THE CRAMÉR-SMIRNOV TEST IN THE PARAMETRIC CASE

D. A. DARLING[1]

University of Michigan and *Columbia University*

Summary. The "goodness of fit" problem, consisting of comparing the empirical and hypothetical cumulative distribution functions (cdf's), is treated here for the case when an auxiliary parameter is to be estimated. This extends the Cramér-Smirnov and von Mises test to the parametric case, a suggestion of Cramér [1], see also [2]. The characteristic function of the limiting distribution of the test function is found by consideration of a Guassian stochastic process.

1. Introduction; position of the problem. Let $X_1, X_2, \cdots, X_n$ be independent observations (random variables) coming from a population whose absolutely continuous cdf is $G(x)$. Let I be a nondegenerate interval on the real axis and suppose, for each ξ contained in the interior of I, that $F(x; \xi)$ is a cdf. In this paper we treat the problem of testing the hypothesis H,

$$H \quad : \quad G(x) = F(x; \xi) \qquad \textit{for some unspecified} \quad \xi \, \varepsilon \, I. \tag{1.1}$$

About the only test of H at present available seems to be the usual χ^2 test, discussed recently by Cochran [3].

For a somewhat different hypothesis H_0,

$$H_0 \quad : \quad G(x) = F(x) = F(x; \xi_0) \qquad \textit{for a specified} \quad \xi_0 \, \varepsilon \, I, \tag{1.2}$$

the following test of Cramér [2], Smirnov [4], and von Mises [5] is available. Let $F_n(x)$ be the empirical cdf of the data; that is, $F_n(x) = k/n$ if k of the X_i, with $i = 1, 2, \cdots, n$, are less than x, for $-\infty < x < \infty$. The test function is then

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x; \xi_0)]^2 \, dF(x; \xi_0), \tag{1.3}$$

and H_0 is rejected if W_n^2 is suitably large. The limiting distribution of W_n^2, if H_0 is true, was given by Smirnov; it has been tabulated recently [6].

This test of H_0 has certain attractive properties not possessed by the usual χ^2 test. It does not require a subjective grouping of the data into classes, it is distribution free for all n (if H_0 is true), and it is consistent (i.e., has limiting power 1). Appraisals of the χ^2 test and this test are given by Cochran [3] and Birnbaum [7].

In an effort to modify the W_n^2 test to treat the hypothesis H of (1.1), we shall consider, following a suggestion of Cramér, the test function

$$C_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x; \hat{\theta}_n)]^2 \, dF(x; \hat{\theta}_n), \tag{1.4}$$

Received Nov. 23, 1953, revised Sept. 21, 1954.

[1] Work sponsored by the Office of Scientific Research of the Air Force. Contract No. AF18(600)-442 Project No. R-345-20-7.

where $\hat{\theta}_n$ is a suitable estimator for the unknown parameter ξ in $F(x; \xi)$, and is a function of $X_1, X_2, \cdots, X_n$. The hypothesis H is to be rejected if C_n^2 is sufficiently large. If $X_1', X_2', \cdots, X_n'$ is a rearrangement of the sample $X_1, X_2, \cdots, X_n$ so that $X_1' < X_2' < \cdots < X_n'$, for computational purposes C_n^2 can be expressed more simply as

$$C_n^2 = \frac{1}{12n} + \sum_{j=1}^{n} \left[F(X_j'; \hat{\theta}_n) - \frac{2j-1}{2n} \right]^2. \tag{1.5}$$

The main analytical task is to calculate the distribution of C_n^2 when H is true; it is accomplished in the present paper. It turns out that matters depend crucially on the estimator $\hat{\theta}_n$ chosen. There are two essentially distinct cases.

If it happens that, among other conditions stated in Theorem 2.1, $\hat{\theta}_n$ is an estimator such that $nE\{(\hat{\theta}_n - \theta)^2\} \to 0$, where θ is the true value of ξ, that is, $G(x) = F(x; \theta)$, then the limiting distribution of C_n^2 is the same as the limiting distribution of W_n^2 given by (1.3). This is known and tabulated.

In the more general case when θ does not admit this so-called superefficient estimate but $\sqrt{n}(\hat{\theta}_n - \theta)$ has a limiting Gaussian distribution with mean 0 and variance $\sigma^2 > 0$, then the limiting distributions of C_n^2 and W_n^2 are not the same; the limiting distribution of C_n^2 is that of $\int_0^1 Y^2(t)\, dt$, where $Y(t)$ is a certain Gaussian stochastic process. The process has a specially simple structure if θ admits a minimum variance (or efficient) estimator $\hat{\theta}_n$, or if the maximum likelihood estimator is asymptotically efficient in the sense of Cramér [1].

In this latter case, we determine the ch.fn. of the limiting distribution of C_n^2 explicitly. We give several illustrations, but unfortunately even in simple, important cases (such as, a normal distribution with an unknown mean) the resulting ch.fn. appears very difficult to invert.

Unlike the W_n^2 test, the C_n^2 test is not distribution free. In general it depends on the unknown true value of ξ, though in important special cases (including the case when ξ is a scale or location parameter) the asymptotic test depends only on the structure of the family $F(x; \xi)$ and not upon the particular value of ξ obtaining; that is, the test is parameter-free, so to speak.

2. The superefficient case. Suppose that H as given by (1.1) is true, and let the true unknown value of the parameter ξ be θ, with θ an interior point of I. Let the density corresponding to $F(x; \xi)$ be $f(x; \xi)$. Denote as before

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

THEOREM 2.1. *Assume that the estimator $\hat{\theta}_n$ and the distribution $F(x; \xi)$ have the following properties:*

1) $nE\{(\hat{\theta}_n - \theta)^2\} \to 0, n \to \infty$;

2) *For $\xi, \xi' \in I$, $F(x; \xi)$ satisfies a Lipschitz condition*

$$| F(x; \xi) - F(x; \xi') | < A(x) | \xi - \xi' |,$$

where $\Pr\{A^2(X_1) > A_0\} = 0$ *for some* $A_0 < \infty$, *the probability according to* $F(x;\theta)$.

