bar{S}_n(1, 1 - \epsilon)}\right|. \tag{5.7}$$

Also we have, for $y \leqq 1 - \epsilon$,

$$E_{\bar{F}}\left\{\frac{n\bar{S}_n(z)}{m'} \,\middle|\, n\bar{S}_n(1, 1 - \epsilon) = m'\right\} = \bar{F}(z) / (1 - \epsilon) = F(x, y / (1 - \epsilon)). \tag{5.8}$$

Hence the conditional d.f. of the first term on the right side of (5.7), given that $n\bar{S}_n(1, 1 - \epsilon) = m'$, is the same as the d.f. of $(1 - \epsilon)D_{m'}$. In what follows define $M' = n\bar{S}_n(1, 1 - \epsilon)$.

If m_1 and m_2 are two positive integers with $m_1 < m_2$, we can think of S_{m_2} as being obtained by adjoining $(m_2 - m_1)$ random vectors Z_i to the set of m_1 random vectors Z_i which gave rise to a corresponding realization of S_{m_1}. Hence, θ' denoting a value with $0 \leqq \theta' \leqq 1$, the corresponding values of $S_{m_1}(z)$ and $S_{m_2}(z)$ differ for all z by no more than

$$|S_{m_2}(z) - S_{m_1}(z)| = \left|\frac{m_1 S_{m_1}(z) + \theta'(m_2 - m_1)}{m_2} - S_{m_1}(z)\right| \leqq \frac{(m_2 - m_1)}{m_2}.$$

Thus, in C^*, where $|M' - m| < n^{1/2}\epsilon^{1/4} + 1$, we have that for each possible value of D_m there is a corresponding set of values of $D_{M'}$ of the same probability (these sets corresponding to different values of D_m arising from disjoint sets in the space of sequences $\{Z_i\}$) with $D_m \leqq D_{M'} + 2\epsilon^{1/4} m^{-1/2}$, provided $m >$ some $M(\epsilon)$.

From (5.3), (5.5), (5.7), (5.8), and the discussion of the previous paragraph, we have

$$\begin{aligned} &P_{\bar{F}}\{\bar{D}_n < r/n^{1/2}\} \\ (5.9) \qquad &\leqq \sup_{|m'-m| < n^{1/2}\epsilon^{1/4}+1} P_F\{(1 - \epsilon)D_{m'} < (r + 4\epsilon^{1/4}) / n^{1/2}\} + P_{\bar{F}}\{C^*\} \\ &\leqq P_F\left\{D_m < \frac{r(1 + \epsilon) + 7\epsilon^{1/4}}{m^{1/2}}\right\} + o(1 \mid \epsilon, N(\epsilon) \mid r, F \varepsilon \mathfrak{F}^I), \end{aligned}$$

which proves Lemma 12.

We now prove

LEMMA 13. *For W satisfying the assumptions of Theorem 1, we have*

$$(5.10) \qquad \sup_{F \varepsilon \mathfrak{F}^*} r_n(F, \phi_n^*) = \sup_{F \varepsilon \mathfrak{F}_\epsilon} r_n(F, \phi_n^*) + o(1 \mid \epsilon, N'(\epsilon)).$$

PROOF. Define $m' = m'(n, \epsilon)$ to be the greatest integer $\leqq (1 - \epsilon)^2 n$. Using Lemma 12 a second time (with a trivial modification since $m'(n, \epsilon)$ may differ by unity from $m[m(n, \epsilon), \epsilon]$) to go from $\bar{F}$ to F^*, we have at once, for any W satisfying (1.4) and the other assumptions of Theorem 1,

$$(5.11) \qquad r_n(F^*, \phi_n^*) \geqq r_{m'}(F, \phi_{m'}^*) + o(1 \mid \epsilon, N(\epsilon) \mid F \varepsilon \mathfrak{F}^I).$$

