THE KOLMOGOROV-SMIRNOV, CRAMÉR-VON MISES TESTS

D. A. DARLING

University of Chicago, University of Michigan

1. Preface. This is an expository paper giving an account of the “goodness of fit” test and the “two sample” test based on the empirical distribution function—tests which were initiated by the four authors cited in the title. An attempt is made here to give a fairly complete coverage of the history, development, present status, and outstanding current problems related to these topics.

The reader is advised that the relative amount of space and emphasis allotted to the various phases of the subject does not reflect necessarily their intrinsic merit and importance, but rather the author’s personal interest and familiarity. Also, for the sake of uniformity the notation of many of the writers quoted has been altered so that when referring to the original papers it will be necessary to check their nomenclature.

2. The empirical distribution function and the tests. Let $X_1, X_2, \cdots, X_n$ be independent random variables (observations) each having the same distribution function $U(x) = \Pr\{X_i < x\}$ and put

$$
\epsilon(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. 
\end{cases}
$$

Then the (random) function

$$(2.1) \quad F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \epsilon(x - X_j)$$

is called the empirical distribution function of the data. Clearly $F_n(x)$ is the proportion of the $X_i, i = 1, 2, \cdots, n$, which are less than $x$.

It is easy to calculate the first and second order moments

$$
E(F_n(x)) = U(x),
$$

$$
\text{Cov}(F_n(x), F_n(y)) = E(F_n(x)F_n(y)) - U(x)U(y)
$$

$$
= \frac{1}{n} c(U(x), U(y)),
$$

where

$$
(2.3) \quad c(s, t) = \min(s, t) - st = \begin{cases} s(1 - t) & s \leq t \\ t(1 - s) & s \geq t, \end{cases}
$$

$$
0 \leq s, t \leq 1.
$$

We quote a few classical consequences of the definition (2.2):

Received April 10, 1957. Special Invited Paper read before the Institute of Mathematical Statistics, August 21, 1956.

1 Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Strong law of large numbers,

\[(2.4) \quad F_n(x) \rightarrow U(x) \text{ with probability 1 for each } x.\]

Law of the iterated logarithm,

\[
\lim_{n \to \infty} \sqrt{n} \left| \frac{F_n(x) - U(x)}{\sqrt{2 \log \log n}} \right| = \sqrt{U(x)(1 - U(x))}
\]

with probability 1 for each \(x\).

Multidimensional central limit theorem,

\[(2.5) \quad \left\{ \sqrt{n}(F_n(x_i) - U(x_i)) \right\} \quad i = 1, 2, \ldots, k
\]

has an asymptotic \((n \to \infty, k \text{ fixed})\) \(k\)-dimensional normal distribution, means 0 and covariance \(c(U(x_i), U(x_i))\) with \(c(s, t)\) given by \((2.3)\).

Cantelli-Glivenko lemma ([29], [88]),

\[(2.6) \quad \sup_{-\infty < x < \infty} |F_n(x) - U(x)| \rightarrow 0 \quad \text{with probability 1.}\]

The last result (2.6), which considerably generalizes (2.4), is itself capable of further extensions. Fortet and Mourier [22] have shown \(1/n \sum_{i=1}^{n} f(X_i) \rightarrow E(f(X_i))\) uniformly with respect to an inclusive family of functions \(\{f\}\) with probability 1. Then (2.6) follows on considering the family \(f_t(x) = \epsilon(\xi - x), \quad -\infty < \xi < \infty, \epsilon(x) \) given by (2.1). Steinhaus [79] showed that the mutual independence of the \(X_i\) could be relaxed to pairwise independence and (2.6) holds. See also Wolfowitz [89].

The following two statistical problems motivate the analysis:

(a) Goodness-of-fit problem. Let the \(X_i\) be the random variables described in the first sentence of this section. The goodness-of-fit problem is to devise a test of the hypothesis

\[(2.7) \quad H_0: U(x) = F(x),\]

where \(F(x)\) is a given continuous distribution function. This is one of the classical problems of statistics for which K. Pearson developed the well known \(\chi^2\) test—cf. Cochran [15].