Then $C_n^2 = W_n^2 + \delta_n$, *where* $\delta_n \to 0$ *in probability.*

PROOF. From (1.5)

$$\begin{aligned} C_n^2 &= \frac{1}{12n} + \sum_{j=0}^{n} \left(F(X_j';\hat{\theta}_n) - \frac{j-\frac{1}{2}}{n} \right)^2 \\ &= \frac{1}{12n} + \sum_{j=0}^{n} \left(\left\{ F(X_j';\theta) - \frac{j-\frac{1}{2}}{n} \right\} + \{F(X_j';\hat{\theta}_n) - F(X_j';\theta)\} \right)^2 \\ &= W_n^2 + 2\sum \left\{ F(X_j';\theta) - \frac{j-\frac{1}{2}}{n} \right\} \{F(X_j';\hat{\theta}_n) - F(X_j';\theta)\} \\ &\qquad\qquad + \sum \{F(X_j';\hat{\theta}_n) - F(X_j';\theta)\}^2 \\ &= W_n^2 + 2\delta_1 + \delta_2 . \end{aligned}$$

Then $\delta_1^2 \leqq (W_n^2 - 1/12n)\delta_2$, and

$$\delta_2 = \sum \{F(X_j';\hat{\theta}_n) - F(X_j';\theta)\}^2 \leqq n(\hat{\theta}_n - \theta)^2 \max_j A^2(X_j).$$

Thus $E(\delta_2) \to 0$ and $E(\delta_1^2) \leqq E\{(W_n^2 - 1/12n)^2\}^{1/2} E(\delta_2^2)^{1/2} \to 0$, and the theorem follows. A trivial consequence of the theorem is

COROLLARY 2.1. *Under the conditions* 1) *and* 2) *of Theorem* 1.1, *the limiting distribution of* C_n^2 *is the same as that of the von Mises statistic* W_n^2.

The limiting distribution of W_n^2 has been tabulated [6]; the problem in this case is solved. Two simple examples in which the estimate $\hat{\theta}_n$ is unbiased follow.

a.) It is easily verified that $\operatorname{Var}(\hat{\theta}_n) = \theta^2/n(n+2)$, and that condition 2) of Theorem 1.1 is satisfied, when

$$I = \{\xi \mid 0 < \delta < \xi < \infty\}; \qquad f(x;\xi) = \begin{cases} 1/\xi, & 0 < x < \xi, \\ 0, & \text{otherwise}; \end{cases}$$

$$\hat{\theta}_n = \frac{n+1}{n} \max\{X_1, X_2, \cdots, X_n\}.$$

b.) Likewise, $\operatorname{Var}(\hat{\theta}_n) = k/2(n+1)(n+2)$, and 2) is satisfied by $f(x;\xi)$, when

$$I = \{\xi \mid -\infty < \xi < \infty\}; \qquad f(x;\xi) = \begin{cases} 1, & \xi - \frac{1}{2} < x < \xi + \frac{1}{2}, \\ 0, & \text{otherwise}; \end{cases}$$

$$\hat{\theta}_n = \tfrac{1}{2}(U + V).$$

Here U is the kth largest X_i and V is the kth smallest X_i, for $i = 1, 2, \cdots n$, and k is independent of n (e.g., $k = 1$ gives $\hat{\theta}_n$ the midrange).

3. The regular estimation case. In general, the rather rapid decrease of $\operatorname{Var}(\hat{\theta}_n)$ to zero as expressed by condition 1) of Theorem 1.1 will not occur. In the case of regular estimation of Cramér ([1] p. 477), we will have $\operatorname{Var}(\hat{\theta}_n) \geqq A/n$ for some positive A—the Cramér-Rao inequality. But it will generally happen that $n^{1/2-\delta}(\hat{\theta}_n - \theta)$ will converge in probability to zero for $\delta > 0$. To cover this situation, we have the following

LEMMA 3.1. *Assume that* $f(x; \xi)$ *and* $\hat{\theta}_n$ *are such that*

1) $nE\{(\hat{\theta}_n - \theta)^4\} \to 0,$

2) $\left| \dfrac{\partial^2 F(x; \xi)}{\partial \xi^2} \right| < g_0(x),$

3) $\left| \dfrac{\partial f(x; \xi)}{\partial \xi} \right| < g_1(x),$

for almost all x, *where* g_0 *and* g_1 *are integrable from* $-\infty$ *to* $+\infty$. *The functions* g_0 *and* g_1 *and the exceptional set do not depend on* ξ.

Then,

$$(3.1)\quad C_n^2 = n \int_{-\infty}^{\infty} \left\{ F_n(x) - F(x) - (\hat{\theta}_n - \theta) \frac{\partial}{\partial \theta} F(x; \theta) \right\}^2 f(x; \theta)\, dx + \delta_n,$$

where $\delta_n \to 0$ *in probability.*

For almost all x we have

$$F(x; \hat{\theta}_n) = F(x; \theta) + (\hat{\theta}_n - \theta) \frac{\partial}{\partial \theta} F(x; \theta) + \tfrac{1}{2}(\hat{\theta}_n - \theta)^2 q_0 g_0(x), \qquad |q_0| < 1;$$

$$f(x; \hat{\theta}_n) = f(x; \theta) + (\hat{\theta}_n - \theta) q_1 g_1(x), \qquad |q_1| < 1.$$

Putting these expansions in expression (2.1), we obtain the lemma after some calculations.