From (5.11) and the fact that $F^* \varepsilon \mathfrak{F}_\epsilon$ if $F \varepsilon \mathfrak{F}^*$, we have

$$(5.12) \qquad \sup_{F \varepsilon \mathfrak{F}_\epsilon} r_n(F, \phi_n^*) \geqq \sup_{F \varepsilon \mathfrak{F}^*} r_{m'}(F, \phi_{m'}^*) + o(1 \mid \epsilon, N(\epsilon)).$$

Now, as in the first part of the proof of Lemma 11, we have that $r_n(F, \phi_n^*)$ approaches a bounded limit as $n \to \infty$, uniformly for F in $\mathfrak{F}_\epsilon$. Hence,

$$(5.13) \qquad \sup_{F \varepsilon \mathfrak{F}_\epsilon} r_n(F, \phi_n^*) = \sup_{F \varepsilon \mathfrak{F}_\epsilon} r_{m'}(F, \phi_{m'}^*) + o_\epsilon(1 \mid n),$$

where $o_\epsilon(1 \mid n)$ denotes a term which, for each ϵ, goes to 0 as $n \to \infty$ (not necessarily uniformly in ϵ). From (5.12), (5.13), and the fact that $\mathfrak{F}_\epsilon \subset \mathfrak{F}^*$, we obtain

$$(5.14) \qquad \sup_{F \varepsilon \mathfrak{F}^*} r_{m'}(F, \phi_{m'}^*) = \sup_{F \varepsilon \mathfrak{F}_\epsilon} r_{m'}(F, \phi_{m'}^*) + o(1 \mid \epsilon, N''(\epsilon)).$$

Since the possible values of m' for $n > N''(\epsilon)$ include all integers $> N''(\epsilon)(1 - \epsilon)^2 - 1 = N'(\epsilon)$ (say), Lemma 13 follows from (5.14).

LEMMA 14. *The statement of Theorem 1 holds with $\mathfrak{F}$ replaced by $\mathfrak{F}^c$.*

PROOF. We have previously alluded to the fact that, if Z_1 has a d.f. F in $\mathfrak{F}^c$, then by appropriate monotonic transformations on the individual coordinates of Z_1 we can obtain a random vector Z_1' (say) such that Z_1' has d.f. F' (say) in $\mathfrak{F}^*$ and $G_n(r; F') = G_n(r; F)$ for all r and n. Hence,

$$\sup_{F \varepsilon \mathfrak{F}^c} r_n(F, \phi_n^*) = \sup_{F \varepsilon \mathfrak{F}^*} r_n(F, \phi_n^*). \tag{5.15}$$

Moreover, in the same way we have

$$\sup_{F \varepsilon \mathfrak{F}_\epsilon} r_n(F, \phi_n^*) = \sup_{F \varepsilon \mathfrak{F}_\epsilon'} r_n(F, \phi_n^*). \tag{5.16}$$

Lemma 14 now follows at once from Lemma 11, (5.16), Lemma 13, (5.15), and the fact that $\mathfrak{F}_\epsilon' \subset \mathfrak{F}^c$.

PROOF OF THEOREM 1. Theorem 1 now follows immediately from Lemma 14 and (1.9).

We remark that the proof of Theorem 1 is clearly valid when $\mathfrak{F}$ is replaced by a suitably large subset.

It is not really necessary to prove Theorem 1 by using (1.9) and proving the result first for $\mathfrak{F}^c$ (in Lemma 14). For Lemma 13 clearly holds if in (5.10) we replace $\mathfrak{F}^*$ by $\mathfrak{F}^I$ and $\mathfrak{F}_\epsilon$ by the class of d.f.'s obtained by substituting $\mathfrak{F}^I$ for $\mathfrak{F}^*$ in the definition of $\mathfrak{F}_\epsilon$; one can carry through the arguments of Sections 2 and 4 with this altered definition of $\mathfrak{F}_\epsilon$ (appropriate results from [1] still hold), and obvious analogues of (5.15) and (5.16) then yield Theorem 1.

In Section 7 we shall discuss various modifications of Theorem 1 obtained by altering the way in which W depends on $F(z) - g(z)$.

