ON TESTS OF NORMALITY AND OTHER TESTS OF GOODNESS OF FIT BASED ON DISTANCE METHODS

By M. Kac, J. Kiefer,[1] and J. Wolfowitz[2]

Cornell University

Summary. The authors study the problem of testing whether the distribution function (d.f.) of the observed independent chance variables $x_1, \cdots, x_n$ is a member of a given class. A classical problem is concerned with the case where this class is the class of all normal d.f.'s. For any two d.f.'s $F(y)$ and $G(y)$, let $\delta(F, G) = \sup_y |F(y) - G(y)|$. Let $N(y \mid \mu, \sigma^2)$ be the normal d.f. with mean μ and variance σ^2. Let $G_n^*(y)$ be the empiric d.f. of $x_1, \cdots, x_n$. The authors consider, inter alia, tests of normality based on $v_n = \delta(G_n^*(y), \ N(y \mid \bar{x}, s^2))$ and on $w_n = \int (G_n^*(y) - N(y \mid \bar{x}, s^2))^2 \, d_y N(y \mid \bar{x}, s^2)$. It is shown that the asymptotic power of these tests is considerably greater than that of the optimum χ^2 test. The covariance function of a certain Gaussian process $Z(t)$, $0 \leqq t \leqq 1$, is found. It is shown that the sample functions of $Z(t)$ are continuous with probability one, and that

$$\lim_{n \to \infty} P\{nw_n < a\} = P\{W < a\}, \quad \text{where} \quad W = \int_0^1 [Z(t)]^2 \, dt.$$

Tables of the distribution of W and of the limiting distribution of $\sqrt{n}\, v_n$ are given. The role of various metrics is discussed.

1. Introduction. Let $x_1, \cdots, x_n$ be n independent chance variables with the same cumulative distribution function $G(x)$ (i.e., $G(x) = P\{x_1 < x\}$) which is unknown to the statistician. It is desired to test the hypothesis that $G(x)$ is a normal distribution. This is an old problem of considerable interest which has received a fair share of attention in the literature.

A commonly used test consists essentially in testing whether the third moment of $G(x)$ about its mean is zero and whether the ratio of the fourth moment about the mean to the square of the second moment about the mean is three. It is obvious that this is not really a test of normality, because there are many non-normal distributions which satisfy these conditions on the moments.

Perhaps the best of the commonly used large sample tests of normality is the χ^2 test due to Karl Pearson; see for example Cramér [1], Sections 30.1 and 30.3, and the recent results of Chernoff and Lehmann [2]. The asymptotic power of the χ^2 test was studied by Mann and Wald [3]. (It is true that these authors

---

Received August 30, 1954, revised January 10, 1955.

[1] Research under contract with the Office of Naval Research. The computations of the tables of the present paper were performed under the auspices of the Office of Naval Research to whom the authors are greatly obliged.

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

studied the problem of goodness of fit for a simple hypothesis, which in our problem would correspond to knowing the mean and variance of $G(x)$. However, it is plausible that the comparison in power which we will make below will be true a fortiori for our problem.) One of their principal results is the following. Define the distance $\delta(H_1, H_2)$ between any two distribution functions H_1 and H_2 by

$$\delta(H_1, H_2) = \sup_x |H_1(x) - H_2(x)|.$$

Suppose one tests the null hypothesis that $G(x) \equiv R(x)$, where $R(x)$ is a given continuous distribution, at some fixed level of significance. In [3] it is shown that the χ^2 test based on n observations and k_n intervals of equal probability under R, where k_n is chosen for each n so as to minimize the value Δ_n for which the minimum of the power function among alternatives R^* with $\delta(R^*, R) \geqq \Delta_n$ is $\frac{1}{2}$, gives $\Delta_n \doteq cn^{-2/5}$ when n is large. In Section 5 we shall show that the result of Mann and Wald just stated also holds if δ is replaced by the measure of discrepancy γ introduced in Section 5.

Let

$$\psi_x(a) = \begin{cases} 0, & x \leqq a; \\ 1, & x > a. \end{cases}$$

Let

$$G_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_x(x_i)$$

be the empiric d.f. of $x_1, \cdots, x_n$, that is, $G_n^*(x)$ is the proportion of x_i's less than x. The asymptotic distribution of $\delta(G, G_n^*)$ was found by Kolmogoroff [4] and it is known that $\delta(G, G_n^*)$ is of the order $n^{-1/2}$ in probability (that is, with probability arbitrarily close to one). If now one bases the test of the hypothesis that $G(x) \equiv R(x)$ on $\delta(R, G_n^*)$, with large values of δ significant, it follows that, for large n, it is sufficient that $\delta(R, R^*)$ be of the order $n^{-1/2}$ for the probability of rejecting the null hypothesis to be appreciable ($\geqq \frac{1}{2}$). To put matters a little differently: Let $\delta(R, R^*) = h$ be small (so that n has to be large in order to distinguish between R and R^*). Then, if n has to be equal to N in order to guarantee that the power at R^* of the χ^2 test of goodness of fit be at least $\frac{1}{2}$, when one uses the test based on δ it is enough that n be of the order of $N^{4/5}$. This is a considerable improvement (for large n). The result just stated holds also for the classical "ω^2" test if δ is replaced in the above by γ; this follows from the results of Section 5.

Let us return to the problem of testing the composite null hypothesis whether $G(x)$ is normal (its mean and variance being unknown). One of the present authors has been developing the minimum distance method for estimation and testing hypotheses in a number of papers (Wolfowitz [5], [6], [7], [8]). In accord with this method it is proposed in [5] (page 149) that this test of normality be based on $\delta(G_n^*, N^{**})$ with the large values critical. Here N^{**} is the class of all normal distributions, and the distance of any d.f. $H(x)$ from the class N^{**} is given by

$$\delta(H, N^{**}) = \inf_{N \varepsilon N^{**}} \delta(H, N).$$

In Sections 2, 3, and 4 we investigate tests based on various "distance" criteria, constructed in the spirit of the above discussion, and in Section 5 we discuss the asymptotic power of these tests. As stated in [5], the minimum distance method is not limited merely to testing normality, and in Section 4 we discuss its application to other tests of goodness of fit.

Throughout this paper we discuss tests, computations, and power considerations in terms of particular (normal or rectangular) examples, but it will be obvious that the results of Sections 2, 3, and 5 may in general be carried over to testing composite hypotheses involving parametric families. The minimum distance method applies in principle to even more complicated families of distribution functions about which one desires to test a hypothesis. For hypotheses of a more complicated nature than our examples, there will often exist the additional complication that the test criterion may not be distribution-free under the null hypothesis.

Tests may also be constructed using other "distances" than those mentioned in Section 2 and above (see also [5], pp. 148–149, and [6], p. 10 in this connection), involving other "estimators" than those of Sections 2 and 3, and involving other modifications of the notion of "distance methods" as motivated in this section.

2. Testing normality. For convenience we divide this section into several subsections.

2.1. The computation of $\delta(G_n^*, N^{**})$ offers considerable difficulty; unpublished work on this subject has been done by Blackman. The distribution of $\delta(G_n^*, N^{**})$ under the null hypothesis is still unknown. (It is easy to verify that, when $G(x)$ is a member of N^{**}, the distribution of $\delta(G_n^*, N^{**})$ does not depend on $G(x)$. Thus the composite null hypothesis determines uniquely the distribution of the test criterion.) In Section 4 we shall give an example of another problem of testing hypotheses where the limiting distribution of the minimum distance criterion is explicitly calculated. In the present Section 2 we shall consider some other "distance" tests. (For the case where either the mean or variance is assumed known, the test corresponding to that discussed in this section is being studied by Darling [18]; the suggestion of using such tests is apparently due to Cramér.) Let

$$\bar{x} = \frac{1}{n} \sum_1^n x_i, \qquad s^2 = \frac{1}{n} \sum_1^n x_i^2 - \bar{x}^2.$$

Let $N(x \mid \bar{x}, s^2)$ be the normal distribution function with mean $\bar{x}$ and variance s^2. One can base the test of normality on $v_n = \delta(G_n^*(x), N(x \mid \bar{x}, s^2))$, with the large values critical. It is easy to see that, when $G(x)$ is actually a member of

N^{**}, the distribution of v_n does not depend on $G(x)$. The distribution of v_n does not seem easy to obtain, except, for example, by Monte Carlo methods. Another test criterion, similar to the above but of the "ω^2" type, is to base the test on

$$w_n = \frac{1}{n} \int_0^1 Z_{r,n}^2 \, dr$$

where $Z_{r,n}$ is defined in 2.2.[3] (The idea of the "ω^2" test, which is defined precisely in Section 6, is due to von Mises, with a modification by Smirnov.) Sections 2.2 to 2.6 are concerned with the limiting distribution of nw_n as $n \to \infty$ when G is normal. The power of the tests based on v_n and w_n is discussed in Section 5. In Section 3 we treat briefly an example which illustrates the construction of test criteria which are similar to v_n and w_n but which may use "inefficient" statistics to estimate which member of N^{**} G is (if it is); such techniques may have obvious practical importance.