(b) Two-sample problem. Let the \(X_i\) be as above with \(U(x)\) known to be continuous and let \(Y_1, Y_2, \ldots, Y_m\) be independent random variables with the common continuous distribution \(V(x) = \Pr\{Y_i < x\},\) all \(n + m\) of these random variables being mutually independent. The two-sample problem is to devise a test of the hypothesis

\[(2.8) \quad H'_0: U(x) = V(x).\]

This is also an old, celebrated problem—cf. [53].

Roughly speaking, the tests proposed here of the null hypotheses \(H_0, H'_0\) are based on certain distribution analogues of the Cantelli-Glivenko lemma (2.6) in the same way that the central limit theorem is a distribution analogue of the law of large numbers.
3. The Cramér-Smirnov tests. In 1928 Cramér [13] suggested for $H_0$ the following test criterion:

$$
\int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dK(x),
$$

where $K(x)$ is suitable nondecreasing weight function. $H_0$ given by (2.7) is to be rejected if this expression is too large. Von Mises [83] independently made an equivalent suggestion and developed a few properties of the test.

Smirnov [71], [72] gave the modification

$$
W_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \psi(F(x)) dF(x).
$$

(3.1)

where $\psi(t)$, $0 \leq t \leq 1$, is a nonnegative weight function to be selected presumably on the grounds of certain power requirements. The test based on $W_n^2$ is distribution free—this is readily seen from (2.2) for, if (2.7) is true, we have, recalling the continuity of $F(x)$,

$$
W_n^2 = n \int_{-\infty}^{\infty} \left( \frac{1}{n} \sum_{j=1}^{n} \epsilon(x - X_j) - F(x) \right)^2 \psi(F(x)) dF(x)
$$

(3.2)

$$
= n \int_{0}^{1} \left( \frac{1}{n} \sum_{j=1}^{n} \epsilon(t - F(X_j)) - t \right)^2 \psi(t) dt,
$$

with probability 1, and since the $F(X_j)$ are independent and uniformly distributed over $(0, 1)$ the result follows.

Besides being distribution free the test is consistent (if $\psi > 0$) and requires no arbitrary grouping of the data—these three desirable properties are not shared by the $\chi^2$ test of $H_0$.

Smirnov’s basic result concerning the distribution of (3.1) if (2.7) is true is that

$$
\lim_{n \to \infty} E[e^{iW^2_n}] = (D(2i\xi))^{-1},
$$

(3.3)

where $D(\lambda)$ is the Fredholm determinant associated with the kernel

$$
k(s, t) = \psi(s)\psi(t)c(s, t), \quad 0 \leq s, t \leq 1,
$$

(3.4)

c(s, t) being given by (2.3).

Smirnov found the distribution function corresponding to this limiting characteristic function in the following form [72]:

$$
\lim_{n \to \infty} \Pr \{W_n^2 < x\} = G(x)
$$

(3.5)

$$
= 1 - \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k-1} \int_{\lambda_{k-1}}^{\lambda_k} \frac{e^{-x^2/2}}{\sqrt{-y^2D(y)}} dy,
$$

where $^2\lambda_j, j = 1, 2, \cdots$ are the (simple) zeros of $D(\lambda)$. Later Smirnov [77] gave simpler proofs of these results.

$^2$ The factor $(-1)^{k-1}$ is missing throughout [72].
von Mises [84] deduced (3.3) and considered a number of extensive generalizations in the direction of nonidentically distributed $X_i$, and quadratic forms other than the mean square.

There now exist quite simple proofs of (3.3) resting on a reduction to a simple stochastic process, basically an idea of Doob [19] and Kac [42]. If we let $F_n^*(t)$ be the empirical distribution function based on $F(X_1), F(X_2), \ldots, F(X_n)$, then from (3.2) we deduce that if

$$x_n(t) = \sqrt{n}(F_n^*(t) - t), \quad 0 \leq t \leq 1,$$

then

$$W_n^2 = \int_0^1 x_n^2(t)\psi(t)\,dt.$$

From the fact that (2.5) has a limiting multidimensional normal distribution, $x_n(t)$ converges in distribution to a Gaussian process $x(t)$ with mean 0 and covariance $c(s, t)$ given by (2.3). If now $Q(f)$ is a "reasonable" functional to the reals it is natural to conjecture that

$$(3.6) \quad \lim_{n \to \infty} \Pr \{Q(x_n(t)) < x\} = \Pr \{Q(x(t)) < x\}.$$

This being true for $Q(f) = \int_a^b f^2(t)\psi(t)\,dt$, Smirnov's result (3.3) follows immediately from a theorem of Kac and Siegert [41].

