DISTRIBUTION FREE TESTS OF INDEPENDENCE BASED ON THE SAMPLE DISTRIBUTION FUNCTION

BY J. R. BLUM,[1] J. KIEFER,[2] AND M. ROSENBLATT[3]

Sandia Corporation and Indiana University; Cornell University; and Brown University

0. Summary. Certain tests of independence based on the sample distribution function (d.f.) possess power properties superior to those of other tests of independence previously discussed in the literature. The characteristic functions of the limiting d.f.'s of a class of such test criteria are obtained, and the corresponding d.f. is tabled in the bivariate case, where the test is equivalent to one originally proposed by Hoeffding [4]. A discussion is included of the computational problems which arise in the inversion of characteristic functions of this type. Techniques for computing the statistics and for approximating the tail probabilities are considered.

1. Introduction. The idea of using various simple functionals of the sample d.f. of vector chance variables in order to test the independence of components, is a natural one. Only the difficult distribution theory prevents the use of such tests and the resulting achievement of improvement in power performance over all currently used tests. Specifically, let Ω be the class of continuous d.f.'s on m-dimensional Euclidean space R^m, and let ω be the subclass consisting of every member of Ω which is a product of its associated one-dimensional marginal d.f.'s. Let $X_1, \cdots, X_n$ be independent random m-vectors with common unknown d.f. F, a member of Ω, and suppose that it is desired to test the hypothesis $H_0: F \varepsilon \omega$ against the alternative $H_1: F \varepsilon \Omega - \omega$. Let S_n be the sample d.f. of $X_1, \cdots, X_n$; i.e., for x in R^m, $S_n(x)$ is n^{-1} times the number of X_i all of whose components are less than or equal to the corresponding components of x, i.e.,

$$S_n(r_1, r_2 \cdots, r_m) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{m} \phi_{r_i}(X_j^{(i)}),$$

where $X_j = (X_j^{(1)}, \cdots, X_j^{(m)})$ and

$$\phi_r(x) = \begin{cases} 1 & \text{if} \quad x \leqq r, \\ 0 & \text{if} \quad x > r. \end{cases}$$

Write S_{nj} for the marginal d.f. associated with the jth component of S_n (i.e., for the sample d.f. of the jth component of the X_i), and let

$$T_n(r) = S_n(r) - \prod_{j=1}^{m} S_{nj}(r_j). \tag{1.1}$$

Received April 24, 1960; revised September 15, 1960.

[1] Research sponsored by the Office of Ordnance Research, U. S. Army, under Contract No. DA-33-008-ORD-965.

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

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

Then many tests based on T_n will have good power properties (see Section 4) and will be similar on ω. For example, the critical region based on large values of

$$A_n = \sup |T_n(r)|,$$

a statistic constructed in the spirit of the Kolmogorov-Smirnov statistics, evidently has such properties. It follows from the results of [8] that the d.f. of $n^{\frac{1}{2}}A_n$ under H_0 differs from unity by less than $c_1 \exp(-c_2 z^2)$ for all n and all arguments $z > 0$, where the c_i are positive constants. It can be shown that the limiting d.f. of $n^{\frac{1}{2}}A_n$ exists (and hence has the same behavior with z); since the proof is somewhat long but uses mainly ideas like those of [8], it will not be given here. The calculation of this asymptotic distribution seems formidable; it is equivalent to the computation of the d.f. of the maximum of a particular Gaussian process with multidimensional *time* parameter. A corresponding calculation of exact (nonasymptotic) distributions for various values of n can, of course, be achieved numerically, but such calculations are extremely laborious even if done by machines for rather small n.

Another critical region, constructed in the spirit of the von Mises-Cramér tests, is that based on large values of

$$B_n = \int [T_n(r)]^2 \, dS_n(r). \tag{1.2}$$

Adapting the well known technique of Kac and Siegert [5] to the present setting (such a multidimensional computation was first carried out in [12]), we shall obtain the characteristic function of the asymptotic distribution of nB_n under H_0 when $m = 2$ (Section 2), in which case the test turns out to be equivalent to one constructed on other heuristic grounds by Hoeffding [4] (see Section 5 below for the form in which Hoeffding stated his test). Certain variants of nB_n in the case $m > 2$ will be considered in Section 3.

In Section 4 questions of distribution under H_1, power, and estimation, and certain modifications, will be taken up. A particularly simple and computationally convenient form of the tests is given in Section 5. In Section 6 an approximation is suggested to the tail of the limiting distribution, which is compared with the exact results; this idea clearly has useful applications in many other problems. Methods for computing distributions of weighted sums of chi-square variables, which are relevant for computing the asymptotic distribution of nB_n as well as many other important distributions in statistics, are discussed in Section 7. The asymptotic distribution of nB_n for the case $m = 2$ is tabulated in Section 8.