2.2. Let $x_1, x_2, \cdots$ be independent, identically distributed Gaussian random variables with zero mean and unit variance. Let $G_n^*(x)$ be as defined in Section 1. We shall use the following notation:

$$\phi(x) = \phi(-x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad I(x) = \int_x^\infty \phi(y) \, dy,$$

$$J(y) = I^{-1}(1 - y) = \{x \mid y = 1 - I(x)\}, \qquad g(x) = \phi(x)/I(x).$$

We also let $[z]$ = smallest integer $\geqq z$. For $0 \leqq r \leqq 1$, we put $U_{r,n} = [rn]$th from the bottom among the ordered $x_1, \cdots, x_n$. Finally, let[4]

$$Z_{r,n} = \sqrt{n}[G_n^*(\sqrt{S_n'}J(r) + \bar{x}_n) - r]$$

where

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \qquad S_n = \frac{1}{n} \sum_{i=1}^n x_i^2, \qquad S_n' = S_n - \bar{x}_n^2.$$

(Remark: S_n and S_n' may be used equally well in the definition of $Z_{r,n}$ for the purpose of obtaining the limiting distribution; in applications S_n' should probably be used.)

2.3. We shall show that for $0 \leqq r_1 < r_2 < \cdots < r_k \leqq 1$, the quantities $Z_{r_i,n}$, where $1 \leqq i \leqq k$, are asymptotically jointly normally distributed with zero means and covariance function (for $Z_{s,n}$ and $Z_{t,n}$ as $n \to \infty$)

$$K(s, t) = \min(s, t) - st - (2\pi)^{-1}(1 + J(s)J(t)/2)e^{-[J(s)^2+J(t)^2]/2}. \tag{2.1}$$

We shall show here that (2.1) follows from the fact proved in Section 2.4, that $\sqrt{n}x_n$, $\sqrt{n}(S_n - 1)$, and $\sqrt{n}(U_{r_i,n} - J(r_i))$, where $1 \leqq i \leqq k$, are jointly

[3] This definition is equivalent to that given in the summary at the beginning of the paper.

[4] Elsewhere in this paper, in the summary, for example, the conventional symbol s^2 is sometimes used instead of the typographically easier (here) symbol S_n'. Both represent the same thing. Also $\bar{x}_n$ and $\bar{x}$ are used interchangeably in a manner to cause no confusion.

asymptotically normal with means 0 and covariance matrix (presented in the same order) given by

$$
(2.2) \qquad \begin{Vmatrix} 1 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 2 & J(r_1) & J(r_2) & \cdots & J(r_k) \\ 1 & J(r_1) & \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & J(r_k) & \lambda_{k1} & \lambda_{k2} & \cdots & \lambda_{kk} \end{Vmatrix},
$$

$$
\lambda_{ij} = \lambda_{ji} = \frac{r_i(1 - r_j)}{\phi(J(r_i))\phi(J(r_j))}, \qquad r_i \leqq r_j .
$$

In fact, all these results are well known ([1], p. 364, 369) except for the joint normality and the last k entries in the first two rows (and columns).

Assume then for the moment that the above is proved. The event $\{Z_{r,n} \leqq z\}$ may be written as $\{[\text{number of } x_i \text{ which are } < \sqrt{S_n'}J(r) + \bar{x}_n] \leqq nr + \sqrt{n}z\}$. This in turn is the same as $\{U_{r+z/\sqrt{n},\, n} \geqq \sqrt{S_n'}J(r) + \bar{x}_n\}$, or

$$
\{\sqrt{n}\bar{x}_n + \sqrt{n}J(r)(\sqrt{S_n'} - 1) - \sqrt{n}(U_{r+z/\sqrt{n},\,n} - J(r + z/\sqrt{n}))
$$
$$
\leqq \sqrt{n}[J(r + z/\sqrt{n}) - J(r)]\}.
$$

As $n \to \infty$, neglecting terms of higher order in probability, we may replace $\sqrt{n}[J(r + z/\sqrt{n}) - J(r)]$ by $zJ'(r)$ and $\sqrt{n}(\sqrt{S_n'} - 1)$ by $\frac{1}{2}\sqrt{n}(S_n - 1)$. We conclude that

$$
\lim_{n\to\infty} P\{Z_{r_1,n} \leqq z_1, \cdots, Z_{r_k,n} \leqq z_k\}
$$

$$
(2.3) \qquad = \lim_{n\to\infty} P\left\{\frac{1}{J'(r_i)}[\sqrt{n}\bar{x}_n + \tfrac{1}{2}\sqrt{n}J(r_i)(S_n - 1)\right.
$$
$$
\left. - \sqrt{n}(U_{r_i+z_i/\sqrt{n},\,n} - J(r_i + z_i/\sqrt{n}))] \leqq z_i, \qquad i = 1, \cdots, k\right\}.
$$

Thus, $Z_{r_1,n}, Z_{r_2,n}, \cdots, Z_{r_k,n}$ are jointly asymptotically normal, with the same limiting distribution as the quantities

$$
[\sqrt{n}\bar{x}_n + \tfrac{1}{2}\sqrt{n}J(r_i)(S_n - 1) - \sqrt{n}(U_{r_i,n} - J(r_i))]/J'(r_i), \quad i = 1, \cdots, k.
$$

From this and (2.2), the result (2.1) follows by direct computation and the fact that $J'(r) = 1/\phi(J(r))$.

2.4. It remains to prove the joint asymptotic normality of the quantities mentioned in the paragraph following (2.1), and to verify the last k entries of the first two rows (and columns) of (2.2). We shall in fact compute here only the limiting distribution of $\sqrt{n}\bar{x}_n$, $\sqrt{n}(S_n - 1)$, and $\sqrt{n}(U_{r,n} - J(r))$ where $0 < r < 1$; from this will follow the desired result of (2.2), and the method of proof of joint asymptotic normality for the $k + 2$ random variables previously mentioned will be evident from that for the case $k = 1$ considered here (note especially p. 369 of [1] in this connection).

We begin by introducing a process which will enable us to simplify this computation. For fixed n, let $Y_{[rn]}$ be a chance variable whose distribution function is that of $U_{r,n}$ above. Let Y_i, where $1 \leqq i \leqq n$ and $i \neq [rn]$, be random variables whose joint conditional distribution, given that $Y_{[rn]} = y$, is such that these $n - 1$ random variables are (conditionally) independent with Y_i having a (conditional) truncated normal distribution given by

$$P\{Y_i < z \mid Y_{[rn]} = y\} = \begin{cases} \min[1, \quad (1 - \mathrm{I}(z))/(1 - \mathrm{I}(y))], & 1 \leqq i < [rn] \\ \max[0, \quad 1 - \mathrm{I}(z)/\mathrm{I}(y)], & [rn] < i \leqq n. \end{cases}$$

If now $Y_1, \cdots, Y_n$ are reordered with probability $1/n!$ for each possible reordering, and the resulting reordered variables are labeled $X_1^1, \cdots, X_n^1$, it is easily verified that $X_1^1, \cdots, X_n^1$ have the same joint distribution as the $x_1, \cdots, x_n$ considered in Section 1. We shall use the process $Y_1, \cdots, Y_n$ to compute the conditional distribution of S_n and $\bar{x}_n$, given that $U_{r,n} = y = J(r) + w/\sqrt{n}$, say. Let

$$([nr] - 1)\bar{X}_n^1 = \sum_1^{[rn]-1} Y_i, \qquad (n - [nr])\bar{X}_n^2 = \sum_{[rn]+1}^{n} Y_i,$$

$$([nr] - 1)S_n^1 = \sum_1^{[rn]-1} Y_i^2, \qquad (n - [nr])S_n^2 = \sum_{[rn]+1}^{n} Y_i^2.$$