6. Integral weight functions. Since for $m > 1$ the procedure ϕ_n^* does not have constant risk for F in $\mathfrak{F}^c$ and any common weight functions of the form given in equation (5.1) of [1], there is no longer any special reason for considering weight functions for which the dependence on F of the integrand is of the form considered there. Therefore, to make the proof of this section as simple as possible, we shall consider here the analogue of the special case of Section 5 of [1] wherein $W(y, z)$ does not depend on z, relegating the consideration of more complicated weight functions to Section 7. Our result is

THEOREM 2. *Let $W(r)$ be a monotonically nondecreasing nonnegative real function of r for $r \geqq 0$ which is not identically zero and which satisfies*

$$\int_0^\infty W(r) r e^{-2r^2} \, dr < \infty. \tag{6.1}$$

Then $\{\phi_n^*\}$ is asymptotically minimax relative to $\mathfrak{F}^c$ and the weight functions

$$W_n(F, g) = \int W(n^{1/2}|F(x) - g(x)|) \, dF(x). \tag{6.2}$$

PROOF. As in Section 5 of [1], the proof of this theorem is essentially easier than that of Theorem 1, since it is centered about the one-dimensional asymptotic result (6.12) (for each z). The analytic details are often like corresponding ones of Section 5 of [1], to which we shall consequently sometimes refer. The proof will be conducted in four numbered paragraphs.

1. From (6.1) and the uniformity of approach to its continuous limit of the d.f. of $n^{1/2}[S_n(z) - F(z)]$ for all z for which $\frac{1}{2} - |F(z) - \frac{1}{2}| > \delta > 0$ and all F in $\mathfrak{F}^c$ (the F-measure of this set of z approaches 1 as $\delta \to 0$, uniformly in F), we conclude at once from (6.1) that $r_n(F, \phi_n^*)$ has a bounded limit uniformly for F in $\mathfrak{F}^c$, and thus that (4.2) is satisfied with $\mathfrak{F}_\epsilon'$ replaced by $\mathfrak{F}^c$, for some F_δ in $\mathfrak{F}^c$ (of course, r_n is now to be computed using (6.2)). We can clearly suppose, and hereafter do, that F_δ is a d.f. on I^m. Let $\mathfrak{F}_{0k}'$ denote the class of d.f.'s defined in paragraph 2 of the proof of Lemma 11, with $\epsilon = 0$; thus, the B_h' of paragraph 3 of that proof now coincides with the B_h there.

2. As in Section 5 of [1], we shall let $\{\xi_{kn}\}$ be a sequence of *a priori* probability measures on B_h (we shall think of $\mathfrak{F}_{0k}'$ and B_h interchangeably), and we shall write $P_z^*\{A\}$ for the probability of an event expressed in terms of $T^{(n,k)} = T^{(n)}$ when the latter has probability function

$$(6.3)\qquad \begin{aligned} &P\{T_i^{(n)} = t_i^{(n)}, 1 \leqq i \leqq h+1\} \\ &\quad = \frac{1}{d(k, n, z)} \int_{B_h} f(z, \pi) P_\pi\{T_i^{(n)} = t_i^{(n)}, 1 \leqq i \leqq h+1\}\, d\xi_{kn}(\pi); \end{aligned}$$

here P_π is defined in (3.1) and $f(z, \pi)$ is the Lebesgue density at z (in I^m) of the d.f. $F(\cdot, \pi)$ in $\mathfrak{F}_{0k}'$ corresponding to a given π in B_h; $d(k, n, z)$ is chosen to make (6.3) a probability function. We take $f(z, \pi)$ to be constant on the interior of each of the k^m cubes in I^m; this determines (6.3) for all z with all irrational components (hereafter called irrational z), to which such z we may limit all further discussion. For each such z and possible value $t^{(n)}$ of $T^{(n,k)}$, we define