2. The case $m = 2$. The statistic B_n is clearly distribution-free for F in ω. As usual, we can therefore carry out our computations when F is the uniform distribution on the unit square I^2. Let $T(x, y)$ be a separable Gaussian process depending on the "time" parameter (x, y) for (x, y) in I^2, and with

$$\begin{aligned} ET(x, y) &= 0, \\ ET(x, y)T(u, v) &= [\min(x, u) - xu]\,[\min(y, v) - yv]. \end{aligned} \tag{2.1}$$

A routine computation (most easily accomplished by writing

$$S_{n1}(x)S_{n2}(y) = xS_{n2}(y) + yS_{n1}(x) - xy + O_p(n^{-1})$$

shows that (2.1) gives the mean and the asymptotic covariance of the random function $n^{\frac{1}{2}}T_n$. It follows from the appropriate analogue in the present case of the corrected argument of [12] or of the argument of Section 2 of [7] (the proof being very similar here) that the asymptotic distribution of nB_n is the same as that of

$$B = \int_0^1 \int_0^1 T^2(x, y)\, dxdy.$$

Writing $s = (x, y)$, $t = (u, v)$, and $K(s, t)$ for the last member of (2.1), we consider the integral equation

$$\int_{I^2} K(s, t)\phi(t)\, dt = \lambda\phi(t). \tag{2.2}$$

It is easily seen that the eigenvalues and (complete set of) eigenfunctions of (2.2) are $1/\pi^4 j^2 k^2$ and $2 (\sin \pi jx) (\sin \pi ky)$; $j, k = 1, 2, \cdots$. Hence, exactly as in [5] and [12], we conclude that

$$Ee^{izB} = \prod_{j,k=1}^{\infty} (1 - 2iz/\pi^4 j^2 k^2)^{-\frac{1}{2}}. \tag{2.3}$$

An equivalent result was first stated by Hoeffding [4], who stated two other different methods for obtaining (2.3). The corresponding d.f. of B is tabled in Section 8.

It is obvious that, because of the factorizability of $K(s, t)$ we can similarly obtain the characteristic function of the limiting d.f. for the case where a weight function of the form $W(S_{n1}(r))W(S_{n2}(r))$ is inserted in the integrand in the expression for B_n; one has merely to use the corresponding one-dimensional results on weighted ω^2 statistics (see, e.g., [1], [5]) to obtain the eigenvalues.

3. The case $m > 2$. For the sake of brevity we shall discuss in detail only the case $m = 3$; the corresponding results for other cases require only obvious changes.

Suppose, then, that F is the uniform distribution on the unit cube. Another routine computation (most easily accomplished in a manner analogous to that suggested in Section 2) yields

$$\begin{aligned} \lim_{n\to\infty} nET_n(x, y, z)T_n(y, v, w) &= \min(x, u) \min(y, v) \min(z, w) \\ &\quad - yzvw \min(x, u) - xzuw \min(y, v) - xyuv \min(z, w) + 2xyzuvw. \end{aligned} \tag{3.1}$$

This kernel does not permit the simple treatment which that of (2.1) did, and the eigenvalues are at present unknown. This suggests that we look for a function T_n' of S_n for which

$$\begin{aligned} &\lim_{n\to\infty} nET_n'(x, y, z)T_n'(u, v, w) \\ &\qquad = [\min(x, u) - xu] [\min(y, v) - yv] [\min(z, w) - zw]. \end{aligned} \tag{3.2}$$

Denoting by S_{njk} the 2-dimensional marginal d.f. of S_n corresponding to the jth and kth coordinates (sample d.f. of the jth and kth components of the X_i), we easily verify that the function T_n' defined by

$$T_n'(x, y, z) = S_n(x, y, z) - S_{n1}(x)S_{n23}(y, z) - S_{n2}(y)S_{n13}(x, z) - S_{n3}(z)S_{n12}(x, y) + 2S_{n1}(x)S_{n2}(y)S_{n3}(z) \tag{3.3}$$

does in fact satisfy (3.2). It follows, in the manner of Section 2, that if

$$B_n' = \int [T_n'(r)]^2 \, dS_n(r),$$

then for F in ω we have

$$\lim_{n\to\infty} Ee^{iznB_n'} = \prod_{j_1, j_2, j_3=1}^{\infty} (1 - 2iz/\pi^6 j_1^2 j_2^2 j_3^2)^{-\frac{1}{2}}.$$