It is easy to compute that for the truncated normal distribution from y to ∞ mentioned above, the first four moments about the origin are

$$\mu_1(y) = g(y), \qquad \mu_2(y) = 1 + y\, g(y),$$

$$\mu_3(y) = (y^2 + 2)\, g(y), \qquad \mu_4(y) = 3 + (y^3 + 3y)\, g(y).$$

The corresponding moments for the truncated distribution from $-\infty$ to y are clearly $-\mu_1(-y)$, $\mu_2(-y)$, $-\mu_3(-y)$, and $\mu_4(-y)$. From these and [1], p. 364, we obtain that

$$A_n = \sqrt{[nr] - 1}\, (\bar{X}_n^1 + g(-y)), \qquad B_n = \sqrt{[nr] - 1}(S_n^1 - 1 + y\, g(-y))$$

are asymptotically conditionally normal with means 0 and covariance matrix

$$\begin{pmatrix} 1 - y\, g(-y) - [g(-y)]^2 & -(y^2 + 1)\, g(-y) - y[g(-y)]^2 \\ -(y^2 + 1)\, g(-y) - y[g(-y)]^2 & 2 - (y^3 + y)\, g(-y) - y^2[g(-y)]^2 \end{pmatrix},$$

and that $C_n = \sqrt{n - [nr]}\, (\bar{X}_n^2 - g(y))$ and $D_n = \sqrt{n - [nr]}\, (S_n^2 - 1 - yg(y))$ are asymptotically conditionally normal with means 0 and covariance matrix

$$\begin{pmatrix} 1 + y\, g(y) - [g(y)]^2 & (y^2 + 1)\, g(y) - y[g(y)]^2 \\ (y^2 + 1)\, g(y) - y[g(y)]^2 & 2 + (y^3 + y)\, g(y) - y^2[g(y)]^2 \end{pmatrix},$$

with (A_n, B_n) conditionally independent of (C_n, D_n). By using the Liapounoff condition one proves that the approach to normality is uniform for any finite w-interval. Let $O_p(\ )$ denote "order of ( ) in probability" (e.g., [9]). Noting that $\sqrt{[nr]} - \sqrt{nr} = O(1/\sqrt{n})$ and that $\sqrt{n}(y/n) = O_p(1/\sqrt{n})$, we obtain as an expression for $\sqrt{n}\bar{x}_n$, given that $U_{r,n} = y$,

$$\sqrt{n}\bar{x}_n = \sqrt{r}A_n + \sqrt{1-r}\,C_n + \sqrt{n}[-r\,g(-y) + (1-r)\,g(y)] + O_p(1/\sqrt{n}).$$

But

$$-r\,g\,(-J(r) - w/\sqrt{n}) + (1-r)\,g\,(J(r) + w/\sqrt{n})$$
$$= \frac{w}{\sqrt{n}} \cdot \frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{n}\right),$$

so that

$$\sqrt{n}\bar{x}_n = \sqrt{r}A_n + \sqrt{1-r}\,C_n + w\,\frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

In similar fashion, we obtain

$$\sqrt{n}(S_n - 1) = \sqrt{r}B_n + \sqrt{1-r}D_n + wJ(r)\,\frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

In both cases the $O_p(1/\sqrt{n})$ term is uniform in every bounded interval of w. Since $W_n = \sqrt{n}(U_{r,n} - J(r))$ is asymptotically normal with mean 0 and variance $r(1-r)/[\phi(J(r))]^2$, the desired asymptotic joint normality follows easily from the last two displayed expressions. The covariance of $\sqrt{n}\bar{x}_n$ and W_n is most easily computed as $E\{W_n \cdot E\{\sqrt{n}\bar{x}_n \mid W_n\}\}$, that of $\sqrt{n}(S_n - 1)$ and W_n being computed similarly; these give the desired results for the last k entries of the first two rows (and columns) of (2.2).

2.5. We now show that there exists a representation of any Gaussian process with mean identically zero and a covariance function like that of (2.1), whose sample functions are continuous with probability one (w.p. 1).

Let $W(t)$, $0 \leqq t \leqq 1$, be the Kac-Siegert representation ([12]) of a Gaussian process with continuous covariance function $K'(s, t)$. Let $\{\lambda_k\}$ be the eigenvalues and $\{\varphi_k(t)\}$ the corresponding normalized eigenfunctions of $K'(s, t)$. Suppose furthermore that $g(t)$ in L^2 is such that

$$K'(s, t) - g(s)g(t) = K''(s, t)$$

is a covariance function. We shall show how to construct explicitly a process $Z(t)$ with covariance function $K''(s, t)$ in terms of the process $W(t)$. (All processes studied in this section are to have mean zero.)

We first prove two lemmas.

LEMMA 1. *A necessary and sufficient condition that $K''(s, t)$ be positive definite is that*

$$\beta^2 = \sum_{k=1}^{\infty} \frac{g_k^2}{\lambda_k} \leqq 1, \qquad g_k = \int_0^1 g(t)\,\varphi_k(t)\,dt.$$

(If a λ_k is 0, it follows immediately from the positivity of $K''(s, t)$ that $g(t)$ is orthogonal to all the eigenfunctions belonging to λ_k; thus the corresponding g_k vanish and we interpret ${g_k}^2/\lambda_k$ as zero.)

Only the necessity of the condition is used in the application of the lemma. The proof of sufficiency is included because it is so brief and sufficiency seems to be of interest.

1. *Necessity*. Set $\psi(t) = \sum_1^n v_k \varphi_k(t)$ and note that by definition

$$\int_0^1 \int_0^1 (K'(s, t) - g(s)\, g(t))\, \psi(s)\, \psi(t)\, ds\, dt \geqq 0.$$

Evaluating the double integral we obtain

$$\sum_1^n v_k^2 \lambda_k \geqq \left(\sum_1^n v_k\, g_k\right)^2,$$

a $\quad = g_k/\lambda_k$

$$\sum_1^n \frac{g_k^2}{\lambda_k} \geqq \left(\sum_1^n \frac{g_k^2}{\lambda_k}\right)^2.$$

Thus $\sum_1^n g_k^2/\lambda_k \leqq 1$ for *every* n and the theorem follows.

2. *Sufficiency*. If $\sum_1^\infty g_k^2/\lambda_k \leqq 1$ we have

$$\left(\sum_1^\infty v_k\, g_k\right)^2 = \left(\sum_1^n v_k \sqrt{\lambda_k}\, \frac{g_k}{\sqrt{\lambda_k}}\right)^2 \leqq \left(\sum_1^\infty v_k^2 \lambda_k\right)\left(\sum_1^\infty \frac{g_k^2}{\lambda_k}\right) \leqq \sum_1^\infty v_k^2 \lambda_k,$$

and positiveness of $K'(s, t) - g(s)\, g(t)$ follows.

LEMMA 2. *The series* $\sum_1^\infty g_k \varphi_k(t)$ *converges uniformly* (*to* $g(t)$).

We have $|\sum_m^n g_k \varphi_k(t)| \leqq \sqrt{\sum_m^n g_k^2/\lambda_k}\, \sqrt{\sum_m^n \lambda_k \varphi_k^2(t)}$. By Mercer's theorem, $\sum_1^\infty \lambda_k \varphi_k^2(t) = K(t, t)$. Hence $|\sum_m^n g_k \varphi_k(t)| \leqq \sqrt{K(t, t)}\, \sqrt{\sum_m^n g_k^2/\lambda_k}$. Since $\sum g_k^2/\lambda_k$ converges (by Lemma 1), Lemma 2 follows.

We are now ready to prove that

$$Z(t) = W(t) - \frac{(1 - \sqrt{1 - \beta^2})}{\beta^2}\, g(t) \sum_{k=1}^\infty \frac{g_k}{\lambda_k} \int_0^1 W(t)\, \varphi_k(t)\, dt \tag{2.4}$$

has the covariance function $K''(s, t)$.

Note first that the chance variables

$$\left\{\frac{1}{\sqrt{\lambda_k}} \int_0^1 W(t) \varphi_k(t)\, dt\right\}$$

are independently and normally distributed, with mean 0 and variance 1. Since $\sum_1^\infty g_k^2/\lambda_k < \infty$, the series

$$\sum_1^\infty \frac{g_k}{\sqrt{\lambda_k}} \cdot \frac{1}{\sqrt{\lambda_k}} \int_0^1 W(t) \varphi_k(t)\, dt$$

converges in the mean (and even w.p. 1) and thus defines a random variable.

To calculate the covariance of $Z(t)$ we need

$$E\left\{W(t) \sum_{k=1}^{\infty} \frac{g_k}{\lambda_k} \int_0^1 W(t)\varphi_k(t)\, dt\right\}.$$

But $E\{W(t) \int_0^1 W(t)\varphi_k(t)\, dt\} = \int_0^1 E\{W(t)W(t)\}\varphi_k(t)\, dt = \int_0^1 K'(t, t)\varphi_k(t)\, dt = \lambda_k\varphi_k(t)$, and hence

$$E\left\{W(t) \sum_{k=1}^{\infty} \frac{g_k}{\lambda_k} \int_0^1 W(t)\varphi_k(t)\, dt\right\} = \sum_{k=1}^{\infty} g_k \varphi_k(t) = g(t).$$

The covariance function of $Z(t)$ now comes out to be

$$K'(s, t) - 2\beta^{-2}(1 - \sqrt{1 - \beta^2})\, g(s)g(t) + \beta^{-2}(1 - \sqrt{1 - \beta^2})^2 g(s)g(t)$$
$$= K'(s, t) - g(s)g(t) = K''(s, t).$$

Since $g(t)$ is continuous (by Lemma 2) it follows, in particular, that the sample functions of $Z(t)$ are continuous w.p. 1 if the sample functions of $W(t)$ are continuous w.p. 1. Now let $K'(s, t) = \min(s, t) - st$. Then the sample functions of $W(t)$ are continuous w.p. 1. The application of the above result twice (once for each remaining negative product in (2.1)) then proves that the representation $Z(t)$, which is Gaussian with mean zero and covariance function (2.1), has continuous sample functions, w.p. 1.