Conditions 2) and 3) in the above lemma could be replaced by the condition, similar to that of Theorem 2.1,

$$\left| \frac{\partial FX; \xi)}{\partial \xi} - \frac{\partial F(X; \xi')}{\partial \xi'} \right| < A\, |\xi - \xi'|.$$

Thus it is necessary to study only the limiting form (if it exists) of the distribution of

$$R_n^2 = n \int_{-\infty}^{\infty} \left\{ F_n(x) - F(x; \theta) - (\hat{\theta}_n - \theta) \frac{\partial}{\partial \theta} F(x; \theta) \right\}^2 f(x; \theta)\, dx.$$

A transformation which is basic in the work to follow,

$$(3.2)\qquad u = F(x; \theta),$$

defines implicitly x as a function of u. Thus $x = x(u; \theta)$ except possibly for an enumerable set of u values at which x can be defined arbitrarily, except to render it monotone nondecreasing. Next we define the function $g(u)$ as

$$(3.3)\qquad g(u) = \frac{\partial}{\partial \theta} F(x; \theta), \qquad 0 \leqq u \leqq 1,$$

and note that $g(u)$ depends in general on θ. Finally, we express the empirical cdf as a function of u. If we introduce the function $\psi_t(v)$ defined as

$$\psi_t(v) = \begin{cases} 1, & v < t, \\ 0, & v \geqq t, \end{cases}$$

then we have, with probability one,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \psi_x(X_j) = \frac{1}{n} \sum_{j=1}^{n} \psi_u(F(X_j\,;\theta)),$$

where u is given by (3.2.). On writing

$$Z_n(u) = \sqrt{n}\left(\frac{1}{n} \sum_{j=1}^{n} \psi_u(F(X_j\,;\theta)) - u\right), \tag{3.4}$$

$$T_n = \sqrt{n}(\hat{\theta}_n - \theta), \tag{3.5}$$

C_n^2 can be written, from (3.1), as

$$C_n^2 = \int_0^1 (Z_n(u) - T_n g(u))^2 \, du + \delta_n$$

where $\delta_n \to 0$ in probability, and $g(u)$ is given by (3.3). Finally by defining the stochastic process $Y_n(u)$ as

$$Y_n(u) = Z_n(u) - T_n g(u) \tag{3.6}$$

where $Z_n(u)$ and T_n are given by (3.4) and (3.5), and $g(u)$ by (3.3), we have

$$C_n^2 = \int_0^1 Y_n^2(u) \, du + \delta_n$$

where $\delta_n \to 0$ in probability.

It follows that the limiting form of the stochastic process $Y_n(u)$ is of central importance, and we next prove

LEMMA 3.2. *Assume, in addition to conditions* 1), 2), 3) *of Lemma* 3.1, *that*

4) $nE(\hat{\theta}_n - \theta) \to 0$;

5) $T_n = \sqrt{n}(\hat{\theta}_n - \theta)$ *is a sum of independent identically distributed random variables having a limiting Gaussian distribution with variance* $\sigma^2 > 0$; *and*

6) *the conditional expectation* $nuE(\hat{\theta}_n - \theta \mid F(X_1\,;\theta) < u)$ *converges to a function* $h(u)$ *with* $0 < u < 1$ *and* $h(0) = h(1) = 0$.

Then the process $Y_n(u)$ *converges in distribution to a Gaussian process* $Y(u)$ *with mean* 0 *and covariance*

$$\begin{aligned} \rho(u, v) &= E(Y(u)Y(v)) \\ &= \min(u, v) - uv - g(u)h(v) - g(v)h(u) + \sigma^2 g(u)g(v). \end{aligned} \tag{3.7}$$

The expression "$Y_n(u)$ converges to $Y(u)$ in distribution" means that for every finite set $u_1, u_2, \cdots, u_k$ the joint cdf of $Y_n(u_1), Y_n(u_2), \cdots, Y_n(u_k)$ converges to the joint cdf of $Y(u_1), Y(u_2), \cdots, Y(u_k)$. The proof of the lemma is quite straightforward. The process $Z_n(u)$ defined by (3.4) is known to converge in distribution to a Gaussian process with mean 0 and covariance $k(u, v) = E(Z_n(u)Z_n(v))$ given by

$$k(u, v) = \min(u, v) - uv, \tag{3.8}$$

with $k(u, v)$ being independent of n (see, for example, [6]). By 5) T_n has a limiting Gaussian distribution with mean 0 and variance σ^2. It follows from the multidimensional central limit theorem that the linear combination $Y_n(u) = Z_n(u) - T_n g(u)$ converges in distribution to a Gaussian process whose mean is 0. Thus will be sufficient to verify that the limiting covariance $\rho_n(u, v) = E(Y_n(u)Y_n(v))$ converges to $\rho(u, v)$ given by (3.7).

Denoting by $h_n(u)$ the function

$$h_n(u) = E(Z_n(u)T_n) \tag{3.9}$$

and using (3.4) and (3.5), we have, by condition 5),

$$\begin{aligned} h_n(u) &= \sqrt{n}E\{[(1/n)\textstyle\sum \psi_u(F(X_j\,;\,\theta)) - u][\sqrt{n}(\hat{\theta}_n - \theta)]\} \\ &= \textstyle\sum E[\psi_u(F(X_j\,;\,\theta))(\hat{\theta}_n - \theta)] - nuE(\hat{\theta}_n - \theta) \\ &= nE\{(\hat{\theta}_n - \theta)\psi_u(F(X_1\,;\,\theta))\} - nuE(\hat{\theta}_n - \theta) \\ &= nuE\{\hat{\theta}_n - \theta \mid F(X_1\,;\,\theta) < u\} - nuE(\hat{\theta}_n - \theta). \end{aligned} \tag{3.10}$$

Consequently by 4) and 6), $h_n(u) \to h(u)$ for $0 \leqq u \leqq 1$. The covariance $\rho_n(u, v)$ is

$$\begin{aligned} \rho_n(u, v) &= E\{[Z_n(u) - T_n g(u)][(Z_n(v) - T_n g(v)]\} \\ &= k(u, v) - h_n(u)g(v) - h_n(v)g(u) + \sigma_n^2 g(u)g(v), \end{aligned} \tag{3.11}$$

where $\sigma_n^2 = \operatorname{Var}(T_n) \to \sigma^2$ and $k(u, v)$ is given by (3.8). Thus we obtain $\rho_n(u, v) \to \rho(u, v)$ for $\rho(u, v)$ as in (3.7), and the lemma is established.