Thus, the asymptotic distribution of nB_n' can be tabulated in the manner of the tabulation of Section 8. However, a test for independence based only on the statistic B_n' is not to be recommended, since the power of any such test will be small for many alternatives which are far from ω; for example, it is clear that $ET_n'(r) = 0$ if F is of the form $F(x, y, z) = F_1(x)F_{23}(y, z)$. A solution to this difficulty can be found in the fact that, *if* the components of the X_i are *pair*wise independent, then $ET_n'(r) = 0$ for all r if and only if $F \varepsilon \omega$. Thus, the three 2-dimensional sample d.f.'s of the components of the X_i can be used to detect departure from pairwise independence, while B_n' detects other possible departures from independence. There are obviously many ways in which these two effects can be combined in constructing a test, and only one of them will be made explicit here. Let $T_{njk}(p, q) = S_{njk}(p, q) - S_{nj}(p)S_{nk}(q)$, and let

$$B_{njk} = \int [T_{njk}(r)]^2 \, dS_{njk}(r).$$

A computation of covariances readily shows that the functions $n^{\frac{1}{2}}T_{n12}$, $n^{\frac{1}{2}}T_{n13}$, $n^{\frac{1}{2}}T_{n23}$, and $n^{\frac{1}{2}}T_n'$ are asymptotically independent. Thus, arguing in the same manner as before, we conclude that the statistic C_n, defined by

$$C_n = n(B_{n12} + B_{n13} + B_{n23} + bB_n'), \tag{3.4}$$

where b is a positive constant, has the asymptotic distribution with characteristic function

$$\lim_{n\to\infty} Ee^{izC_n} = \prod_{j,k} (1 - 2iz/\pi^4 j^2 k^2)^{-3/2} \prod_{j_1, j_2, j_3} (1 - 2biz/\pi^6 j_1^2 j_2^2 j_3^2)^{-1/2}. \tag{3.5}$$

The corresponding asymptotic distribution can be tabulated in the manner of Section 8. The power properties of a critical region consisting of large values of C_n can be obtained as in Section 4.

4. Asymptotic distribution under H_1; power; estimation; modifications. We consider the case $m = 2$ throughout this section; the analogous results obviously all hold when $m > 2$.

If $F(x, y)$ is *not* of the form $G(x)H(y)$, where G and H are the two continuous marginal d.f.'s of X_1, the limiting d.f. of $n^{\frac{1}{2}}B_n$ can be obtained by noting that $n^{\frac{1}{2}}B_n$ is asymptotically

$$n^{\frac{1}{2}} \iint [EU_n(x, y)]^2 \, d[S_n(x, y) - F(x, y)] + n^{\frac{1}{2}} \iint [U_n(x, y) - EU_n(x, y)] \; dF(x, y) + o_p(1),$$

where

$$U_n(x, y) = S_n(x, y) - G(x)S_{n2}(y) - H(y)S_{n1}(x) + G(x)H(y).$$

Writing

$$\Delta(x, y) = F(x, y) - G(x)H(y),$$

$$\epsilon(x, y) = \iint [\phi_u(x) - G(u)][\phi_v(y) - H(v)]\Delta(u, v) \, dF(u, v),$$

we obtain that $n^{\frac{1}{2}}[B_n - \iint \Delta^2(x, y) \, dF(x, y)]$ *is asymptotically normal with mean* 0 *and the same variance as the random variable* $\Delta^2(X, Y) + 2\epsilon(X, Y)$, *where* (X, Y) *is distributed according to* F. An equivalent form of this result was given by Hoeffding [4].

Of greater interest for most applications is the limiting d.f. of nB_n when we consider a sequence $F^{(n)}$ of alternatives on I^2 for which

$$n^{\frac{1}{2}}[F^{(n)}(x, y) - G^{(n)}(x)H^{(n)}(y)] \to q(x, y)$$

(finite and continuous) as $n \to \infty$. We obtain, using arguments similar to those of Section 2, that the limiting d.f. of $n\mathrm{B}_n$ is the same as the d.f. of

$$B' = \iint [T(x, y) + q(x, y)]^2 \, dxdy.$$

Recalling the eigenvalues and eigenfunctions of K obtained in Section 2, we can write

$$T(x, y) = \sum_{j,k=1}^{\infty} 2\pi^{-2}j^{-1}k^{-1} (\sin \pi xj) \, (\sin \pi ky)X_{jk},$$

where the X_{jk} are independent normal variates with means 0 and variances 1. Hence, writing

$$q_{jk} = \iint 2q(x, y)(\sin \pi jx)(\sin \pi ky) \, dxdy,$$

we obtain for the limiting characteristic function of nB_n,

$$Ee^{B'it} = \left\{\prod_{j,k}\left(1 - \frac{2it}{\pi^4 j^2 k^2}\right)^{-\frac{1}{2}}\right\} \cdot \exp\left\{-\tfrac{1}{2}\sum_{j,k} q_{jk}^2\, \pi^4 j^2 k^2 + \tfrac{1}{2}\sum_{j,k} q_{jk}^2 \pi^4 j^2 k^2 \left(1 - \frac{2it}{\pi^4 j^2 k^2}\right)^{-1}\right\}.$$

For simple $q(x, y)$'s (e.g., where all but a finite number of the q_{jk} are zero), one could easily compute tables of the power, in the manner of Section 8. Even for general $q(x, y)$, an argument like that of Section 6 would yield information.

Without obtaining such quantitative results, we can easily give a lower bound on the power. The power properties of tests based on the sample d.f. have been discussed in detail in [6] and [7], and it will suffice to state briefly the analogous results for the problems treated in the present paper. Such results will clearly apply for arbitrary m, and for the sake of clarity and brevity we shall only state them for the case $m = 2$, the extensions to $m > 2$ being obvious.