2.6. From the results of Sections 2.2 to 2.5, it is easily verified that the demonstration given by Kac on pp. 197–198 of [10] for the case $K(s, t) = \min(s, t) - st$ carries over with only slight modifications to the case now under discussion where $K(s, t)$ is given by (2.1). We conclude that

$$\lim_{n\to\infty} P\{nw_n < a\} = P\{W < a\}, \tag{2.5}$$

where $W = \int_0^1 [Z(t)]^2\, dt$ and $Z(t)$ is Gaussian with covariance function given by (2.1) and sample functions continuous w.p. 1. Modifying slightly the technique of Kac and Siegert [11], [12] as applied on pp. 199–200 of [10], we now study the distribution of Z. We write

$$K(s, t) = K^*(s, t) - h_1(s)\, h_1(t) - h_2(s)\, h_2(t) \qquad 0 \leqq s, t \leqq 1, \tag{2.6}$$

where $K^*(s, t) = \min(s, t) - st$ and $h_k(s) = (2\pi)^{-1/2}[J(s)/\sqrt{2}]^{k-1}e^{-[J(s)]^2/2}$ for $k = 1, 2$. Let $\lambda_1, \lambda_2, \cdots$ be the eigenvalues (all positive, since K is positive definite) of the integral equation

$$\int_0^1 K(x, y)\, \varphi(y)\, dy = \lambda\, \varphi(x). \tag{2.7}$$

Following the demonstration of [10], we conclude that the characteristic function of W is

$$Ee^{iW\xi} = \prod_{j=1}^{\infty} (1 - 2i\xi\lambda_j)^{-1/2}. \tag{2.8}$$

Thus, we may express W as

$$W = \sum_{j=1}^{\infty} R_j, \tag{2.9}$$

where the R_j are independent and R_j has density function

$$\frac{1}{\sqrt{2\pi r\lambda_j}}\, e^{-r/2\lambda_j}, \qquad r > 0. \tag{2.10}$$

(We remark that it may of course be easier to obtain the convolution with itself of the distribution of W rather than the latter itself. This could be used if one based a test of normality on the sum of two W_m's, each computed from half the sample of size $n = 2m$.)

We now give a procedure for finding the λ_j. For any eigenvalue λ and corresponding eigenfunction φ (not necessarily normalized) of (2.7), write

$$C_i = \int_0^1 h_i(y)\, \phi(y)\, dy, \qquad i = 1, 2. \tag{2.11}$$

We can rewrite (2.7) as

$$\int_0^1 K^*(x, y)\, \varphi(y)\, dy - C_1 h_1(x) - C_2 h_2(x) = \lambda\, \varphi(x). \tag{2.12}$$

Differentiating twice with respect to x and writing $\mu^2 = 1/\lambda$ (we may consider $\mu > 0$ in the sequel), we obtain

$$\varphi''(x) + \mu^2\varphi(x) = -C_1\mu^2 h_1''(x) - C_2\mu^2 h_2''(x). \tag{2.13}$$

Any eigenvalue λ and eigenfunction $\phi(x)$ of (2.7) satisfy (2.13), (2.11), and

$$\varphi(0) = \phi(1) = 0. \tag{2.14}$$

Conversely if λ and $\phi(x)$ satisfy these conditions they are an eigenvalue and eigenfunction of (2.7). For let

$$\int_0^1 K(x, y)\, \phi(y)\, dy = \lambda\, \theta(x).$$

As we obtained (2.13) we get

$$\theta''(x) + \mu^2\phi(x) = -C_1\mu^2 h_1''(x) - C_2\mu^2 h_2''(x).$$

Hence $\theta''(x) = \phi''(x)$, and since also $\theta(0) = \theta(1) = 0$ we have $\theta(x) = \phi(x)$.

Our problem now is to find a value μ for which there exist a function $\phi(x)$ and constants C_1 and C_2 satisfying (2.13), (2.14), and (2.11). For given C_1 and C_2 the general solution of (2.13) can be written

$$\phi(x) = A \sin \mu x + B \cos \mu x - \mu C_1 g_1(x) - \mu C_2 g_2(x), \tag{2.15}$$

where

$$g_i(x) = \int_x^{1/2} h_i''(t) \sin \mu(t - x)\, dt. \tag{2.16}$$

Applying conditions (2.14) and (2.11) gives

$$0 = B - \mu C_1 g_1(0) - \mu C_2 g_2(0), \qquad 0 = A \sin \mu + B \cos \mu - \mu C_1 g_1(1) - \mu C_2 g_2(1),$$

$$\begin{aligned} C_i = A \int_0^1 h_i(x) \sin \mu x \, dx + B \int_0^1 h_i(x) \cos \mu x \, dx \\ -\mu C_1 \int_0^1 h_i(x)\, g_1(x)\, dx - \mu C_2 \int_0^1 h_i(x)\, g_2(x)\, dx. \end{aligned} \tag{2.17}$$

These equations have a nontrivial solution for A, B, C_1, and C_2 if and only if the determinant $D(\mu)$ of the coefficient of these four quantities is zero. Hence the eigenvalues of (2.7) are determined by the roots of $D(\mu) = 0$.

The following method of computing $D(\mu)$ is due to R. J. Walker, to whom the authors are greatly obliged.

We first note some pertinent properties of $h_i(x)$ and $g_i(x)$.

$$h_1(1 - x) = h_1(x), \quad h_1(0) = 0, \quad h_1(\tfrac{1}{2}) = 1/\sqrt{2\pi}, \quad h_1'(\tfrac{1}{2}) = 0;$$

$$h_2(1 - x) = -h_2(x), \quad h_2(0) = 0, \quad h_2(\tfrac{1}{2}) = 0, \quad h_2'(\tfrac{1}{2}) = 1/\sqrt{2};$$

$$\begin{aligned} g_1(1 - x) &= \int_{1-x}^{1/2} h_1''(t) \sin \mu(t - 1 + x)\, dt = \int_x^{1/2} h_1''(1 - s) \sin \mu(s - x)\, ds \\ &= g_1(x); \end{aligned}$$

and similarly

$$g_2(1 - x) = -g_2(x).$$

It follows that $g_1(1) = g_1(0)$ and $g_2(1) = -g_2(0)$, and

$$\int_0^1 h_i(x)\, g_j(x)\, dx = \begin{cases} 2 \int_0^{1/2} h_i(x) g_i(x), & j = i; \\ 0, & j \neq i. \end{cases}$$

Also, using $\sin \mu x = \sin \frac{1}{2}\mu \cos \mu(\frac{1}{2} - x) - \cos \frac{1}{2}\mu \sin \mu(\frac{1}{2} - x)$, we get

$$\int_0^1 h_1(x) \sin \mu x\, dx = 2 \sin \tfrac{1}{2}\mu \int_0^{1/2} h_1(x) \cos \mu(\tfrac{1}{2} - x)\, dx,$$

$$\int_0^1 h_2(x) \sin \mu x\, dx = -2 \cos \tfrac{1}{2}\mu \int_0^{1/2} h_2(x) \sin \mu(\tfrac{1}{2} - x)\, dx,$$

with similar reductions for the coefficients of B in (2.17).

Introducing these simplifications we get by direct computation $D(\mu) = -2D_1(\mu)\, D_2(\mu)$, where

$$\begin{aligned} D_1(\mu) = \cos \tfrac{1}{2}\mu \left[1 + 2\mu \int_0^{1/2} h_1(x)\, g_1(x)\, dx \right] \\ -2\mu g_1(0) \int_0^{1/2} h_1(x) \cos \mu(\tfrac{1}{2} - x)\, dx, \end{aligned} \tag{2.18}$$

$$D_2(\mu) = \sin \tfrac{1}{2}\mu \left[1 + 2\mu \int_0^{1/2} h_2(x)\, g_2(x)\, dx\right] - 2\mu g_2(0) \int_0^{1/2} h_2(x) \sin \mu(\tfrac{1}{2} - x)\, dx. \tag{2.19}$$

These equations can be put in a form more suitable for computation. Integrating (2.16) by parts gives, for $i = 1$,

$$g_1(x) = -\frac{\mu}{\sqrt{2\pi}} \cos \mu(\tfrac{1}{2} - x) + \mu h_1(x) - \mu^2 \int_x^{1/2} h_1(t) \sin \mu(t - x)\, dt$$

$$g_1(0) = -\frac{\mu}{\sqrt{2\pi}} \cos \tfrac{1}{2}\mu - \mu^2 \int_0^{1/2} h_1(t) \sin \mu t\, dt.$$