It might be concluded that the limiting distribution of C_n^2 is the distribution of $C^2 = \int_0^1 Y^2(u)\,du$ where $Y(u)$ is a Gaussian process with mean 0 whose covariance is given by (3.7), following Doob's [8] heuristic approach. But we shall prove this fact in the next section only when the estimator $\hat{\theta}_n$ is further specialized.

Two further properties of the function $h_n(u)$ conclude this section.

LEMMA 3.3. *For the function $h_n(u)$ as defined by* (3.9),

1) $|h_n(u)| \leqq \sigma_n\sqrt{u(1 - u)}, \qquad 0 \leqq u \leqq 1;$

2) $h_n'(u) = nE\{\hat{\theta}_n - \theta \mid F(X_1\,;\,\theta) = u\} - nE(\hat{\theta}_n - \theta).$

The first relation follows immediately by Schwarz' inequality from (3.9) by using (3.8) with $u = v$ and $\operatorname{Var}(T_n) = \sigma_n^2$. For the second relation we use (3.10) to give

$$h_n(u + \delta) - h_n(u) = n\delta E\{(\hat{\theta}_n - \theta) \mid [u < F(X_1\,;\,\theta) \leqq u + \delta]\} - n\delta E\{\hat{\theta}_n - \theta\}.$$

By letting $\delta \to 0$, 2) follows.

For $n \to \infty$, $h_n(u)$ converges to $h(u)$ and σ_n^2 converges to σ^2. Thus we have $| h(u) | \leqq \sigma\sqrt{u(1 - u)}$. However, we cannot conclude directly that $\lim h_n'(u)$ exists.

4. Case of an efficient estimator. As yet, we have given no special attention to the choice of the estimator $\hat{\theta}_n$. It might be thought that, paralleling the principle of minimum χ^2, we should choose $\hat{\theta}_n$ so as to make C_n^2, as given by (2.1), a minimum. However, as is often the case with minimum χ^2, this does not lead to usable results. However, precisely as in the χ^2 case, the maximum likelihood principle does lead to a certain ideal properties for C_n^2, at least asymptotically.

In this section, we assume that Cramér's conditions ([1], pp. 477–489) for a regular, unbiased efficient (or minimum variance) estimate are satisfied. Following Cramér, we simply term the estimate efficient. Then all the conditions 1) through 6) of Lemma 3.2 are satisfied, as noted below, with the possible exception of condition 2), which we shall further presume satisfied. Condition 2) is also postulated by Cramér for the maximum likelihood case which we consider in the next section. The efficient estimator is unbiased so that condition 4) is satisfied, and implies besides that the likelihood function

$$L = \prod_{j=1}^{n} f(X_j\,;\xi)$$

has the property that if $\xi = \hat{\theta}_n$ is a root of $(\partial/\partial\xi)\log L = 0$, then, defining

$$\sigma^2 = 1/E\left\{\left[\frac{\partial}{\partial\theta}\log f(X_1\,;\theta)\right]^2\right\} = \left(\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2 f(x;\theta)\,dx\right)^{-1}, \tag{4.1}$$

we have

$$\frac{\partial}{\partial\xi}\log L = \sum_{j=1}^{n}\frac{\partial}{\partial\xi}\log f(X_j\,;\xi) = \frac{n}{\sigma^2}(\hat{\theta}_n - \xi).$$

By putting $\xi = \theta$, this yields

$$\frac{n}{\sigma^2}(\hat{\theta}_n - \theta) = \sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f(X_j\,;\theta) = \frac{\sqrt{n}}{\sigma^2}T_n. \tag{4.2}$$

It then follows that the variance of T_n is σ^2, independent of n, given by (4.1), and that conditions 5) and 6) are satisfied.

For condition 7) of Lemma 3.2, we multiply through the last equality by σ^2 and take conditional expectations of both sides under the condition that $F(X_1\,;\theta) = u$. Then on using 2) of Lemma 3.3, and the fact that $E[(\partial/\partial\theta)\log f(X_j\,;\theta)] = 0$, we have

$$\begin{aligned} h_n'(u) &= \sigma^2\sum_{j=1}^{n} E\left\{\frac{\partial}{\partial\theta}\log f(X_j\,;\theta) \mid F(X_1\,;\theta) = u\right\} \\ &= \sigma^2 E\left\{\frac{\partial}{\partial\theta}\log f(X_1\,;\theta) \mid F(X_1\,;\theta) = u\right\} \\ &= \sigma^2\frac{\partial}{\partial\theta}\log f(x;\theta), \end{aligned}$$

using transformation (3.2). Thus $h_n'(u)$ is *independent* of n. We denote it by $h'(u)$, given by

$$h'(u) = \sigma^2 \frac{\partial}{\partial \theta} \log f(x;\theta). \tag{4.3}$$

Because of this simple formula for $h(u)$, making it proportional to $g(u)$ as we see immediately, $\rho(u, v)$ simplifies and renders the process $Y(u)$ of Lemma 3.2 manageable.

From the definition of $g(u)$ in (3.3) we have immediately $g(0) = g(1) = 0$ and

$$g'(u) = \frac{\partial}{\partial \theta} f(x;\theta) \frac{dx}{du} = \frac{1}{f(x;\theta)} \frac{\partial}{\partial \theta} f(x;\theta) = \frac{\partial}{\partial \theta} \log f(x;\theta) = \frac{1}{\sigma^2} h'(u).$$

Since $h(0) = 0$ by 1) of Lemma 3.3, by integrating we obtain $h(u) = \sigma^2 g(u)$. Thus it follows that we can define a function $\varphi(u)$,

$$\varphi(u) = \frac{h(u)}{\sigma} = \sigma g(u), \tag{4.4}$$

where $h'(u)$ is given by (4.3) and $g(u)$ by (3.3), σ being given by (4.1) and x by (3.2).