Let F be a d.f. on R^2 and let F_1 and F_2 be the corresponding marginal d.f.'s. Write

$$\delta_F = \sup_{x,y} |F(x, y) - F_1(x)F_2(y)|$$

and

$$\gamma_F = \left\{\int_{R^2} [F(x, y) - F_1(x)F_2(y)]^2\, dF_1(x)\, dF_2(y)\right\}^{\frac{1}{2}}.$$

(A similar treatment applies if the integrating measure is replaced by F in the definition of γ_F.) Then, for $0 < \alpha, \beta < 1$, there is a constant $C(\alpha, \beta)$ such that, for each $d > 0$, there is a critical region based on large values of A_n with $n < C(\alpha, \beta)\, d^{-2}$ and which has size $\leqq \alpha$ on ω and power $\geqq \beta$ for all alternatives F for which $\delta_F \geqq d$. Thus, the behavior of the required sample size as a function of d is of the same order as in common parametric (e.g., Gaussian) examples. The same conclusion for B_n holds if δ_F is replaced by γ_F in the above.

It is clear that this guaranteed behavior of the power function against *all* alternatives is far superior to that of the other nonparametric tests previously described in the literature (outside of [4]). Many of the latter have zero efficiency compared with tests based on A_n or B_n. Perhaps the best of these classical tests is the chi-square test with the observations divided into the k_n^2 classes determined by $k_n - 1$ equally spaced values of $S_{n1}(x)$ and $S_{n2}(y)$. The optimum choice of k_n has not been investigated, but it is reasonable to suppose that the power function for the optimum choice will behave no better, and possibly worse, than that of the best chi-square test of goodness of fit (see [10], [6]). If this is so, we would conclude that, if N observations are required by the test based on A_n (resp., B_n) to achieve a goal in terms of δ_F (resp., γ_F) like that described in the previous paragraph, then at least $\bar{C}(\alpha, \beta)N^{5/4}$ observations are required by the best chi-square test.

We remark that the relationship between δ_F and γ_F is easily seen to be $\delta_F \geqq \gamma_F \geqq C\delta_F^2$, where $C > 0$.

In many applications it is desirable not merely to test for dependence, but rather to estimate the type of dependence. There are many possible formulations of this problem. If it is desired to estimate the entire function $F - F_1F_2$, then, for almost any reasonable weight function, a modification of the arguments of [9] shows that $S_n - S_{n1}S_{n2}$ is asymptotically a minimax estimator (as $n \to \infty$). Similar results hold for the problem of estimating various functionals of F, F_1, F_2.

These results on power and estimation also apply under such obvious modifications as that of considering the probabilities and empiric frequencies in all rectangles instead of only in third quadrants, of inserting a weight function in the definition of A_n and B_n, etc. Also, as in [6], [7], [8], the results on size and minimum power are not materially affected if discontinuous distributions are admitted. We note also that, just as in [7], the results are unaffected if the integrating measure S_n is replaced by $S_{n1}S_{n2} \cdots S_{nm}$ in the definition of B_n (many other functions could be used, too); in fact, the limiting d.f. is exactly the same with this modification.

5. Computation of the statistics. The statistic B_n (or one of its variants, such as those mentioned at the end of Section 4) is rather unwieldy for practical computations in its form (1.2), even if the integral is rewritten as a sum to take account of the atomicity of the integrating measure. The form originally suggested by Hoeffding for his statistic (which differs slightly from B_n) for $n \geqq 5$ was

$$D_n = \frac{1}{4n(n-1)(n-2)(n-3)(n-4)} \tag{5.1}$$
$$\cdot \sum'' \prod_{j=1}^{2} [\phi_{X_{i_1}^{(j)}}(X_{i_2}^{(j)}) - \phi_{X_{i_1}^{(j)}}(X_{i_3}^{(j)})][\phi_{X_{i_1}^{(j)}}(X_{i_4}^{(j)}) - \phi_{X_{i_1}^{(j)}}(X_{i_5}^{(j)})],$$

where ϕ is defined as in Section 1 and $\sum''$ denotes the sum over all 5-tuples $(i_1, \cdots, i_5)$ of different integers, $1 \leqq i_q \leqq n$. Another form of D_n, for use in computations, was given by Hoeffding in Section 5 of his paper.