Putting these in (2.18) gives

$$\begin{aligned} D_1(\mu) = \cos \tfrac{1}{2}\mu &\left[1 + 2\mu^2 \int_0^{1/2} h_1^2(x)\, dx\right] \\ &+ 2\mu^3 \left[\iint_{S+T} h_1(x)\, h_1(t) \cos \mu(\tfrac{1}{2} - x) \sin \mu t\, dA \right. \\ &\left. - \iint_T h_1(x)\, h_1(t) \sin \mu(t - x) \cos \tfrac{1}{2}\mu\, dA\right], \end{aligned}$$

where S is the triangle bounded by the lines $t = 0$, $t = x$, and $x = \frac{1}{2}$, and T the triangle bounded by $x = 0$, $x = t$, and $t = \frac{1}{2}$. The sum of the integrals over T reduces to

$$\iint_T h_1(x)\, h_1(t) \cos \mu(\tfrac{1}{2} - t) \sin \mu x\, dA,$$

which equals the integral over S. Hence, finally

$$D_1(\mu) = \cos \tfrac{1}{2}\mu \left[1 + 2\mu^2 \int_0^{1/2} h_1^2(x)\, dx\right] + 4\mu^3 \int_0^{1/2} h_1(x) \sin \mu x\, dx \int_x^{1/2} h_1(t) \cos \mu(\tfrac{1}{2} - t)\, dt. \tag{2.20}$$

Similarly, (2.19) becomes

$$D_2(\mu) = \sin \tfrac{1}{2}\mu \left[1 + 2\mu^2 \int_0^{1/2} h_2^2(x)\, dx\right] + 4\mu^3 \int_0^{1/2} h_2(x) \sin \mu x\, dx \int_x^{1/2} h_2(t) \sin \mu(\tfrac{1}{2} - t)\, dt. \tag{2.21}$$

The method just developed for obtaining the eigenvalues of (2.7) seems more accurate and computationally simpler than other methods, such as those employing trial functions. In Section 6 the smallest few zeros of the functions $D_1(\mu)$ and $D_2(\mu)$ of (2.20) and (2.21) are tabulated, and an approximation for the distribution of W is thereby obtained. This approximation is compared with empirical distribution functions of nw_n obtained by sampling. Empirical distribution functions of nw_n and $\sqrt{n}v_n$ are tabulated, and certain other interesting results of the sampling experiments are noted (for example, the joint distribution of the classical Kolmogoroff and von Mises statistics D_n and ω_n^2, which are defined precisely in Section 6).

3. Tests using quantiles. The relative ease of computing the limiting distributions of various possible test criteria of the type considered in this paper will of course depend on the particular problem. Thus, the use of the sample mean and variance in non-normal cases may lead to more complicated results than those of Section 2. A tool which may be used in all cases (where the hypothesized family has finitely many natural parameters, is normal or not, where a simple sufficient statistic does or does not exist, etc.) of density functions, with about equal complexity in all cases, is the use of sample quantiles (as many as necessary) to estimate the "true" d.f. if the null hypothesis is true. (Of course, the more unknown parameters and hence quantiles which must be used, the messier will be the result.) For computational reasons, tests constructed in this manner will sometimes be more practical to use than those involving a sufficient statistic. The results on power in Section 5 apply also here.

As an example, suppose the pth sample quantile $U_{p,n}$, $0 < p < 1$, is to be used to estimate the corresponding population parameter of a family of d.f.'s $F(x - \theta)$ for $-\infty < \theta < \infty$, in testing whether or not the "true" d.f. is a member of this family. We suppose without loss of generality that $F(0) = p$, and we denote by f the density function of F. We assume f to be continuous and positive in a neighborhood of 0.

Then letting $\psi = F^{-1}$ and $Z_{r,n} = \sqrt{n}[F_n(\psi(r) + U_{p,n}) - r]$, an argument like that of Sections 2.3 and 2.4 leads easily to the conclusion that, for $0 \leqq r_1 \leqq \cdots \leqq r_k \leqq 1$, the limiting distribution of $Z_{r_i,n}$, $1 \leqq i \leqq k$, is the same as that of

$$\sqrt{n}[U_{p,n} - [U_{r_i,n} - \psi(r_i)]] \,/\, \psi'(r_i), \qquad 1 \leqq i \leqq k,$$

and, in particular, is Gaussian. Putting $\gamma(r) = f(\psi(r))$ and

$$g(r) = \sqrt{p(1-p)}\,\frac{\gamma(r)}{\gamma(0)} - \frac{\min\,[p, r] - pr}{\sqrt{p(1-p)}},$$

we obtain for the limiting covariance function, for $0 \leqq s,t \leqq 1$,

$$\begin{aligned} K(s, t) &= \gamma(s)\,\gamma(t)\left[\frac{p(1-p)}{[\gamma(0)]^2} - \frac{\min\,(p, s) - ps}{\gamma(0)\,\gamma(s)}\right. \\ &\qquad\qquad \left. - \frac{\min\,(p, t) - pt}{\gamma(0)\,\gamma(t)} + \frac{\min\,(s, t) - st}{\gamma(s)\,\gamma(t)}\right] \\ &= g(s)\,g(t) - \frac{[\min\,(p, s) - ps][\min\,(p, t) - pt]}{p(1-p)} + \min\,(s, t) - st \end{aligned}$$

$$= g(s)\,g(t) + \begin{cases} \min(s, t) - st/p & s, t \leqq p, \\ \min(s-p, t-p) - \dfrac{(s-p)(t-p)}{1-p} & s, t \geqq p, \\ 0 & \text{otherwise.} \end{cases}$$

For computing purposes (e.g., by Monte Carlo methods), the last form of the covariance function is useful. Let $X_1(t)$ and $X_2(t)$ be Gaussian, each with the familiar covariance function $\min(s, t) - st$, and let X be a normal random variable with mean zero and variance 1, where X, X_1, and X_2 are independent. Defining $X_1(t) = X_2(t) = 0$ if $t < 0$ or $t > 1$, let

$$Z(t) = g(t)X + \sqrt{p}\, X_1\left(\frac{t}{p}\right) + \sqrt{1-p}\, X_2\left(\frac{t-p}{1-p}\right), \quad 0 \leqq t \leqq 1.$$

Then Z has the covariance function $K(s, t)$. As an example, if F is normal with unit variance and $p = \frac{1}{2}$, we have

$$(Zt) = [\tfrac{1}{2}e^{-\psi(t)^2/2} = \min(t, 1-t)]X + \frac{1}{\sqrt{2}} X_1(2t) + \frac{1}{\sqrt{2}} X_2(2t-1).$$

4. The rectangular distribution. In Section 4.1 we give an example where the minimum distance method statistic can be computed explicitly; in Section 4.2 we comment on the limiting distribution of test criteria of the type treated in Section 2 in the present case.

4.1. Let

$$F(x;\theta) = \begin{cases} 0 & x < \theta - \frac{1}{2}, \\ x - \theta + \frac{1}{2} & \theta - \frac{1}{2} \leqq x \leqq \theta + \frac{1}{2}, \\ 1 & \theta + \frac{1}{2} < x, \end{cases}$$

and let R denote the family of all such distribution functions for $-\infty < \theta < \infty$. It is desired to test the hypothesis that $x_1, x_2, \cdots, x_n$ are independently and identically distributed according to some member of R. We will be concerned with computations when this hypothesis is true (see Section 5 for remarks on power which apply also here), and denote by G the true member of R (i.e., the distribution function of X_1). G_n^* is as defined in Section 1. Let

$$D_n^+ = \sup_x (G_n^*(x) - G(x)), \qquad D_n^- = \sup_x (G(x) - G_n^*(x)).$$

In the present example the minimum distance criterion is easily seen to be

$$\delta(G_n^*, R) = \inf_\theta \sup_x |G_n^*(x) - F(x;\theta)| = \tfrac{1}{2}(D_n^+ + D_n^-). \tag{4.1}$$

The joint limiting distribution of $\sqrt{n}D_n^+$ and $\sqrt{n}D_n^-$ (as $n \to \infty$) is given by Doob ([13], p. 403) for $x,y > 0$ as

$$\begin{aligned} &\lim_{n\to\infty} P\{\sqrt{n}D_n^- \leqq x, \sqrt{n}D_n^+ \leqq y\} \\ &= 1 - \sum_{m=1}^{\infty} \{e^{-2(mx+(m-1)y)^2} + e^{-2(my+(m-1)x)^2} - 2e^{-2m^2(x+y)^2}\} \qquad (4.2) \\ &= G(x, y), \quad \text{say.} \end{aligned}$$