$$(6.4)\qquad r_{kn}(z, \phi, t^{(n)}) = \int_{B_h} E_\phi W(n^{1/2}|g(z) - F(z, \pi)|)\, d_\pi \xi_{kn}^*(\pi, z, t^{(n)}),$$

where, for Borel subsets B of B_h,

$$(6.5)\qquad \xi_{kn}^*(B, z, t^{(n)}) = \frac{\int_B f(z, \pi) P_\pi\{t^{(n)}\}\, d\xi_{kn}(\pi)}{\int_{B_h} f(z, \pi) P_\pi\{t^{(n)}\}\, d\xi_{kn}(\pi)};$$

we have used $P_\pi\{t^{(n)}\}$ to denote the function of (3.1).

For each n and k, if F is restricted to be in $\mathfrak{F}_{0k}'$, we may, as in Section 4, restrict our consideration to procedures ϕ in $\mathfrak{D}_{n,k}$. Denoting expectation with respect to P_z^* by E_z^*, we have as in (5.10) of [1],

$$(6.6)\qquad \int r_n(F, \phi)\, d\xi_{kn} = \int_{I^m} E_z^* r_{kn}(z, \phi, T^{(n,k)})\, d(k, n, z)\, dz,$$

where dz denotes the differential element of Lebesgue measure on I^m. For fixed n, k, z, and $t^{(n)}$, let $r_{kn}^*(z, t^{(n)})$ denote the infimum of (6.4) over $\mathfrak{D}_{n,k}$. In order to prove Theorem 2, according to (6.6) and the discussion of paragraph 1 of this proof, it clearly suffices to show that, for some $\{\xi_{kn}\}$,

$$\lim_{k\to\infty} \lim_{n\to\infty} \int_{I^m} E_z^* r_{kn}(z, \phi, T^{(n,k)})\, d(k, n, z)\, dz \geqq \lim_{n\to\infty} r_n(F_\delta, \phi_n^*). \tag{6.7}$$

3. Fix k. Let $\pi_{\delta k}$ be such that $F(z, \pi_{\delta k}) = F_\delta(z)$ for z in A_k^m. We may assume $\pi_{\delta k}$ is an interior point of B_h; for, if $\pi_{\delta k}$ were not an interior point, letting $F_\delta' = (1 - \delta')F_\delta + \delta' U$ where U is the uniform d.f. on I^m, we see easily that the right side of (6.7) can be decreased by at most a quantity which approaches 0 as $\delta' \to 0$ if F_δ is replaced by F_δ' there; we could thus replace $\pi_{\delta k}$ by the interior point π corresponding to F_δ' (for δ' small but positive) in what follows. Let ξ_{kn}, $n = 1, 2, \cdots$, be a sequence of *a priori* measures on B_h which "shrink down" on $\pi_{\delta k}$ as the ξ_n of Lemma 8 shrink down on π^*; e.g., ξ_{kn} is uniform on a sphere of radius $n^{-1/4}$ about $\pi_{\delta k}$. It follows at once that

$$\lim_{n\to\infty} d(k, n, z) = f(z, \pi_{\delta k}) \tag{6.8}$$

at all irrational z. Suppose we show that, for any irrational z and any $\epsilon > 0$, there is an $N = N(\epsilon, z, k)$ such that, for $n > N$, P_z^* assigns probability at least $1 - \epsilon$ to a set of $T^{(n,k)}$ values for which

$$r_{kn}^*(z, T^{(n,k)}) + \epsilon > \int_{-\infty}^{\infty} W(y) q(y, \sigma(z, k))\, dy, \tag{6.9}$$

where $q(y, \sigma) = (2\pi\sigma^2)^{-1/2} \exp(-y^2/2\sigma^2)$ and where $\sigma(z, k)$ is continuous in z and

$$\sigma(z, k) = F(z, \pi_{\delta k})[1 - F(z, \pi_{\delta k})] + o(1 \mid k \mid z) \leqq \tfrac{1}{4}. \tag{6.10}$$