A more convenient form than (1.2) for computational purposes is obtained by noting that, when $m = 2$,

$$n^2T_n(X_j^{(1)}, X_j^{(2)}) = N_1(j)N_4(j) - N_2(j)N_3(j),$$

where $N_1(j)$, $N_2(j)$, $N_3(j)$, $N_4(j)$ are the numbers of points lying, respectively, in the regions $\{(x, y) \mid x \leqq X_j, y \leqq Y_j\}$, $\{(x, y) \mid x > X_j, y \leqq Y_j\}$, $\{(x, y) \mid x \leqq X_j, y > Y_j\}$, $\{(x, y) \mid x > X_j, y > Y_j\}$. Thus, we have only to count the number of points lying in each of the four regions determined by the vertical and horizontal lines through $X_j = (X_j^{(1)}, X_j^{(2)})$, and compute

$$B_n = n^{-4} \sum_{j=1}^{n} [N_1(j)N_4(j) - N_2(j)N_3(j)]^2. \tag{5.2}$$

Similarly, when $m > 2$ a statistic such as that of (3.4) can easily be written in terms of the numbers of points in each of the 2^m orthants determined by the

m hyperplanes through $X_j = (X_j^{(1)}, \cdots, X_j^{(m)})$ and parallel to the coordinate hyperplanes. Thus, for $m = 3$ the statistic C_n can be written in terms of quantities $N_1(j), \cdots, N_8(j)$. We omit the details.

6. Approximation to the tail probabilities of the limiting distribution. We again limit ourselves to the case $m = 2$, although the discussion which follows even has obvious applications to problems outside this paper.

The Laplace transform of the asymptotic distribution of the test statistic nB_n under the null hypothesis is

$$\prod_{j,k=1}^{\infty}\left[1 + \frac{2t}{\pi^4 j^2 k^2}\right]^{-\frac{1}{2}}. \tag{6.1}$$

The singularity of this expression in the complex t-plane which has largest real part is located at $t = -(\pi^4/2)$. In the neighborhood of $t = -(\pi^4/2)$ the expression (6.1) has the same behavior as

$$\left(1 + \frac{2t}{\pi^4}\right)^{-\frac{1}{2}} \prod_{(j,k)\neq(1,1)} \left[1 - \frac{1}{j^2k^2}\right]^{-\frac{1}{2}}. \tag{6.2}$$

Making use of the relation $[(\sin z)/z] = \prod_{n=1}^{\infty}\{1 - [z^2/(\pi^2 n^2)]\}$ we see that

$$\prod_{(j,k)\neq(1,1)} \left[1 - \frac{1}{j^2k^2}\right]^{-\frac{1}{2}} = \sqrt{2}\prod_{n=2}^{\infty}\left(\frac{\pi/n}{\sin(\pi/n)}\right)^{\frac{1}{2}}.$$

We have been unable to invert (6.1) directly. However, some of the Tauberian theorems for Laplace transforms (see e.g., [2], p. 269) suggest that if we invert (6.2) we should approximate the tail of our distribution reasonably well. Thus we are led to approximate our distribution in the tail by

$$2^{\frac{1}{2}}\prod_{n=2}^{\infty}\left(\frac{\pi/n}{\sin(\pi/n)}\right)^{\frac{1}{2}} P\left\{\frac{X^2}{\pi^4} \geq t\right\}, \tag{6.3}$$

where X is a normal random variable with mean zero and unit variance.

The tabulation which follows gives the exact value of $1 - F(y)$ and the corresponding tail approximation, where F is the limiting distribution of $\pi^4 nB_n/2$, tabled in Section 8:

y	1 − F(y)	Tail approximation
2	.145	.115
3	.0414	.0361
4	.0130	.0118
5	.00424	.00395
6	.00142	.00134
7	.00048	.00046

Thus, the agreement is quite good for even moderate values of the size. A similar approximate computation for the asymptotic distribution of the von Mises statistic tabled in [1] also gave good agreement.

7. Some remarks on computations. The computation of the d.f. of a weighted sum of independent chi-square variables, such as that whose characteristic function is given by (2.3) or by (3.5), arises too frequently to require further mention of examples. Unfortunately, the computational techniques now available in the literature for such problems are often extremely poor in applications. While the authors have no panacea to suggest, it does seem appropriate to make a few remarks whose content has proved helpful in considering the computations of the present and other papers (e.g., [6], [7]).

A. *Useful inequalities for estimating truncation error.* In inverting expressions like (2.3), it is usually convenient to work with a finite product, and it is therefore necessary to have a bound on the error introduced by truncating the infinite product. To this end, we consider the random variable

$$Z = \sum_{k=1}^{\infty} c_k Y_k, \tag{7.1}$$

where $c_i > 0$ and the Y_i are independent chi-square variables with one degree of freedom (it will be obvious that the case where Y_i has n_i degrees of freedom can be reduced to this case). We seek an upper bound on the quantity

$$p = P\{Z > \epsilon\}, \tag{7.2}$$

where $\epsilon > 0$. The usual Chebyshev inequality is not very good here, and any of several modifications yields great improvement. The details of one such modification will now be given. We have, for $0 < T < (2 \max_k c_k)^{-1}$,

$$\begin{aligned} p &= P\{e^{TZ} > e^{\epsilon T}\} < e^{-\epsilon T} E e^{TZ} \\ &= \exp\{-\epsilon T - \tfrac{1}{2} \sum_k \log(1 - 2Tc_k)\}. \end{aligned} \tag{7.3}$$