Let $g(x, y) = \partial^2 G(x, y)/\partial x \partial y$. The series in (4.2) is uniformly and absolutely convergent and differentiable with respect to x and y outside of any neighborhood of the origin. Using the fact that the mixed derivative of the expression $e^{-2(ax+by)^2}$ is $4ab(u^2 - 1)e^{-u^2/2}$ where $u = (2ax + 2by)^2$, we obtain for the mth term of the series for $\int_0^z g(y, z - y)\, dy$ the expression

$$4m(m-1)[H(2mz) - H(2(m-1)z)] + 4m(4m^2z^2 - 1)H(2mz), \tag{4.3}$$
$$H(x) = \int_x^\infty (u^2 - 1)e^{-u^2/2}\, du = xe^{-x^2/2}.$$

The density function corresponding to the limiting distribution function of $B_n = \sqrt{n}(D_n^+ + D_n^-)$ is a sum of expressions given in (4.3). For $\sqrt{n}\delta(G_n^*, R) = \frac{1}{2}B_n = U$, say, the expression corresponding to (4.3) is

$$8m(m-1)[H(4mu) - H(4(m-1)u)] + 8m(16m^2u^2 - 1)H(4mu). \tag{4.4}$$

The series is absolutely convergent and in the sum of (4.4) from 1 to ∞, the coefficient of the expression $H(4ku)$ for $k \geqq 0$ is $8k(16k^2u^2 - 3)$, so that the density function of the limiting cumulative distribution function of U is

$$32u \sum_{m=1}^{\infty} m^2(16m^2u^2 - 3)e^{-8m^2u^2}. \tag{4.5}$$

Outside any given neighborhood of the origin, all terms of (4.5) are positive except for a finite number. Thus, integrating (4.5) we obtain for $u > 0$

$$\lim_{n\to\infty} P\{\sqrt{n}\delta(G_n^*, R) \leqq u\} = 1 - \sum_{m=1}^{\infty} (32m^2u^2 - 2)e^{-8m^2u^2}. \tag{4.6}$$

It is also of interest that in this example we can compute the limiting distribution of the minimum distance *estimator* of θ, namely, the random variable $T_n = T_n(x_1, \cdots, x_n)$ defined by

$$\delta(G_n^*(y), F(y; T_n)) = \delta(G_n^*, R), \tag{4.7}$$

which is satisfied by $T_n - \theta = \frac{1}{2}(D_n^- - D_n^+)$ when θ is the true parameter value. An analysis similar to that given above shows that, for $t > 0$,

$$\lim_{n\to\infty} P\{\sqrt{n}\,|T_n - \theta| \leqq t\} = 1 - 2\sum_{m=1}^{\infty} \frac{1}{4m^2 - 1} e^{-8m^2t^2}. \tag{4.8}$$

Of course, in this simple parametric example there are estimators of order $1/n$ in probability; in *estimation* problems, the minimum distance method is most useful in examples of a more nonparametric nature, where it often yields consistent estimators when other methods do not.

4.2. It is interesting to note that in the example of Section 4.1 and other similar cases it is simple to design along the lines of Section 2 a test of whether or not the unknown d.f. belongs to the specified class, where the limiting distribution of the test criterion when the null hypothesis is true is already known. These are the so-called "irregular" cases of estimation where an estimator of the unknown parameter(s) indexing the class exists whose deviation from the true parameter value is of lower order in probability than the usual $1/\sqrt{n}$ encountered in "regular" cases; for example, of order $1/n$ for the rectangular distribution with unknown location and/or range, or for the exponential distribution with known scale but unknown location parameter. For the sake of definiteness, we fix our attention on a Kolmogoroff-type criterion for the latter example, although the result applies equally to other distributions and other criteria (ω^2-type tests with different "weight"-functions, etc). Thus, the problem is to test, on the basis of n observations, whether or not the true d.f. is of the form

$$0, \quad x < \theta, \qquad 1 - e^{-(x-\theta)}, \quad x \geqq \theta,$$

for some real θ. Let $T_n = \min(x_1, \cdots, x_n)$. Then $P_\theta\{\lim_{n\to\infty}\sqrt{n}(T_n - \theta) = 0\} = 1$. Hence, if we compare the sample d.f. with the exponential c.d.f. as estimated by using T_n, by computing

$$B_n = \sqrt{n} \sup_{0<r<1} \left| F_n\left(T_n + \log \frac{1}{1-r}\right) - r \right|,$$

we may conclude that, when the null hypothesis is true, B_n has the same limiting distribution as the Kolmogoroff statistic.

Remarks on power like those of Section 5 apply also to the present case. The present remarks may also be modified to apply to situations where one but not all parameters have irregular estimators, for example, for the case of the exponential distribution with unknown location *and* scale. For the case of the rectangular distribution with unknown range studied in Section 4.1, it seems intuitively reasonable that a test constructed in the manner of the present section may be more powerful than the one considered there.

5. Asymptotic power of the tests of normality. The results of this section are carried out for the tests of normality mentioned in Section 2, but the remarks below concerning v_n may also be carried through for the minimum distance test, the test of Section 3, and in many other examples, and the remarks concerning v_n and w_n may be extended to many other "distance" criteria.

First we consider the test of size α based on v_n. The critical region is of the form $\{v_n > b(\alpha)/\sqrt{n}\}$, where $b(\alpha)$ is a constant, except for terms of lower order in n. Suppose $G(x) \equiv R_0(x)$, and that $\delta(R_0, N^{**}) = d/\sqrt{n}$. From the theorem of Kolmogoroff [4] we have that $\delta(R_0, G_n^*)$ is of the order $1/\sqrt{n}$ in probability (uniformly in R_0). Hence, for $0 < \beta < 1$ there is a number $d^* = d^*(\alpha, \beta)$ such that $d > d^*$ implies that $\delta(G_n^*, N^{**}) > b(\alpha)/\sqrt{n}$ with probability $\geqq 1 - \beta$. From the definition of $\delta(G_n^*, N^{**})$ we have

$$\delta(G_n^*, N(y \mid \bar{x}, s^2)) \geqq \delta(G_n^*, N^{**}).$$

Thus, if we are using the test of size α, the power is at least $1 - \beta$ for any alternative R_0 whose distance from N^{**} is $\geqq d^*(\alpha, \beta)/\sqrt{n}$, and the power of the test at R_0 approaches one as $\sqrt{n}\delta(R_0, N^{**})$ increases indefinitely. This is a result of the same order as obtains for testing goodness of fit of a simple hypothesis by use of Kolmogoroff's distribution (Section 1).

We now consider the "ω^2-type" test criterion w_n.[5] We consider the function

$$\gamma(F, H) = \left\{ \int_{-\infty}^{\infty} [F(x) - H(x)]^2 \, d\left(\frac{F(x) + H(x)}{2} \right) \right\}^{1/2} \tag{5.1}$$

as a possible measure of discrepancy between two d.f.'s. This measure has been used by Lehmann [14] and others. (We remark that γ is not a metric, since it does not satisfy the triangle inequality. Another undesirable property of γ is that the discrepancy between the d.f.'s of two random variables X and Y may not be the same as that between the d.f.'s of $-X$ and $-Y$. Also, the failure of the formula for integration by parts in expressions like (5.1) necessitates slight complications, for example, in the second following paragraph. Neither of the last two difficulties is present if both d.f.'s are continuous. If the d.f.'s have jumps which are nowhere dense, one could eliminate these last difficulties by redefining γ, for example, by replacing each jump by a constant density over an interval of width ϵ about the jump and letting $\epsilon \to 0$ after integrating. The development which follows would not be materially altered by such a change in the definition of γ.)

If F and H are continuous,

$$\int_{-\infty}^{\infty} [F(x) - H(x)]^2 \, d[F(x) - H(x)] = 0,$$

and the integration in (5.1) may be carried out with respect to either F or H instead of $\frac{1}{2}(F + H)$. Hence, if $G(x) \equiv R_0(x)$ is continuous,

$$\begin{aligned} w_n^{1/2} &= \left\{ \int [G_n^*(x) - N(x \mid \bar{x}, s^2)]^2 \, d_x N(x \mid \bar{x}, s^2) \right\}^{1/2} \\ &\geqq \left\{ \int [R_0(x) - N(x \mid \bar{x}, s^2)]^2 \, d_x N(x \mid \bar{x}, s^2) \right\}^{1/2} \\ &\quad - \left\{ \int [R_0(x) - G_n^*(x)]^2 \, d_x N(x \mid \bar{x}, s^2) \right\}^{1/2} \\ &\geqq \gamma(R_0, N^{**}) - \delta(R_0, G_n^*), \end{aligned} \tag{5.2}$$

where

$$\gamma(R_0, N^{**}) = \inf_{N \varepsilon N^{**}} \gamma(R_0, N^{**}).$$

Now, $\delta(R_0, G_n^*)$ is of order $1/\sqrt{n}$ in probability (uniformly in R_0) and the critical region based on w_n is of the form $\{\sqrt{w_n} > c(\alpha)/\sqrt{n}\}$, where $c(\alpha)$ is constant except for terms of lower order in n. Hence, using (5.2), an argument like that

[5] In an unpublished manuscript, T. W. Anderson considers similar criteria for testing a simple hypothesis, and obtains similar results on asymptotic power. See also his abstract (*Ann. Math. Stat.*, Vol. 25 (1954), p. 174).

of the previous paragraph shows that there is a value $d'(\alpha, \beta)$ such that the test of size α has power $\geqq 1 - \beta$ for any continuous alternative R_0 for which $\gamma(R_0, N^{**}) \geqq d'(\alpha, \beta)/\sqrt{n}$.