Putting these values for $h(u)$ and $g(u)$ in (3.11), we obtain

LEMMA 4.1. *In the case of an efficient estimator, the process $Y_n(u)$ given by* (3.6) *has mean* 0 *and covariance*

$$\rho(u, v) = k(u, v) - \varphi(u)\varphi(v), \tag{4.5}$$

independent of n, where $k(u, v)$ is given by (3.8) *and $\varphi(u)$ by* (4.4). *The function $\varphi(u)$ has the properties*

$$\text{a)} \quad |\varphi(u)| \leqq \sqrt{u(1-u)}, \qquad \text{b)} \quad \int_0^1 \varphi'^2(u)\,du = 1.$$

Condition a) follows from 1) of Lemma 3.3. To prove b), from (4.4) and (4.3) we have

$$\int_0^1 \varphi'^2(u)\,du = \frac{1}{\sigma^2} \int_0^1 h'^2(u)\,du = \sigma^2 \int_0^1 \left[\frac{\partial}{\partial \theta} \log f(x;\theta)\right]^2 du$$
$$= \sigma^2 \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} \log f(x;\theta)\right]^2 f(x;\theta)\,dx = 1,$$

when we use (4.1).

Thus we have shown that, when an efficient estimator $\hat{\theta}_n$ exists, the limiting distribution of C_n^2 is the same as the limiting distribution of $\int_0^1 Y_n^2(u)\,du$, where $Y_n(u)$ has mean 0 and covariance $\rho(s, t)$ given by (4.5), and approaches a Gaussian process $Y(u)$ in distribution. We need to show that this limiting distribution is the same as the distribution of $C^2 = \int_0^1 Y^2(u)\,du$ where $Y(u)$ is a Gaussian process with mean 0 and covariance $\rho(s, t)$. To state the theorem formally,

THEOREM 4.1. *In the case of an efficient estimator,*

$$\lim_{n\to\infty} \Pr \{C_n^2 < x\} = \Pr\left\{\int_0^1 Y^2(u)\, du < x\right\},$$

where $Y(u)$ *is a Gaussian process with mean* 0 *and covariance* $\rho(u, v)$ *given by* (4.5).

Actually we are going to suppose that $\varphi(u)$ satisfies the

AUXILIARY ASSUMPTION. $\varphi''(u)$ exists almost everywhere for $0 \leqq u \leqq 1$ and

$$\int_0^1 |\varphi''(u)|\, u(1 - u) \log \log \frac{1}{u(1 - u)}\, du < \infty. \tag{4.6}$$

We first show the $Y_n(u)$ process can be expressed in terms of $Z_n(u)$ and that we have the representation

$$Y_n(u) = Z_n(u) + \varphi(u) \int_0^1 \varphi''(t) Z_n(t)\, dt. \tag{4.7}$$

From (3.6) and (4.2) we have merely to show

$$\int_0^1 \varphi''(t) Z_n(t)\, dt = -\frac{\sigma}{\sqrt{n}} \sum_{j=1}^n \frac{\partial}{\partial\theta} \log f(X_j; \theta).$$

By using (3.4) for $Z_n(t)$, integrating by parts, and using (4.3) to obtain $\varphi'(t) = \sigma(\partial/\partial\theta) \log f(x; \theta)$, where $t = F(x; \theta)$, the result follows.

Now the process $Z(u)$, which is the limit in distribution of $Z_n(u)$ given by (3.4), is Gaussian with mean 0 and, by (3.8), covariance $k(s, t) = \min(s, t) - st$. It has been shown [6] that the random variable $\int_0^1 \varphi''(u) Z(u)\, du$ exists (and is Gaussian with mean 0) when the auxiliary assumption (4.6) is satisfied. Then the process

$$Y(u) = Z(u) + \varphi(u) \int_0^1 \varphi''(t) Z(t)\, dt \tag{4.8}$$

is Gaussian with mean 0 and covariance

$$\begin{aligned} E(Y(u)Y(v)) = k(u, v) &+ \varphi(u) \int_0^1 \varphi''(t) k(v, t)\, dt + \varphi(v) \int_0^1 \varphi''(t) k(t, u)\, dt \\ &+ \varphi(u)\varphi(v) \int_0^1 \int_0^1 \varphi''(t)\varphi''(s) k(t, s)\, dt\, ds. \end{aligned}$$

Since $\int_0^1 \varphi''(t) k(v, t)\, dt = -\varphi(v)$, the covariance becomes

$$\begin{aligned} E(Y(u)Y(v)) &= k(u, v) - 2\varphi(u)\varphi(v) + \varphi(u)\varphi(v) \int_0^1 \varphi''(t)\varphi(t)\, dt \\ &= k(u, v) - 2\varphi(u)\varphi(v) + \varphi(u)\varphi(v) \int_0^1 \varphi'^2(t)\, dt \\ &= k(u, v) - \varphi(u)\varphi(v) = \rho(u, v), \end{aligned}$$

where we have used b) of Lemma 4.1.

Hence (4.8) is a *representation* of the process $Y(u)$ of Theorem 4.1 in terms of the process $Z(u)$. Then since $\int_0^1 Y_n^2(u)\,du$ is a functional of $Z_n(u)$ continuous in the uniform topology by (4.7), it follows from a theorem of Donsker [9] that its limiting distribution is the same as the distribution of $\int_0^1 Y^2(u)\,du$ for the process (4.8) or the process $Y(u)$ of the theorem, and the proof is complete.