Kac [43] justified (3.6) for this $Q$ when $\psi = 1$, and Donsker [18] proved (3.6) for a wide class of $Q$. There now exist very extensive generalizations of this so-called invariance principle ([66], [57]).

The essential result of the line of attack in [41] is that for $z(t), a \leq t \leq b$, Gaussian,

$$E(z(t)) = 0$$

$$E(z(t)z(s)) = \Gamma(s, t),$$

the distribution of $W_n^2 = \int_a^b z^2(t)\,dt$ is that of

$$(3.7) \quad \sum_{j=1}^\infty \frac{G_j^2}{\lambda_j},$$

where $G_1, G_2, \cdots$ are independent, normally distributed, means 0, variances 1, and $\lambda_1, \lambda_2, \cdots$ are the eigenvalues of the kernel $\Gamma(s, t)$—i.e., the zeros of the Fredholm determinant $D(\lambda)$ of the integral equation

$$f(t) = \lambda \int_a^b \Gamma(t, s)f(s)\,ds.$$

For the kernel (3.4), this result yields (3.3) immediately.

A systematic study of the limiting distribution of (3.1) was made in [1], and it turns out that $D(\lambda)$ can be determined from an initial value equation. If $\psi(t)$ is continuous in $0 \leq t \leq 1$, then

$$\varphi''(t) + \lambda \psi(t)\varphi(t) = 0, \quad \varphi(0) = 0, \quad \varphi'(0) = 1,$$
has a unique solution $\phi_\lambda(t)$ and the Fredholm determinant $D(\lambda)$ of (3.4) is

$$D(\lambda) = \frac{\phi_\lambda(1)}{\phi_0(1)}.$$  

For the important case $\psi = 1$, the limiting characteristic function (3.3) is

$$\left(\sqrt{2i\xi} \csc \sqrt{2i\xi}\right)^1.$$  

This was inverted in [1] in a form different from (3.5) and a table given of the limiting distribution of $W_n^2$. For the statistically appealing weight function

$$\psi(t) = \frac{1}{t(1-t)},$$

the limiting characteristic function is $\sqrt{-2\pi i \xi \left[\cos\left(\frac{1}{2}\pi(1 + 8i\xi)^{1/2}\right)\right]}^{-1}$ which was also inverted [1] and a few significance points given [2].

There is no multivariate analogue to the $W_n^2$ test which is distribution free (unless the components are independent). There is, however, a transformation of a multivariate distribution to a uniform distribution over the unit cube due to Lévy, and Rosenblatt [68] suggested an analogue to $W_n^2$ for it and obtained [69] a few results for the corresponding limiting distribution.

For $H'_0$ of (2.8) a corresponding distribution free test exists—cf. Lehmann [53]. The natural analogue to (3.1) is

$$\frac{mn}{m+n} \int_{-\infty}^{\infty} (F_n(x) - G_n(x))^2 \psi \left(\frac{nF_n + mG_m}{m+n}\right) d \left(\frac{nF_n + mG_m}{m+n}\right),$$

where $F_n(x)$ and $G_n(x)$ are respectively the empirical distribution functions of the $X$'s and the $Y$'s. It is easy to prove when (4.4) below holds that this has the same limiting distribution (if $H'_0$ is true) as $W_n^2$ of (3.1)—cf. [69] for the case $\psi = 1$.