Thus, for given c_k and ϵ, the best bound of this type is achieved by minimizing the expression in braces with respect to T. It is easier to obtain an *explicit* bound by first invoking an inequality such as

$$-\log(1 - 2Tc_k) \leqq -(c_k/c^*) \log(1 - 2Tc^*), \tag{7.4}$$

where $c^* = \max_k c_k$. Substituting the expression on the right side of (7.4) into the last expression of (7.3) and then minimizing with respect to T, and writing $S_j = \sum_k c_k^j$ and $\epsilon = S_1(1 + \delta)$, we obtain

$$P\{Z > (1 + \delta)S_1\} < \exp\{-(S_1/2c^*)[\delta - \log(1 + \delta)]\}. \tag{7.5}$$

This can be improved by using a sharper inequality in place of (7.4). For example, the substitution

$$-\log(1 - 2Tc_k) \leqq Tc_k - (c_k^2/c^{*2})[\log(1 - 2Tc^*) + Tc^*] \tag{7.6}$$

yields, in place of (7.5), the better bound

$$P\{Z > (1 + \delta)S_1\} < \exp\left\{-\frac{S_2}{2c^*}\left[\delta - \frac{S_1}{c^*S_1}\log\left(1 + \frac{c^*S_1\,\delta}{S_2}\right)\right]\right\}. \tag{7.7}$$

Further improvements can be made similarly. Of course, the usual Chebyshev inequality is

$$P\{Z > (1 + \delta)S_1\} \leqq 1/(1 + \delta). \tag{7.8}$$

As an example, suppose we want to truncate the product in (2.3) by considering only terms for which $jk \leqq 10$. To estimate the error involved in doing this, we seek an upper bound on p where the set $\{c_k\}$ consists of the λ_{jk} of Section 2 for which $jk > 10$. Routine computations yield $e^{-11\delta^2}$ and $e^{-43\delta^2}$ for the bounds of (7.5) and (7.7), respectively, when δ is small. In any event, we see that ϵ in (7.2) must be between S_1 and $2S_1$ for c_k of this sort, in order to make p fairly small. Since $S_1 = .0043$ and since $EB = .027$ (where B is as in (2.3)), we can only conclude that an approximate computation of the d.f. of B obtained by this truncation, at a value x of the argument, may actually yield the true value of the d.f. at a value as far away as $x + .2EB$, and this would probably be unsatisfactory. A larger truncation value is thus indicated. If the value 10 determining this truncation is increased to L, S_1 varies approximately inversely with L.

Since the ratio of S_1 to EB is the critical factor in determining the adequacy of a truncation in computations like that just mentioned, and since S_1 often decreases very slowly with increasing truncation value in such examples, a large number of terms in the product (2.3) will have to be used for even fair accuracy. An improvement would probably result from substituting for the ignored terms a multiple of a chi-square variable with appropriate low moments, but it seems difficult to *guarantee* an appreciable improvement in accuracy in this way. We shall return to these considerations in Section 7C.

B. Some methods of expansion and inversion. One of the most commonly used techniques for inverting characteristic functions of the form

$$\prod_{j=1}^{k} (1 - a_j it)^{-\frac{1}{2}m_j} \tag{7.9}$$

where the m_j are positive integers and the a_j are positive, is that of Pitman and Robbins [11]. Although this technique and variants of it which represent the solution in slightly different form are sometimes useful, these methods suffer from three defects in many problems: (1) the solution is given in the form of an infinite series which converges rather slowly; (2) the terms of the series are quantities such as incomplete gamma functions, which may not be convenient for some machine computations; (3) the methods do not distinguish simple cases for which a simple inversion in finite terms is possible. For a trivial example of (3), we note that, if $k = 2$, $m_1 = m_2 = 2$, $a_1 = 1$, and $a_2 = 2$, the distribution in question is immediately found by a routine convolution of two exponentials to be $2(e^{-x} - e^{-2x})$, whereas the method of [11] expresses the result as the sum of an infinite series of incomplete gamma functions.

This suggests that it will often be efficient to factor out of the expression (7.9) the corresponding expression wherein each $\frac{1}{2}m_j$ is replaced by its integral part n_j (say), to expand $t^{-1} \prod (1 - a_j it)^{-n_j}$ into partial fractions (the extra factor t^{-1} being introduced so as to give the Fourier transform of the d.f. rather than of the density), and then to invert term by term. Thus, for example, in inverting the expression discussed in the paragraph following (7.8), we can factor out and invert such an expression, leaving only the factors corresponding to λ_{11}, λ_{22}, and λ_{33}; the d.f. corresponding to these terms must then be found by other means and can then be convolved with the d.f. corresponding to the other terms. It should also be noted that the partial fraction technique will often be easy to apply in cases where (7.9) is replaced by an infinite product. For example, the expression $t^{-1}\prod_{j=1}^{\infty}(1 + 2t/\pi^2j^2)^{-1}$, which is the Laplace transform of the d.f. "B_2" which was computed by other means in Section 4 of [7], can easily be rewritten as $t^{-1} + \sum_{j=1}^{\infty}(-1)^j 4/\pi^2j^2(1 + 2t/\pi^2j^2)$, which we can invert at once to give, for $z > 0$,