We now consider what happens if R_0 is not continuous. It is easy to show, by consideration of the contribution to the integral of (5.1) at discontinuities of H, that, if F is continuous,

$$\int (F - H)^2 \, dF \geqq \tfrac{1}{4} \int (F - H)^2 \, dH,$$

so that

$$\int (F - H)^2 \, dF \geqq \tfrac{2}{5} \int (F - H)^2 \, d\left(\frac{F + H}{2}\right).$$

Hence, if R_0 is not continuous, the argument of the previous paragraph need only be altered by inserting the factor $\sqrt{2/5}$ before γ in the last expression of (5.2). The ensuing discussion of power then proceeds as before.

It is interesting to compare the measurements of distance δ and γ. Clearly, $\gamma(F, G) \leqq \delta(F, G)$ for all F,G. On the other hand, if there is a value x_0 for which $F(x_0) - G(x_0) = \delta$, it is clear from the monotonicity of F and G, using (5.1), that $[\gamma(F, G)]^2 \geqq \frac{1}{6}[\delta(F, G)]^3$. Thus, we have

$$\sqrt{1/6}\, \delta^{3/2} \leqq \gamma \leqq \delta, \tag{5.3}$$

where both equalities are attainable. We conclude that, whereas for any given $\beta > 0$, the power exceeds $1 - \beta$ for alternatives whose distance from N^{**} is of order $1/\sqrt{n}$ either according to δ for the test based on v_n or according to γ for the test based on w_n, the distance in terms of δ must be of order $1/\sqrt[3]{n}$ (in the worst case) to insure this for the latter test. For the former test, γ-distance of order $1/\sqrt{n}$ suffices.

We next verify the property of the χ^2-test relative to γ which was stated at the end of the third paragraph of the introduction. For brevity we shall use the notation of [3] without redefining symbols here; the reader may also refer to [3] for details of the argument which we omit. Suppose then that we are testing the hypothesis $G(x) = R(x) \equiv x$ for $0 \leqq x \leqq 1$ by means of the χ^2-test based on N observations and k_N intervals of equal length on the unit interval. If now

$$G(x) = F_{k_N}(x) \equiv \sum_{i=1}^{k_N} \frac{1}{k_N} \psi_x \left(\frac{2i - 1}{2k_N}\right)$$

(see Section 1 for the definition of ψ_x), then $\gamma(F_{k_N}, R) = 1/k_N \sqrt{6}$. Since F_{k_N} assigns the same probability as R to each of the k_N intervals, the power of the test against the alternative F_{k_N} is just the size of the test (assumed to be $< \frac{1}{2}$). We conclude that if the test gives power $\geqq \frac{1}{2}$ for all alternatives R^* satisfying $\gamma(R^*, R) \geqq \Gamma_N$, then we must have

$$k_N > 1/\Gamma_N\sqrt{6}. \tag{5.4}$$

(This is not the best possible inequality, but it suffices for our proof.) Consider now the distribution function

$$H_a(x) = \begin{cases} (1 + 2a)x, & 0 \leqq x \leqq \frac{1}{2}, \\ 2a + (1 - 2a)x, & \frac{1}{2} \leqq x \leqq 1, \end{cases} \tag{5.5}$$

where $0 < a < \frac{1}{2}$.

A simple computation shows that $\gamma(H_a, R) = a/\sqrt{3} = \gamma_a$, say, and that (assuming for simplicity that k_N is even) when $G = H_a$ we have $\sum p_i^2 = (1 + 4a^2)/k_N = (1 + 12\gamma_a^2)/k_N$. Hence, the function ψ of [3] is given, when $G = H_a$, by

$$\sigma'\psi(k_N) = 12(N - 1)\,\gamma_a^2 - C\sqrt{2(k_N - 1)}. \tag{5.6}$$

In order that the χ^2-test based on k_N intervals have power $\geqq \frac{1}{2}$ for all alternatives R^* with $\gamma(R^*, R) \geqq \Gamma_N$, it is necessary that the expression (5.6) be $\geqq 0$ asymptotically when we put $\gamma_a = \Gamma_N$, and that (5.4) be satisfied. We thus obtain, when N is large,

$$\Gamma_N \geqq C'N^{-2/5} \tag{5.7}$$

where C' is a positive constant. From the result of [3] and the fact that $\gamma \leqq \delta$, we see that the reverse inequality to (5.7) is (for a different C') also true. Thus, $N^{-2/5}$ is indeed the smallest order of Γ_N which will give appreciable power for all alternatives R^* with $\delta(R, R^*) \geqq \Gamma_N$.

We shall now summarize the results proved thus far in this section. It is not known how the power function of the χ^2-test for composite hypotheses behaves, but it is plausible that the power function when testing a composite hypothesis by means of the χ^2-test (in any of its variations) is no better (in the sense we have used in measuring the goodness of a power function) than when testing a simple hypothesis. We have shown that *if* Γ *and* Δ *are small and the* χ^2*-test of size* $< \frac{1}{2}$ *of a simple hypothesis* $G = R$ *requires* N *observations to insure power* $\geqq \frac{1}{2}$ *at alternatives* R^* *for which* $\gamma(R^*, R) = \Gamma$ *(or* $\delta(R^*, R) = \Delta$*), so that* $N = C_1\Gamma^{-5/2}$ *(or* $N = C_2\Delta^{-5/2}$*), then the numbers of observations required by the Kolmogoroff and* ω^2 *tests to achieve the same minimum power at* Γ *(or* Δ*) are at most*

$$\begin{aligned} &\textit{Kolmogoroff}: \quad n = C_3N^{4/5} = C_4\Gamma^{-2} \qquad (n = C_5N^{4/5} = C_6\Delta^{-2}) \\ &\omega^2: \quad n = C_7N^{4/5} = C_8\Gamma^{-2} \qquad (n = C_3N^{6/5} = C_{10}\Delta^{-3}). \end{aligned}$$

The numbers of observations n *required by the tests based on* v_n *and* w_n *in testing composite hypotheses about parametric families are the same functions of* Γ *or* Δ *as for the Kolmogoroff and* ω^2 *tests of simple hypotheses.* The test based on v_n may thus be expected to be superior to the χ^2-test in the sense of both γ and δ, and that based on w_n may be expected to be superior in the sense of γ, at least for large N.

It might be supposed that the χ^2-test would show up to better advantage relative to the test based on v_n in terms of a metric like

$$\eta(R_0, N^{**}) = \inf_{N \epsilon N^{**}} \int d\, |R_0 - N|.$$

However, this is not so; for fixed n, even if $\eta(R_0, N^{**})$ is near its maximum of 2, neither the best χ^2-test of [3] nor any of the other tests we have mentioned need have appreciable minimum power ($\geqq \frac{1}{2}$), and $\delta(R_0, N^{**})$ and $\gamma(R_0, N^{**})$ can be arbitrarily small. In fact, it is easy to see that no test can have the infimum of its power function over all alternatives R_0 with $\eta(R_0, N^{**}) = C > 0$ greater than the size of the test. (If N^{**} were a simple hypothesis, this would still be true.) In order better to compare the behavior of tests in terms of the metric η, we might therefore restrict our consideration to alternatives R_0 belonging to some regular class, for example, the class of d.f.'s with densities which cross that of each member of N^{**} at most M times. Under such a comparison the test based on v_n may be shown to be superior to that based on the χ^2-test, in the same sense as under our previous comparison.

The discussion of this section suggests very strongly that there is a "natural" distance with respect to which the power characteristics of a particular "distance" test criterion should be measured, and that a comparison of the power of such tests in terms of their own and other distances indicates that tests corresponding to strong metrics have the best global power characteristics. It is hoped to investigate this idea further.

6. Numerical results. In this section we list some pertinent experimental and computational results. The main purpose of the sampling experiments was to obtain estimates of the distributions of nw_n and $\sqrt{n}v_n$ which may be used in applications to test normality. As a check on the sampling experiments the same data were used to compute experimentally the d.f. of $\bar{x}$ and the d.f. of $\sqrt{n}D_n$ with n large, D_n being defined by

$$D_n = \delta(L(x), L_n^*(x))$$

where $L_n^*(x)$ is the empiric d.f. of n independent chance variables with the d.f. $L(x)$. Also, as another check on the experimentally obtained d.f. of nw_n, an approximation to the d.f. of W was computed, using (2.9).