4. The Kolmogorov-Smirnov tests. In 1933 Kolmogorov [45] suggested a test of $H_0$ of (2.7) based on the statistic

$$K_n = \sqrt{n} \sup_{-\infty < x < \infty} \left| F_n(x) - F(x) \right|;$$

$H_0$ is to be rejected if $K_n$ is sufficiently large. The distribution of $K_n$ is independent of $F(x)$ if (2.7) is true (i.e., the test is distribution free) and denoting its distribution by $\Phi_n(x)$ Kolmogorov proved that

$$\lim \Pr \{K_n < x\} = \lim \Phi_n(x) = \Phi(x)$$

$$= \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 x^2}, \quad 0 < x < \infty.$$  

If $F(x)$ is not continuous, $\Pr\{K_n < x\} \geq \Phi_n(x)$, so the test could be used conservatively even if the $X$'s have not a continuous distribution. Smirnov [74] gave a simpler proof of (4.2) and also a distribution free test of $H'_0$. He proved that the random variable
\( D_{mn} = \sqrt{\frac{mn}{m+n}} \sup_{-\infty < z < \infty} |F_n(x) - G_m(x)|, \)

with distribution function \( \Phi_{m,n} \) had, if

\( 0 < a \leq \frac{m}{n} \leq b < \infty, \quad m \to \infty, \quad n \to \infty, \)

a limiting distribution \( \Phi \) given by (4.2).

For the corresponding one-sided tests define

\( K_+^n = \sqrt{n} \sup_{-\infty < z < \infty} (F_n(x) - F(x)), \)

\( D_{mn}^+ = \sqrt{\frac{mn}{m+n}} \sup (F_n(x) - G_m(x)), \)

\( D_{mn}^- = \sqrt{\frac{mn}{m+n}} \sup (G_m(x) - F_n(x)). \)

Smirnov ([74], [75]) gave limiting distributions of these random variables under condition (4.4)

\( \lim \Pr\{K_+^n < x\} = \lim \Pr\{D_{mn}^+ < x\} = 1 - e^{-2z^2}, \quad 0 \leq x < \infty, \)

\( \lim \Pr\{D_{mn}^+ < x, D_{mn}^- < y\} = \Phi(x, y) = 1 + \sum_{j=1}^{\infty} \{2e^{-2(j+1)x^2} - e^{-2(j+x+y)^2} - e^{-2(j+y+(j-1)x)^2}\}, \quad 0 \leq x, y < \infty. \)

The early work of Kolmogorov and Smirnov is summarized in [46] and [75]. A short table of the distribution \( \Phi \) of (4.2) was given in [74] and amplified in [76]. Corrections to the tables are in [50], [51], and extensive percentage points in [65].

Wald and Wolfowitz ([85], [86]), in connection with a problem of finding confidence limits for an unknown distribution function considered independently the distribution of \( K_n \) of (4.1), giving methods of calculating its distribution for finite \( n \). For elementary expository remarks and applications, cf. [39].

Feller [21] rederived (4.2). A strong counterpart of (4.2) was given by Chung [12] who proved that infinitely many inequalities

\( \sup_{-\infty < z < \infty} \sqrt{n} |F_n(x) - F(x)| > \lambda_n, \)

occur with probability zero or one according as

\( \sum \frac{\lambda_n^2}{n} e^{-2\lambda_n^2} \)

converges or diverges.
Doob [19] showed that $\Phi$ of (4.2) is given by

$$\Phi(x) = \Pr \{ \sup_{0 < t < 1} |x(t)| < x\},$$

(4.10)

where $x(t)$ is the Gaussian process of (3.6). Doob omitted the justification of (3.6) for $Q(f) = \sup |f|$, which was supplied by Donsker [18]. Doob observed that the Gaussian process $x(t)$ with mean 0 and covariance (2.3) was simply transformable to the Wiener process $w(t)$, $0 \leq t < \infty$, and that the probability (4.10) is a simple first passage probability for that process. Similarly for the limiting distributions of (4.3), (4.5), and (4.6).

Using this last observation a generalization of the $K_n$ test was proposed [1] as follows:

$$K_n^* = \sup_{-\infty < x < \infty} \sqrt{n} |F_n(x) - F(x)| \psi(F(x)),$$

(4.11)

where $\psi \geq 0$ is a preassigned weight function. The limiting distribution of $K_n^*$ can be obtained then as the solution to a boundary value problem associated with the simple diffusion equation. If $\psi(t) = (at + \beta)^{-1}$ in a piecewise way the classical methods give the limiting distribution in quadratures; [1].