$$B_2(z) = 1 + 2\sum_{j=1}^{\infty}(-1)^j e^{-\pi^2 j^2 z/2}.$$

(Incidentally, this proves the following interesting relationship: if W_1 and W_2 are independent and each is distributed according to the limiting $n\omega_n^2$ distribution, then $\pi(W_1 + W_2)^{\frac{1}{2}}/2$ is distributed according to the limiting Kolmogorov-Smirnov distribution.)

We must still discuss the inversion of general expressions like (7.9) or, with the aid of a factorization like that just discussed, of expressions like (7.9) with all $m_j = 1$. There are many possible expansions akin to that of [11], and for the sake of brevity we shall illustrate only a few such possibilities in the simple case of (7.9) where $k = 2$, $a_1 = 1$, $a_2 = c^2$ with $0 < c < 1$, and $m_1 = m_2 = 1$. Writing t for $-it$ in (7.9) (i.e., working with the Laplace transform), this expression becomes $q(t) = (1 + t)^{-\frac{1}{2}}(1 + c^2t)^{-\frac{1}{2}}$. Factoring out $(1 + ct)^{-1}$, $(1 + c^2t)^{-1}$, or $[1 + (1 + c^2)t/2]^{-1}$, respectively, and then using the binomial expansion on the remaining factor, we obtain the three expressions for $q(t)$,

$$\text{(7.10)}\quad \begin{aligned} &\text{(a)}\quad q(t) = \frac{1}{1 + ct}\sum_{j=0}^{\infty}(-1)^j c_j \frac{(1 - c)^{2j}t^j}{(1 + ct)^{2j}},\\ &\text{(b)}\quad q(t) = \frac{c}{1 + c^2t}\sum_{j=0}^{\infty} c_j \frac{(1 - c^2)^j}{(1 + c^2t)^j},\\ &\text{(c)}\quad q(t) = \frac{1}{1 + \left(\frac{1 + c^2}{2}\right)t}\sum_{j=0}^{\infty} c_j \frac{\left(\frac{1 - c^2}{4}\right)^j t^{2j}}{\left[1 + \left(\frac{1 + c^2}{2}\right)t\right]^{2j}},\end{aligned}$$

where $c_j = (2j)!/2^{2j}(j!)^2$. The second of these corresponds to the method of [11]. Thus we see that various expansions are available which differ in speed of convergence and difficulty of inversion. If suitable partial fraction or other routines are available for inverting the individual terms, an expression like (a) might be useful for some values of c; in other cases, (b) might be satisfactory. Without giving detailed calculations of examples, we can see how ill-advised it is always to use, mechanically, the same routine in every case.

C. Other inversion techniques. Because of the large number of terms which must be kept in (2.3) in order to obtain reasonable accuracy (as discussed in Section 7A) when applying the techniques we have discussed, and because of the other shortcomings of these methods (see Section 7B), it is reasonable to investigate other inversion techniques. For example, in the problem of Section 2, if we first take the product with respect to k, we obtain $\prod_j \{\sinh [(2\pi)^{\frac{1}{2}}/\pi j]/[(2\pi)^{\frac{1}{2}}/\pi j]\}^{-\frac{1}{2}}$ for the Laplace transform, and one can try various manipulations with this expression. Another possibility, which seems more fruitful in this and many other problems, is that of direct numerical integration to invert the expression of (2.3).

In order to perform such an integration, one must first tabulate the function (2.3) for various values of the argument. A method which seems to be much more efficient than that of directly multiplying together an appropriately large number of terms of the product is to use the fact that, in a neighborhood of $v = 0$, we have

$$-\tfrac{1}{2} \sum_{i,j \geqq h} \log\left(1 + \frac{v}{i^2 j^2}\right) = \sum_{k=1}^{\infty} a_k v^k, \tag{7.11}$$

where $a_k = (-1)^k \left(\sum_{j \geqq h} j^{-2k}\right)^2/2k$ (these coefficients can be written in terms of Bernoulli numbers). On the basis of preliminary estimates of

$$g(v) = v^{-1} \prod_{i,j} (1 + v/i^2 j^2)^{-\frac{1}{2}}$$

on the proposed line of integration, the value of h can be chosen so as to make the series (7.11) convergent over that (finite) portion of the line where the integration will actually be performed. The series can then be evaluated for appropriate complex v, exponentiated, and the result multiplied by the remaining factor of $g(v)$, which can be expressed in terms of hyperbolic sines and of powers of linear functions of v. The numerical integration can then be performed. This was the method used to obtain the tables of Section 8.