Then, writing $V(z, k)$ for the expression on the right side of (6.9), we will clearly have (from (6.10), (6.1), and the continuity of q)

$$\begin{aligned} \lim_{n\to\infty} r_n(F_\delta, \phi_n^*) &= \int_{I^m} \lim_{k\to\infty} V(z, k)\, dF_\delta(z) \\ &= \lim_{k\to\infty} \int_{I^m} V(z, k)\, dF_\delta(z) \\ &= \lim_{k\to\infty} \int_{I^m} V(z, k) f(z, \pi_{\delta k})\, dz. \end{aligned} \tag{6.11}$$

Thus, an application of Fatou's lemma to the left side of (6.7) shows that (6.11) and (6.8) will imply (6.7). Thus, it remains to prove (6.9) for the appropriate values of the arguments there.

4. The proof of (6.9) is similar to that of Lemma 8. For fixed z, the expression of (6.5) is like the *a posteriori* probability measure of π when ξ_{kn} is the *a priori*

measure, except for the factor $f(z, \pi)$. In fact, by the shrinking property of ξ_{kn} as $n \to \infty$ and the nature of $f(z, \pi)$, one obtains in the manner of [7] (see [1] for details) that, for any $\epsilon' > 0$ and for n suitably large, with probability $>1 - \epsilon'$ under P_z^*, the joint density according to ξ_{kn}^* of the quantities $\bar{\gamma}_i = n^{1/2}(p_i - t_i^{(n)}/n)$, $1 \leqq i \leqq h$ (where we have written $\pi = (p_1, \cdots, p_{h+1})$), in a spherical region of probability $>1 - \epsilon'$ under ξ_{kn}^*, is at least $(1 - \epsilon')$ times the appropriate normal density for which the $\bar{\gamma}_i$ have means 0, var $\bar{\gamma}_i = p_{\delta i}(1 - p_{\delta i})$, $\text{cov}(\bar{\gamma}_i, \bar{\gamma}_j) = -p_{\delta i}p_{\delta j}$ (the $p_{\delta i}$ being the components of $\pi_{\delta k}$). For $\epsilon'' > 0$, an elementary computation (the details being like those of [1], p. 661, except that now $m > 1$) then shows that, for a fixed arbitrary irrational z, the corresponding distribution of $n^{1/2}[F(z, \pi) - J_{n,k}(z)]$, where $J_{n,k}(z)$ is the obvious best linear estimator in $\mathfrak{D}_{n,k}$ of $F(z, \pi)$ for π in $\mathfrak{F}_{0k}'$ (not in general $S_n(z)$, unless $z \in A_k^m$), has, with probability $>1 - \epsilon''$ under P_z^*, an absolutely continuous component the magnitude of whose Lebesgue density is at least

$$(1 - \epsilon'')\, q(y, \sigma(z, k)) \tag{6.12}$$

on the interval $-1/\epsilon'' < y < 1/\epsilon''$, where $\sigma(z, k)$ is continuous in z and satisfies (6.10). Since ϵ'' is arbitrary, (6.9) follows easily from (6.12) and the trivial one-dimensional case of [8] (see [1] for details; the argument here is easier, since we have not yet included the additional dependence of W on other quantities as in [1] and Section 7 below). Thus, Theorem 2 is proved.

It is clear that Theorem 2 remains valid if $\mathfrak{F}^c$ is replaced by a suitably large subset. Further generalizations will be discussed in the next section.

7. Other loss functions. We list a few of the extensions of Theorems 1 and 2 which may be proved by the same methods with only minor modifications and no essential new difficulties in the proof. In fact, our treatment of the case $m > 1$ (compared with the argument of [1]) has been concentrated on the difficulty engendered by the nonconstancy of $r_n(F, \phi_n^*)$, and that nonconstancy (in the counterpart of modification F, below) is the only real new difficulty in any of the corresponding generalizations of Section 6 of [1] (the difficulty is more trivial there, where $m = 1$ and the nonconstancy is easier to deal with than in Theorems 1 and 2 above).