These latter include the case of detecting discrepancies over a central portion of the interval ([1], [55], [56]) where

$$\psi_l(t) = \begin{cases} 
1 & a < t < b \\
0 & \text{otherwise}, 
\end{cases}$$

(4.12)

and over the tails $1 - \psi_l(t)$, [4], and the cases $\psi_1 = 1/t$, $\psi_2 = 1/(1 - t)$ for $t$ in a subinterval of $(0, 1)$; see [67], [11], and [54]—cf. also [27].

For $\sqrt{\psi}$ where $\psi$ is given by (3.8), the distribution of $K_n^*$ was given in [1]; and when $m = n \to \infty$, the limiting distribution of $D_{mn}$ of (4.3) has been treated [52] analogously with the weight function $\psi_0$ of (4.12).

Interest of late has been in calculating the distribution of these random variables for finite sample sizes (always under the assumption that $H_0$, $H'_0$ of (2.7) and (2.8) are true). In [85] a method of calculating the distribution of $K_n$ of (4.1) was given, applicable when $n$ is small. A series of recurrence relations were given in [45] for calculating the distribution of $K_n$, and it was suggested much later [5] that these may be amenable to high-speed calculation—the program was subsequently carried out ([58], [7]) giving tables of the distribution of (4.1). For $D_n^+$ similarly, cf. [80].

Birnbaum and Tingey [6] proved that for (4.5)

$$\Pr \{ K_n^+ > \epsilon \sqrt{n} \} = (1 - \epsilon)^n + \epsilon \sum_{1 \leq j \leq n(1 - \epsilon)} \binom{n}{j} \left( 1 - \epsilon - \frac{j}{n} \right)^{n-j} \left( \epsilon + \frac{j}{n} \right)^{j-1}.$$

Gnedenko and his students have recently studied systematically (4.3), (4.5), (4.6), and (4.7), mainly in the case of equal sample sizes $m = n$. We abbreviate in this case $D_{mn} = D_n$, $D_{mn}^+ = D_n^+$, $D_{mn}' = D_n'$. The distribution of $D_n$, $D_n^+$ and $D_n^-$ can be reduced to first passage problems associated with simple random walks.
Consider, e.g., the distribution of $D_n^+$. If the pooled sample of size $2n$, $X_1, X_2, \ldots, X_n, Y_1, \ldots, Y_n$, is arranged in increasing magnitude and we denote by $z_i, i = 1, 2, \ldots, 2n$ a random variable equal to $+1$ or $-1$ according as the $i$th member of it is an $X$ or $Y$ respectively, then if $H'_0$ of (2.8) is true and $S_j = z_1 + z_2 + \cdots + z_j, (S_0 = 0)$,

$$\Pr \{D_n^+ < x\} = \Pr \left\{ \max_{1 \leq j < 2n} S_j < x\sqrt{2n}\right\}.$$  

The set $S_0, S_1, \ldots, S_{2n} = 0$ form a Markov chain, and the probability in question is given by a simple reflection principle [3]. One obtains in fact

$$\Pr \{D_n^+ < x\} = 1 - \frac{\binom{2n}{n} + [nx\sqrt{2n}]}{\binom{2n}{n}}, \quad 0 \leq x \leq \sqrt{\frac{n}{2}},$$

and similar simple formulas for the distributions of $D_n$ and the joint distribution of $D_n^+$ and $D_n^-$ for finite $n$.

There exist many other results in this direction, too numerous to treat in detail; we mention several of the simpler in their limiting form:

$$\lim_{n \to \infty} \sqrt{2n} \Pr \left\{ D_n^+ + D_n^- = \sqrt{\frac{1}{2n} [z\sqrt{2n}]} \right\} = 8z \sum_{j=1}^{\infty} \left( 4j^2z^2 - 3j^2 \right) e^{-2j^2z^2}, \quad 0 < z < \infty,$$

cf. [35], [38];

$$\lim_{n \to \infty} \rho(D_n^+, D_n^-) = \frac{2\pi^2 - 3\pi - 12}{3(4 - \pi)} = -0.6547,$$

cf. [34];

$$\lim_{n \to \infty} \Pr \{F_n(x) > G_n(x) \text{ for all } x \text{ such that } \alpha < U(x) < \beta\} = \frac{1}{\pi} \sin^{-1} \sqrt{\frac{\alpha(1 - \beta)}{\beta(1 - \alpha)}},$$

cf. [37], [24], and [70]; where $\rho$ is the correlation coefficient and $U(x)$ the common distribution of $X_i$ and $Y_j$. Much of the work of Gnedenko and his coworkers is summarized in [37] and [38].