By virtue of Theorem 4.1 we can now concentrate on the process $Y(u)$ and attempt to find the distribution of $C^2 = \int_0^1 Y^2(u)\,du$.

5. Case of a maximum likelihood estimator. When an efficient estimator for θ, for finite n, does not exist, it will generally happen that a maximum likelihood estimator for θ will exist and be asymptotically efficient in the sense of Cramér. In this case, the results of Section 4 will still be valid.

We shall assume that the conditions of Cramér ([1] pp. 500–501) are satisfied. These conditions imply that conditions 1)–6) of Lemmas 3.1 and 3.2 are satisfied with the exception of condition 4). We shall suppose this condition to be satisfied also, calling it a condition of "*weak-biasedness*" on $\hat{\theta}_n$. In some cases, as with the efficient estimators, $\hat{\theta}_n$ is actually unbiased and 4) is trivially satisfied.

The following analysis parallels exactly the development in the preceding section. The formulas are all the same except for additional terms which approach 0 in the mean and with probability one. We merely indicate the development leading to theorem 5.1.

Defining σ^2 as in (4.1), we can write

$$\frac{\partial}{\partial \xi} \log L = \sum_{j=1}^{n} \frac{\partial}{\partial \xi} \log f(X_j\,;\xi) = \frac{n}{\sigma^2}(\hat{\theta}_n - \xi)(1 + \epsilon_n),$$

where $\epsilon_n \to 0$ in the mean and with probability one. Then (4.2) becomes, with $\epsilon_n' \to 0$ strongly,

$$\frac{n}{\sigma^2}(\hat{\theta}_n - \theta) = \sum_{j=1}^{n} \frac{\partial}{\partial \theta} \log f(X_j\,;\theta)(1 + \epsilon_n').$$

Consequently, $h_n'(u) \to h'(u)$ and $\sigma_n^2 = \operatorname{Var}(\sqrt{n}(\hat{\theta}_n - \theta)) \to \sigma^2$ for $n \to \infty$, where $h_n'(u)$ is given by 2) of Lemma 3.3 and $h'(u)$ by (4.3). Accordingly, $\rho_n(u, v)$ given by (3.11) converges to $\rho(u, v)$ defined by (4.5). Thus we can generalize Theorem 4.1 to

THEOREM 5.1. *If $\hat{\theta}_n$ is a weakly-biased maximum likelihood estimator satisfying Cramér's conditions, and if $\varphi(u)$ satisfies* (4.6), *then*

$$\lim \Pr\{C_n^2 < x\} = \Pr\left\{\int_0^1 Y^2(u)\,du < x\right\}$$

where $Y(u)$ is a Gaussian process with mean 0 *and covariance $\rho(s, t)$ given by* (4.5).

The theorem says that C_n^2 converges in distribution to $C^2 = \int_0^1 Y^2(u)\,du$.

6. The limiting distribution of C_n^2. In the preceding sections we have reduced the problem of finding the limiting distribution of C_n^2 under general conditions to that of finding the distribution of $\int_0^1 Y^2(u)\,du$; where $Y(u)$ is a Gaussian process with mean 0 and covariance $\rho(s, t)$ given by

$$\rho(u, v) = k(u, v) - \varphi(u)\varphi(v), \qquad k(u, v) = \min(u, v) - uv; \tag{6.1}$$

$$\varphi(u) = \sigma \frac{\partial}{\partial\theta} F(x; \theta), \qquad u = F(x; \theta);$$
$$\sigma^2 = \left[\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log f(x; \theta)\right)^2 f(x; \theta)\,dx\right]^{-1}. \tag{6.2}$$

The function $\varphi(u)$ can also be determined from

$$\varphi'(u) = \sigma \frac{\partial}{\partial\theta} \log f(x; \theta), \qquad u = F(x; \theta).$$

In determining the distribution of $\int_0^1 Y^2(u)\,du$ we shall use a basic theorem due essentially to Kac-Siegert [10]. It states that the distribution of this random variable is the same as the distribution of a sum of weighted χ^2,

$$C^2 = \int_0^1 Y^2(u)\,du = \sum_{n=1}^{\infty} \frac{G_n^2}{\lambda_n},$$

where $G_1, G_2, \cdots$ are independent normally distributed random variables with mean 0 and variance 1, and $\lambda_1, \lambda_2, \cdots$ are the eigenvalues of the kernel $\rho(s, t)$. Proof of this result for nonstationary processes follows almost immediately from [10]; it may be found, for example, in Fortet's paper [11]. We have for the ch.fn.

$$E\left\{\exp\left[it \int_0^1 Y^2(u)\,du\right]\right\} = \prod_{j=1}^{\infty} \left(1 - \frac{2it}{\lambda_j}\right)^{-1/2}.$$

This expression is merely $(D(2it))^{-1/2}$, where $D(\lambda)$ is the Fredholm determinant associated with the kernel $\rho(u, v)$ in (6.1). To state matters formally, we have

THEOREM 6.1. *Under the conditions of Theorem* 5.1,

$$\lim_{n\to\infty} E\{\exp(itC_n^2\} = [D(2it)]^{-1/2},$$

where $D(\lambda)$ is the Fredholm determinant of the integral equation

$$f(x) = \lambda \int_0^1 \rho(x, y) f(y)\,dy$$

in which $\rho(x, y)$ is given by (6.1).

Thus, for $\varphi(u)$ given, it is necessary only to determine $D(\lambda)$ and invert the ch.fn. $(D(2it))^{-1/2}$ in order to obtain the limiting distribution of C_n^2. It turns out to be possible to get a fairly explicit formula for $D(\lambda)$, but the resulting ch.fn.'s seem very difficult to invert.

We first prove a result more general than needed but which may be of independent interest.