A. In Theorem 2, the form of W may be extended. For $m = 1$, the more general form $W(n^{1/2}|F(z) - g(z)|, F(z))$ was considered in Section 5 of [1]. The same form can be considered here, but perhaps the dependence on the second variable is no longer so natural; it may be replaced or supplemented, for example, by a dependence on the value of the marginal d.f.'s at the point z. The regularity condition which must be imposed on W in order for our method of proof to hold is, in any event, exactly the obvious analogue of that of Section 5 of [1]. For example, continuity and an appropriate integrability condition (the analogue of (5.5) of [1]) is more than enough.

B. In Theorem 2, W can be replaced by a measure (rather than a density) in the second argument of the W of A above (or its replacements, just above).

For example, when $m = 2$, one might be interested only in the estimation of the deciles of the marginal d.f.'s F_1 and F_2 (say) and, at each decile r of F_1, the deciles of the d.f. $F(r, y) / F_1(r)$ (and its counterpart with x and y interchanged).

C. An analogue of Theorem 2 (with any of the modifications noted above) for $\mathfrak{F}$ rather than $\mathfrak{F}^c$ is perhaps not too natural (see [1] for further comments), but can be given under suitable assumptions. An analogue of Theorem 1 or Theorem 2 for the class of purely discrete d.f.'s (e.g., on R^m, or on the integral lattice points of R^m) can also be given; for example, the former essentially follows from the fact that there is a discrete d.f. at which ϕ_n^* has almost the same risk as at F_δ when n is large (see (1.4) through (1.8)).

D. In Theorem 1, one can replace D_n by $\sup_z[|g(z) - F(z)|h(F(z))]$, where h is a suitably regular nonnegative function whose dependence on $F(z)$ may be replaced, e.g., by a dependence on the marginal d.f.'s, as in A above; a linear combination of such functions can also be employed. If h takes on only the values 0 and 1, this modification amounts to taking the supremum of the deviation over a suitable subset of R^m whose description depends on F.

E. In Theorem 1, one could consider the *measures* P, Q_n, and g^* corresponding to F, S_n, and g, and could let W depend on $\sup_A|P(A) - g^*(A)|$ where the supremum is taken over a suitable family of sets, e.g., rectangles with sides parallel to the coordinate axes. This presents no new difficulties.

F. The function h of D above, the second argument of W in A above, and the integrating measure of (6.2), can all be changed so as to depend only on z and not on $F(z)$ (or they can depend on both). This requires no new arguments, only obvious regularity conditions as on p. 664 of [1]. It is again the existence of an F_δ which is the crucial point.

G. The remarks on the *sequential* asymptotic minimax character of ϕ_n^* for suitable weight functions, which are contained on pp. 664–665 of [1], hold here without change.

H. Obvious combinations of the types of dependence of W_n on F and z which occur in Theorems 1 and 2 and in the previous remarks can be considered with no essential new difficulty. In fact, the asymptotic minimax character of ϕ_n^* seems to hold for a very general class of weight functions. The discussion of p. 664 of [1] indicates the possible breadth of that class, but we are even further than we were in the case $m = 1$ of [1] from being able to give a single simple, unified proof.

REFERENCES

[1] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 642–669.

[2] J. Kiefer, and J. Wolfowitz, "On the deviations of the empiric distribution function of vector chance variables," *Trans. Amer. Math. Soc.*, Vol. 87, Jan. 1958, pp. 173–186.

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

[4] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–402.

[5] C. G. Esseen, "Fourier analysis of distribution functions," *Acta. Math.*, Vol. 77 (1945), pp. 1–125.
[6] H. Bergstrom, "On the central limit theorem in the space R^k, $k > 1$," *Skand. Aktuarietids*, (1949), pp. 106–127.
[7] R. v. Mises, "Fundamentalsätze der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol. 4 (1919), pp. 1–97.
[8] T. W. Anderson, "The integral of a symmetric unimodal function," *Proc. Amer. Math. Soc.*, Vol. 6 (1955), pp. 170–176.