The random walk method of treating $D_n^+$, $D_n$ was employed independently in [20], and tables of the distribution of $D_n$ were constructed ([61], [62]), for finite $n$ using methods unrelated to the above.

For unequal sample sizes, the distributions of $D_{n,n,p}$, $D_{n,n,p}^+$, $D_{n,n,p}^-$, $p$ integral, can be again reduced to a random walk problem ([48], [49], [10]) of a somewhat more complex kind, but still amenable to the reflection principle.

The exact formulas lead to asymptotic expansions of $K_n^+$, $K_n$, $D_n^+$, $D_n$ of which (4.2), (4.8), (4.9) are the leading terms—but Smirnov’s original analysis required only (4.4) to hold for $D_{mn}^+$, $D_{mn}$ rather than equal sample sizes $m = n$. 

5. Other tests. Besides the tests described in the preceding two sections, there are a number of others based on the behavior of the empirical distribution function.

Smirnov [73] discussed the number of crossings $N_n$ of $F_n(x)$ and $F(x)$. If (2.7) is true he proved that

$$\lim \Pr \left\{ N_n < t \sqrt{n} \right\} = 1 - e^{-t^2/2},$$

and gave generalizations. The distribution of the number of crossings of $F_n(x)$ and $G_n(x)$ is known [64] for $m = n$ finite.

For the case of two samples of size $n$, $m = np$ respectively, $p \geq 1$ integral, Gnedenko and Mihalavič ([31], [36]) proved that if $J$ is the number of "positive jumps" of $F_m(x)$—i.e., the number of $X_k$, $k = 1, 2, \ldots, m$ such that $F_m(X_k - 0) = (k - 1)/m \geq G_n(X_k)$—then $J$ has the simple distribution

$$\Pr \{ J = j \} = \frac{1}{m + 1}, \quad j = 0, 1, \ldots, m.$$

From this last result it follows (letting $p \to \infty$) that if $\Delta_n$ is the sum of the vertical parts of the graph of $F_n(x)$ which exceed $F(x)$—i.e.,

$$\Delta_n = \int_{-\infty}^{\infty} (F_n(x) - F(x)) \epsilon(F_n(x) - F(x)) \, dF_n(x),$$

where $\epsilon(x)$ is given by (2.1)—then $\Delta_n$ is uniformly distributed over $(0, 1)$

$$\Pr \{ \Delta_n < x \} = x, \quad 0 \leq x \leq 1.$$

The limiting form of this theorem was found earlier by Kac [42], who also gave a general method for finding the limiting distribution of

$$\int_{-\infty}^{\infty} V(F_n(x) - F(x)) \, dF(x),$$

for quite general functions $V$. Kac also considers the statistic corresponding to $K_n$ of (4.8) when the sample size $n$ is chosen at random with a Poisson distribution whose parameter goes to infinity.

Smirnov [78] considered using $F_n(x)$ to construct confidence limits, not for $U(x)$, but for its density by using a statistic similar to $K_n$.

The effect of grouping the data on the tests has been discussed for the $D_n$, $D_n^+$ tests in [24], [25], and [28]; the $K_n$, $K_n^+$ tests in [40], [23], and [33]; and the $W_n^2$ test in [87].

6. The parametric case. The two null hypotheses $H_0$, $H_0'$ of (2.7) and (2.8) are simple, and it is desirable to extend the tests to composite null hypotheses [14]. Some attention has been given to this problem lately for the hypothesis $H_0$.