THEOREM 6.2. *Let $k_1(x, y)$ be a symmetric, bounded positive definite kernel over the unit square $0 \leqq x, y \leqq 1$ whose Fredholm determinant $d_1(\lambda)$ has simple zeroes $0 < \lambda_1 < \lambda_2 < \cdots$, and let the corresponding eigenfunctions be $f_1(x), f_2(x), \cdots$. Let*

$$a_j = c(f_j) = \int_0^1 f_j(x)\varphi_1(x)\,dx, \qquad \varphi_1(x) \in L_2(0, 1).$$

Put $c(g) = \int_0^1 g(x)\varphi_1(x)\,dx$. Then the Fredholm determinant $D_1(\lambda)$ for the kernel

$$\rho_1(x, y) = k_1(x, y) - \varphi_1(x)\varphi_1(y)$$

is given by

$$D_1(\lambda) = d_1(\lambda)\left(1 + \lambda \sum_{n=1}^{\infty} \frac{a_n^2}{1 - \lambda/\lambda_n}\right).$$

PROOF. The integral equation

$$g(x) = \lambda \int_0^1 [k_1(x, y) - \varphi_1(x)\varphi_1(y)]\,g(y)\,dy \tag{6.3}$$

can be written

$$g(x) = -\lambda\varphi_1(x)c(g) + \lambda \int_0^1 k_1(x, y)\,g(y)\,dy. \tag{6.4}$$

We have the familiar series ([12] p. 228)

$$g(x) = -\lambda c(g) \sum \frac{a_j}{1 - \lambda/\lambda_j} f_j(x), \qquad \lambda \neq \lambda_j. \tag{6.5}$$

This is not a solution to (6.3), since g appears on both sides. But we may multiply both sides by $\varphi_1(x)$ and integrate; termwise integration of the series is easily justified. We obtain

$$c(g) = -c(g)\lambda \sum a_j^2/(1 - \lambda/\lambda_j).$$

Hence either g is such that $c(g)$ vanishes, or else λ must be a root of the function

$$P(\lambda) = 1 + \lambda \sum a_j^2/(1 - \lambda/\lambda_j).$$

If $\lambda \neq \lambda_j$, then $c(g)$ cannot be zero, since otherwise (6.4) would be a homogeneous equation with a nontrivial solution existing for λ not an eigenvalue of the kernel. Consequently, the only possible values of λ for which (6.3) has a solution are either the roots of $P(\lambda)$ or else $\lambda = \lambda_j$. That is, λ must be a root of $D_1(\lambda) = d_1(\lambda)P(\lambda)$.

Now $P(\lambda)$ is analytic save for possible simple poles at $\lambda = \lambda_j$. Hence $D_1(\lambda)$ is an entire function of λ with $D_1(0) = 1$, and is of genus zero. We note that

$$\sum 1/\lambda_j = \int_0^1 k_1(s, s)\,ds < \infty, \qquad \sum a_j^2 = \int_0^1 \varphi_1^2(t)\,dt < \infty.$$

Consequently, if for any root $\lambda = \lambda^*$ of $D_1(\lambda)$, (6.3) has a solution $g^*(x)$ with $\int_0^1 g^{*2}(x)\,dx = 1$, then $D_1(\lambda)$ is indeed the Fredholm determinant of the kernel $\rho_1(x, y) = k_1(x, y) - \varphi_1(x)\varphi_1(y)$ in (6.3).

Let now $\lambda = \lambda^*$ be a root of $D_1(\lambda) = d_1(\lambda)P(\lambda)$. We have the following two cases

$$(1)\quad \lambda^* \neq \lambda_j\,; \qquad (2)\quad \lambda^* = \lambda_j\,, \quad a_j = 0.$$

In case (2), if $\lambda^* = \lambda_j$ and $a_j \neq 0$, then $D_1(\lambda^*) \neq 0$, since the roots of $d_1(\lambda)$ are simple.

In case (1), λ^* is a root of $P(\lambda)$; we show immediately that λ^* is real and simple. If we substitute the function

$$g^*(x) = -\sum \frac{a_j f_j(x)}{1 - \lambda^*/\lambda_j} \Big/ \sqrt{\sum \frac{a_j^2}{(1 - \lambda^*/\lambda_j)^2}} \tag{6.6}$$

in (6.3), we verify easily that it satisfies the equation. Since the kernel in (6.3) is symmetric, λ^* is real. Then, since

$$P'(\lambda) = \sum \left(\frac{a_j}{1 - \lambda/\lambda_j}\right)^2, \qquad \lambda \neq \lambda_j \text{ if } a_j \neq 0,$$

is positive for real λ, λ^* is simple; the case when $\int_0^1 \varphi_1^2(x)\,dx = 0$ is disposed of easily, of course. Thus, for any root under case (1) we have the solution $g^*(x)$ given by (6.6).

In case (2) there are two subcases, $P(\lambda^*) \neq 0$ and $P(\lambda^*) = 0$. If $P(\lambda^*) \neq 0$, then $\lambda^* = \lambda_j$ is a simple zero of $D_1(\lambda)$; in this case $f_j(x)$ is readily seen to satisfy (6.3), and $f_j(x)$ is orthogonal to $\varphi_1(x)$. If $P(\lambda^*) = 0$, then $D_1(\lambda)$ has a double root at $\lambda = \lambda^*$, and besides the solution $f_j(x)$ we have the solution $g^*(x)$ of (6.6), which is orthogonal to $f_j(x)$. Accordingly, to each root of $D_1(\lambda)$ we obtain solutions to (6.3) of the proper multiplicity. Theorem 6.2 is thus proved.

There is only a slight complication to prove an analogous theorem in the case when $d_1(\lambda)$ has repeated roots.

By plotting the function $P(\lambda)$ for real λ, it is easy to prove that $D_1(\lambda)$ has at most one negative zero λ_0^*; the condition for its existence is that $\sum_{j=1}^{\infty} \lambda_j a_j^2 > 1$, or that the series diverges. If the positive zeros are $0 < \lambda_1^* \leqq \lambda_2^* \leqq \cdots$, repeated according to their multiplicity, we verify easily that $\lambda_j^* \geqq \lambda_j$ for $j = 1, 2, \cdots$. This remark will be useful in the sequel.