Define

$$\omega_n^2 = \int (L(x) - L_n^*(x))^2\, dL(x).$$

The sampling experiments were conducted using 400 samples of size $n = 100$ and 400 samples of size $n = 25$, of random standard (mean zero, variance one) normal deviates from the well-known Rand Corporation series. From each sample the values of the sample mean ($\bar{x}$), sample variance (s^2), v_n, w_n, D_n, and ω_n^2, were computed. Thus, there were obtained 400 "observations" on $\sqrt{n}v_n$ and nw_n for $n = 25$ and $n = 100$, and the sample d.f.'s based on these observations serve as estimates of the d.f.'s of $\sqrt{n}v_n$ and nw_n. The known d.f. of $\bar{x}$

and the known d.f. of $\sqrt{n}D_n$ were compared with the experimentally obtained d.f.'s of $\bar{x}$ and $\sqrt{n}D_n$ as a check on the experimentally obtained d.f.'s of $\sqrt{n}v_n$ and nw_n.

It was found that the experimentally obtained d.f. of $\bar{x}$ agreed well with the known d.f., and the experimentally obtained d.f. of $\sqrt{n}D_n$ agreed well with the d.f. of $\sqrt{n}D_n$ as tabulated in [15]. In each case, the maximum difference between the experimentally obtained and the known actual d.f. was found to be small according to the tables [15]; this is the basis for saying the agreement was good. The agreement between the limiting d.f. of $n\omega_n^2$ (tabulated in [16]) and the experimentally obtained d.f. was found to be fairly close for $n = 100$ but not close for $n = 25$; the upper tails of the distributions seem to be the parts which come into agreement most rapidly with n. The distributions of $\sqrt{n}v_n$ and nw_n are concentrated closer to the origin than those of $\sqrt{n}D_n$ and $n\omega_n^2$, respectively. This is not entirely surprising since the covariance function (2.1) is smaller than that for the process corresponding to D_n and ω_n^2 (namely, min $(s, t) - st$). Scatter diagrams of the sample values of $(\sqrt{n}D_n, n\omega_n^2)$ indicate that the regression of $n\omega_n^2$ on $\sqrt{n}D_n$ is roughly parabolic with the conditional variance of $n\omega_n^2$ increasing with the value of $\sqrt{n}D_n$; this is not entirely surprising in view of the way in which ω_n^2 and D_n are computed from a sample. Although v_n and w_n are correlated with D_n and ω_n^2, they seem to be more strongly related to $\bar{x}$.

In Tables I and II, respectively, are given the estimates of the d.f.'s of $\sqrt{n}v_n$ and nw_n; commonly used percentage points are listed for convenience in Table III.

We now turn to the limiting distribution of nw_n. The first eight zeros of $D(\mu)$ are alternately zeros of $D_1(\mu)$ and $D_2(\mu)$. Their values, obtained by numerical computation of these functions for various values of μ, are 7.38, 8.62, 13.66, 15.14, 19.91, 21.52, 26.16, 27.87. The corresponding values of λ_j for $1 \leqq j \leqq 8$ are .01836, .01346, .00536, 00436, .00252, .00216, .00146, .00129. It is to be noted that λ_{2j-1} and λ_{2j} are both approximately c/j^2 (the corresponding property for the eigenvalues arising in the computation of $n\omega_n^2$ is that $\lambda_j = 1/\pi^2 j^2$). This (inferred) speed of convergence of the λ_j to 0 implies that the distribution of $W^* = \sum_1^4 \lambda_j \gamma_j$ should be a fairly good approximation to that of $W = \sum_1^\infty \lambda_j \gamma_j$,

TABLE I

Estimate of $Q_n(x) = P\{\sqrt{n}v_n \leqq x\}$.

x	$Q_{25}(x)$	$Q_{100}(x)$	x	$Q_{25}(x)$	$Q_{100}(x)$
.30	0	0	.75	.8100	.8425
.35	.0125	.0025	.80	.8600	.9025
.40	.0550	.0250	.85	.9125	.9350
.45	.1500	.0800	.90	.9500	.9600
.50	.2525	.1975	.95	.9725	.9775
.55	.3675	.3300	1.00	.9850	.9900
.60	.5025	.4775	1.05	.9950	.9925
.65	.6400	.6750	1.10	1.0000	.9975
.70	.7225	.7325	1.15	1.0000	1.0000

TABLE II

Estimate of $R_n(x) = P\{nw_n \leqq x\}$ and distribution function $H(x) = P\{W^* \leqq x\}$.

x	$R_{25}(x)$	$R_{100}(x)$	$H(x)$
.01	0	0	.108
.02	.0375	.0225	.297
.03	.1325	.1400	.471
.04	.2125	.3175	.609
.05	.4300	.4775	.712
.06	.5825	.6250	.787
.07	.6625	.7175	.843
.08	.7250	.7850	.883
.09	.7925	.8600	.914
.10	.8400	.8900	.937
.11	.8975	.9175	.953
.12	.9150	.9350	.966
.13	.9425	.9700	.975
.14	.9525	.9825	.982
.15	.9650	.9850	.987
.16	.9700	.9850	.997
.17	.9800	.9875	
.18	.9875	.9875	
.19	.9875	.9900	
.20	.9900	.9950	
.21	.9950	.9950	
.22	.9975	.9975	
.23	1.0000	1.0000	

TABLE III

Estimates of common percentage points of Q_n and R_n.

p	$Q_{25}^{-1}(p)$	$Q_{100}^{-1}(p)$	$R_{25}^{-1}(p)$	$R_{100}^{-1}(p)$
.20	.7435	.729	.0909	.0824
.10	.8225	.797	.1145	.1019
.05	.8980	.878	.1352	.1240
.02	.9685	.954	.1671	.1386
.01	1.0145	.989	.1957	.1859
.005	1.0465	1.062	.2053	.1957

and the distribution of W^* was therefore computed, using the method of [17] (here $\lambda_j\gamma_j$ is the R_j of (2.9)). The results are given in the last column of Table II.

From the fact noted above regarding the speed of convergence of the d.f. of $n\omega_n^2$, and the fact that Table II indicates (in the difference $R_{100}(x) - R_{25}(x)$) a much slower approach to the limiting distribution for $R_n(x)$ than that noted in our experiment for the ω_n^2-distribution, we would expect that the d.f. $H(x)$ of W^* should lie above the estimates $R_n(x)$ of Table II, being close to $R_{100}(x)$ only in the upper tail. This is what the last column of Table II actually shows, and an idea of how good the agreement is in the tail may be obtained by computing $M(x) = 20[H(x)]^{-1/2}[H(x) - R_{100}(x)]$; for $x = .13, .14, .15, .20$, one obtains $M(x) < 1$, which indicates very good agreement. Thus, in applications where n is large, it seems reasonable to use the last column of Table II, especially in the upper tail (which is the region that matters for statistical tests).

We are indebted to Prof. R. J. Walker and to Mr. R. C. Lesser of the Cornell Computing Center for carrying out the sampling and the computation of the λ_j's. The experimental data are available in the files of the Cornell Computing Center.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] H. Chernoff and E. L. Lehmann, "The use of maximum likelihood estimates in χ^2 tests for goodness of fit," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 579–586.
[3] 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.
[4] A. N. Kolmogoroff, "Sulla determinazione empirica di una leggi di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 83–91.
[5] J. Wolfowitz, "Consistent estimators of the parameters of a linear structural relation," *Skand. Aktuarietids.*, Vol. 35 (1952), pp. 132–151.
[6] J. Wolfowitz, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math., Tokyo*, Vol. 5 (1953), pp. 9–23.
[7] J. Wolfowitz, "Estimation by the minimum distance method in nonparametric stochastic difference equations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 203–217.
[8] J. Wolfowitz,[6] "Estimation of the components of stochastic structures," *Proc. Nat. Acad. Sci., U.S.A.*, Vol. 40 (1954), pp. 602–606.
[9] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217–226.
[10] M. Kac, "On some connections between probability theory and differential and integral equations," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 180–215.
[11] M. Kac and A. J. F. Siegert, "On the theory of noise in radio receivers with square law detectors," *J. Applied Physics*, Vol. 18 (1947), pp. 383–397.
[12] M. Kac and A. J. F. Siegert, "An explicit representation of stationary Gaussian processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 438–442.
[13] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnoff theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.
[14] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.
[15] Z. W. Birnbaum, "Numerical tabulation of the distribution of Kolmogoroff's statistic for finite sample size," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 425–441.
[16] 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.
[17] H. Robbins and E. J. G. Pitman, "Application of the method of mixtures of quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.
[18] D. A. Darling, "The Cramér-Smirnov test in the parametric case" (abstract), *Ann. Math. Stat.*, Vol. 24 (1953), p. 493.

[6] The sentence at the bottom of page 602 of [8] which reads, "It is well known that $F(z)$ determines α . . ." should read, "It is well known that the distribution of (x_i, x_{i+1}) determines α. . . ."