We suppose, instead of (2.7),

$$(6.1) \quad H_0^* : U(x) = F(x, \theta), \quad \theta \in \Theta,$$
where the parameter \( \theta \) ranges over a set \( \Theta \). For the case when \( \Theta \) consists of an interval of the reals, \( a \leq \theta \leq b \), a test of \( H_0^a \) analogous to \( W^2_n \) of (3.1) was introduced in [16]:

\[
C_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x, \hat{\theta}_n))^2 \, dF(x, \hat{\theta}_n),
\]

where \( \hat{\theta}_n \) is an estimator of \( \theta \). \( H_0^a \) is to be rejected if \( C_n^2 \) is sufficiently large. The chief result here [16] is that, under suitable regularity conditions, if \( \text{Var} (\hat{\theta}_n) \) goes to zero sufficiently rapidly (the superefficient case), the limiting distributions of \( C_n^2 \) and \( W^2_n \) (with \( \psi = 1 \) in (3.1)) are the same, and if \( \theta \) admits a "regular estimator" \( \hat{\theta}_n \), then the limiting distribution of \( C_n^2 \) is that of \( \int_0^1 y^2(t) \, dt \), where \( y(t) \) is a Gaussian process with mean 0 and covariance

\[
k(s, t) = c(s, t) - \varphi(s)\varphi(t),
\]

with \( c(s, t) \) given by (2.3) and

\[
\varphi(F(x, \theta)) = \lim_{n \to \infty} \sqrt{n \, \text{Var}(\hat{\theta}_n)} \frac{\partial}{\partial \theta} F(x, \theta),
\]

\( \hat{\theta}_n \) being an asymptotically unbiased minimum variance estimator. The limiting distribution of \( C_n^2 \) is then given by (3.5) for \( D(\lambda) \) the Fredholm determinant of the kernel (6.3).

The test criterion \( C_n^2 \) of (6.2) is in its limiting form not generally distribution free—i.e., the limiting distribution of (6.2), if (6.1) is true, depends in general on the true unknown value of \( \theta \) and the structure of the family \( F(x, \theta) \), unlike the \( W^2_n \) test of (2.7). In the important special cases where \( \theta \) is a location, scale, or exponential parameter, the limiting distribution is independent of the particular value of \( \theta \) obtaining, which makes the test usable.

We quote one result: Let

\[
F(x, \theta) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} (x - \theta), \quad -\infty < x < \infty, \quad -\infty < \theta < \infty;
\]
i.e., we want to test if a sample of data came from some Cauchy distribution with unspecified median. Then [16]

\[
D(\lambda) = \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}} + \left( \frac{4\pi}{4\pi^2 - \lambda} \right)^2 (1 - \cos \sqrt{\lambda}),
\]

and the limiting distribution of (6.2) is that of (3.7), where the \( \lambda_j \) are the zeros of this \( D(\lambda) \).

In [44] the case where \( F(x, \theta) \) is a normal family with unknown mean and variance is treated in some detail, similar to the above analysis, and important power comparisons with the classical \( \chi^2 \) test were made (cf. Sec. 7). In [26] the problem was treated from a different viewpoint, with grouped data using an analogue of the \( K_n \) test of (4.1) and under a condition that the normalized estimator converged to a fixed value.
There seem to be no results for finite sample sizes, or a corresponding test of \( H_0 \), or a direct analogue of the \( K_n \) test. And there does not seem to be a single example where the limiting distribution of \( C_n^2 \) is known in a reasonable analytic form.

**7. Power of the tests.** In the research quoted thus far, the principal effort has been to obtain distributions and limiting distributions under the null hypotheses, with occasional fleeting and unsystematic remarks on the power of the tests. This important facet of the problem has only lately been studied and the results are still quite fragmentary concerning the optimum choice and relative power of the tests.

The choices of the weight functions (3.8), (4.12), etc., were made on more or less intuitive grounds to maximize the power of the tests against a rather vaguely defined class of alternatives; and indeed not only for the present tests but with other related distribution free tests (Wilcoxon, run, ranking, sign tests, etc.), there are fundamental and as-yet-unsolved problems as to delineating the classes of alternate hypotheses and of establishing realistic power comparisons.