A recent paper by Grenander, Pollak, and Slepian [3] discusses an interesting computational technique for obtaining an approximation to limiting distributions such as those discussed above by solving a set of linear equations whose solution approximates that of an integral equation for the limiting d.f. or c.f. The reader is referred to [3] for details and related discussion.

8. Tables. The inversion of (2.3) was carried out by the method outlined in the second paragraph of Section 7C, which was calculated to require much less machine time than any of the other available methods. The authors are grateful to Professor R. J. Walker for carying out the computations on the Cornell Computing Center's 220. Table I gives values (under H_0) of

$$F(x) = \lim_{n \to \infty} P\{\tfrac{1}{2}\pi^4 n B_n \leqq x\},$$

while Table II gives values of $F^{-1}(p)$.

It is not very difficult to program a computing machine to evaluate the statistic B_n or the modifications of it mentioned in Section 4. It may be worthwhile, especially for small n, to reduce the error introduced when using the

TABLE I

$$F(y) = \lim_{n\to\infty} P_{H_0}\{\tfrac{1}{2}\pi^4 nB_n \leqq y\}$$

y	$F(y)$	y	$F(y)$	y	$F(y)$
.30	.00000	2.10	.87275	3.90	.98546
.35	.00010	2.15	.88084	3.95	.98627
.40	.00086	2.20	.88835	4.00	.98702
.45	.00389	2.25	.89534	4.05	.98774
.50	.01158	2.30	.90185	4.10	.98841
.55	.02614	2.35	.90791	4.15	.98905
.60	.04867	2.40	.91357	4.20	.98965
.65	.07899	2.45	.91885	4.25	.99022
.70	.11594	2.50	.92377	4.30	.99075
.75	.15784	2.55	.92838	4.35	.99126
.80	.20293	2.60	.93268	4.40	.99174
.85	.24960	2.65	.93670	4.45	.99219
.90	.29652	2.70	.94047	4.50	.99261
.95	.34267	2.75	.94400	4.55	.99301
1.00	.38730	2.80	.94730	4.60	.99339
1.05	.42994	2.85	.95039	4.65	.99375
1.10	.47027	2.90	.95329	4.70	.99409
1.15	.50816	2.95	.95602	4.75	.99441
1.20	.54354	3.00	.95857	4.80	.99471
1.25	.57645	3.05	.96097	4.85	.99499
1.30	.60697	3.10	.96322	4.90	.99527
1.35	.63521	3.15	.96533	4.95	.99552
1.40	.66131	3.20	.96732	5.00	.99576
1.45	.68540	3.25	.96918		
1.50	.70763	3.30	.97094	5.50	.99755
1.55	.72813	3.35	.97259	6.00	.99858
1.60	.74704	3.40	.97414	6.50	.99918
1.65	.76449	3.45	.97561	7.00	.99952
1.70	.78060	3.50	.97698	7.50	.99972
1.75	.79547	3.55	.97828	8.00	.99983
1.80	.80922	3.60	.97949	8.50	.99990
1.85	.82193	3.65	.98064	9.00	.99994
1.90	.83369	3.70	.98172	9.50	.99997
1.95	.84459	3.75	.98274	10.00	.99998
2.00	.85469	3.80	.98370	10.50	.99999
2.05	.86406	3.85	.98461	11.00	1.00000

TABLE II

p	$F^{-1}(p)$	p	$F^{-1}(p)$
.9	2.286	.998	5.68
.95	2.844	.999	6.32
.98	3.622	.9995	6.96
.99	4.230	.9998	7.82
.995	4.851	.9999	8.47

limiting d.f. (in particular, in the limiting covariance function) by using $(n - 1)B_n$ instead of nB_n.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] G. Doetsch, *Theorie und Anwendung der Laplace Transformation*, Dover Pub., New York, 1943.

[3] U. Grenander, H. O. Pollak, and D. Slepian, "The distribution of quadratic forms in normal variates," *J. S. I. A. M.*, Vol. 7 (1959), pp. 374–401.

[4] Wassily Hoeffding, "A nonparametric test of independence," *Ann Math. Stat.*, Vol. 19 (1948), pp. 546–557.

[5] M. Kac, "On some connections between probability theory and differential and integral equations," *Proceedings of the Second Berkely Symposium of Mathematical Statistics and Probability*, University of California Press, 1951, pp. 180–215.

[6] M. Kac, J. Kiefer, and J. Wolfowitz, "On tests of normality and other tests of goodness of fit based on distance methods," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 189–211.

[7] J. Kiefer, "*K*-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 420–447.

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

[9] J. Kiefer and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function for vector chance variables," *Ann Math. Stat.*, Vol. 30 (1959), pp. 463–489.

[10] H. B. Mann and A. Wald, "On the choice of the number of class intervals in the application of the chi square test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 306–317.

[11] E. J. G. Pitman and Herbert Robbins, "Application of the method of mixtures of quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[12] M. Rosenblatt, "Limit theorems associated with variants of the von Mises statistic," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 617–623.