Massey ([59], [63]) showed that the \( K_n \) test was consistent and biased, and he gave a lower bound for the power. Birnbaum [9] considered the \( K_n^+ \) test of (4.5) and a class of alternate hypotheses to (2.7) of the form

\[
\sup_{-\infty < z < \infty} (U(x) - F(x)) = \delta,
\]

and obtained best possible upper and lower bounds for the power for finite \( n \), and for \( n \to \infty \). The power of the \( D_{nm} \) test of (4.3) was compared with the \( \chi^2 \) test [60], and in the case of a normal family with unknown mean and variance, the \( C_n^2 \) test of (6.2) was found [44] to have considerable power advantage over the \( \chi^2 \) test for alternatives to (2.7) of the form

\[
\int_{-\infty}^{\infty} (U(x) - F(x))^2 \, dU(x) \geq \delta,
\]

(7.1)

\[
\sup_{-\infty < z < \infty} |U(x) - F(x)| \geq \delta.
\]

(7.2)

For example [44], when the class of alternatives (7.1) is considered for \( \delta \) sufficiently small, the size of the test being \( < \frac{1}{2} \), if it takes a sample size \( N \) for the \( \chi^2 \) test to achieve a minimum power \( \frac{1}{2} \) against all alternatives (7.1), then the \( C_n^2 \) test with the same size will need asymptotically only \( \alpha N^{4/5} \) observations to attain the same minimum power. Similar remarks hold for the alternatives (7.2) with a parametric extension of the \( K_n \) test.

The asymptotic power of the tests of \( H_0 \) of (2.7) can be studied by considering, e.g., alternatives to (2.7) of the form

\[
U(x) = F(x) + \frac{1}{\sqrt{n}} G(x),
\]

(7.3)
where \( G(x) \) is a specified function, and the merits of the various tests can be compared by considering the limiting probabilities with which (2.7) is rejected if (7.3) is true; and if the asymptotically most powerful test of (2.7) against (7.3) exists (and is known), one has the concept of asymptotic efficiency against the sequence of alternatives (7.3).

In the case of a normal distribution with mean 0 and variance 1, the alternatives being normal distributions with means \( \theta \), variances 1, \( \theta \neq 0 \), the known uniformly most powerful unbiased test of (2.7) was compared with the \( K_n \) test of (4.1) in [54], with the \( K_n \) test showing up fairly poorly, as might be expected. For the \( W_n^2 \) test (with \( \psi = 1 \)), the limiting distribution of (3.1) when (7.3) is true has been found under certain regularity conditions on \( F(x), G(x) \) by T. W. Anderson\(^3\), and is that of

\[
(7.4) \quad \int_0^1 [x(u) - k(u)]^2 \, du, 
\]

where \( x(u), 0 \leq u \leq 1 \), is a Gaussian process mean 0, covariance \( c(s, t) \) of (2.3), and \( k(u) \) is a certain function depending on \( F(x) \) and \( G(x) \). The distribution of (7.4) can be studied by methods similar to those in Sec. 3.

Alternatives to \( H_0^k \) of (2.8) of the form \( U(x) = V^k(x), k = 2, 3, \ldots \) have been investigated ([53], [82]) and power comparisons made for a number of tests including the \( D_{mn} \) test of (4.3).

For very small sample sizes, the exact distributions of \( K_n, K_n^+, D_{mn}, D_{mn}^+ \) can be computed by brute force when \( H_0, H_0^k \) are not necessarily true; and there has been some recent work of rather special character on their power. If \( F(x) \) is normal mean 0, variance 1, and \( U(x) \) is normal mean \( \mu > 0 \), variance 1, the \( K_n^+ \) test of (4.5), \( n = 2, 3, 5 \) has been compared [81] with the classical uniformly most powerful test. For \( U(x), V(x) \) normal, different means, variance \( \sigma^2 \), the test of \( H_0^k \) has similarly been investigated: \( \sigma^2 \) known [17], \( \sigma^2 \) unknown [82], and comparisons have been made with various other distribution-free tests. The \( K_n \) and \( D_{mn} \) tests do not perform exceptionally well, as might be surmised, and for increasing \( m, n \), their relative power is conjectured [81] to decrease.

Of course, essentially nothing in the way of an absolute judgement of the merits of the tests can be attained by such studies, since the alternatives against which the tests described here are supposed to have good power have little relation to the above alternatives against which the classical tests have maximum power.

REFERENCES


\(^3\) Personal communication.

\(^4\) This bibliography contains only the papers cited in the text. The titles in Russian have been translated.
KOLMOGOROV TESTS  