In Theorem 6.2, in order to find $D_1(\lambda) = d_1(\lambda)P(\lambda)$, it is necessary to know the λ_j and the Fourier coefficients a_j of $\varphi_1(x)$. The following theorem gives an evaluation of $D_1(\lambda)$ through the solution of a nonhomogeneous integral equation without specifically bringing these quantities into evidence.

THEOREM 6.3 *Let $k_1(x, y)$ and $\varphi_1(x)$ be as in Theorem 5.2. Regarding s as a parameter only, let $u(t, s)$ be the solution of*

$$u(t, s) = \varphi_1(t)\varphi_1(s) + \lambda \int_0^1 k_1(\xi, t)u(\xi, s)\,d\xi. \tag{6.7}$$

Then the Fredholm determinant $D_1(\lambda)$ *for the kernel* $k_1(x, y) - \varphi_1(x)\varphi_1(y)$ *is*

$$D_1(\lambda) = d_1(\lambda)\left(1 + \lambda \int_0^1 u(t, t)\, dt\right).$$

The theorem follows quite simply from the series solution ([12], p. 228), using the previous notation,

$$u(t, s) = \sum_{j=1}^{\infty} \frac{a_j f_j(t)}{1 - \lambda/\lambda_j} \varphi_1(s)$$

so that $\int_0^1 u(t, t)\, dt = \sum a_j^2/(1 - \lambda/\lambda_j)$. The result follows by using Theorem 6.2.

For the analysis of the present problem, we have, from (6.1), $k(x, y) = \min(x, y) - xy$. In this case it is well known that

$$\lambda_j = \pi^2 j^2, \qquad f_j(x) = \sqrt{2} \sin \pi j x, \qquad d(\lambda) = (\sin \sqrt{\lambda})/\sqrt{\lambda}.$$

Using Theorem 6.2 we have, concerning the limiting ch.fn of C_n^2,

COROLLARY 6.4 *The function* $D(\lambda)$ *of Theorem* 6.1 *is*

$$D(\lambda) = \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}}\left(1 + \lambda \sum_{j=1}^{\infty} \frac{a_j^2}{1 - \lambda/\pi^2 j^2}\right)$$

where $a_j = \sqrt{2} \int_0^1 \varphi(x) \sin \pi j x\, dx$.

It turns out to be possible to replace this expression by a quadrature, and we have

THEOREM 6.5 *The function* $D(\lambda)$ *of Theorem* 6.1 *is*

$$D(\lambda) = \frac{\sin \sqrt{\lambda}}{\sqrt{\lambda}} - 2 \int_0^1 \int_0^t \varphi'(s)\varphi'(t) \cos \sqrt{\lambda}(1 - t) \cos \sqrt{\lambda}\, s\, ds\, dt. \tag{6.8}$$

To prove the theorem we first note that

$$\pi k a_k = \sqrt{2} \int_0^1 \varphi'(t) \cos k\pi t\, dt.$$

Hence, supposing initially that $\lambda \neq \pi^2 k^2$,

$$\begin{aligned} P(\lambda) &= 1 + \lambda \sum \frac{a_k^2}{1 - \lambda/\pi^2 k^2} \\ &= 1 + 2\lambda \int_0^1 \int_0^1 \phi'(t)\varphi'(s) \sum_1^{\infty} \frac{\cos k\pi t \cos k\pi s}{\pi^2 k^2 - \lambda}\, dt\, ds \\ &= 1 + \int_0^1 \int_0^1 \varphi'(t)\varphi'(s)\left\{1 - \frac{\sqrt{\lambda}}{\sin \sqrt{\lambda}} L(t, s)\right\} dt\, ds \\ &= 1 - \frac{\sqrt{\lambda}}{\sin \sqrt{\lambda}} \int_0^1 \int_0^1 \varphi'(t)\varphi'(s) L(t, s)\, dt\, ds, \end{aligned}$$

$$L(t, s) = \begin{cases} \cos \sqrt{\lambda}(1 - t) \cos \sqrt{\lambda}\, s, & t \geqq s; \\ \cos \sqrt{\lambda}(1 - s) \cos \sqrt{\lambda}\, t, & t \leqq s. \end{cases}$$

Hence, since $D(\lambda) = (1/\sqrt{\lambda}) \sin \sqrt{\lambda}\; P(\lambda)$, we obtain the conclusion to the theorem, $L(t, s)$ being symmetric.

By using Theorem 6.3 we can obtain an alternative expression for $D(\lambda)$ in terms of the solution to a differential equation.

THEOREM 6.6 *Regarding* λ *and* h *as parameters, let* $v = v(t, h)$ *be the solution to*

$$v'' + \lambda v = h\varphi''(t), \qquad v(0, h) = v(1, h) = 0. \tag{6.9}$$

Then the function $D(\lambda)$ *of Theorem* 6.1 *is*

$$D(\lambda) = (1/\sqrt{\lambda}) \sin \sqrt{\lambda} \left(1 + \lambda \int_0^1 v(t, \varphi(t))\, dt\right). \tag{6.10}$$

In fact, the integral equation (6.7) is equivalent to the differential equation, with s parametric,

$$v'' + \lambda v = \varphi(s)\varphi''(t),$$

which vanishes at $t = 0$ and $t = 1$. Hence, using Theorem 6.3 we obtain (6.10) immediately.

The result of Theorem 6.5 follows directly from Theorem 6.6, for in solving (6.9) we determine the Green's function $G(\xi, t)$ for the differential expression $v'' + \lambda v$ over the interval $(0, 1)$. This is

$$G(\xi, t) = \begin{cases} \dfrac{\sin \sqrt{\lambda}\xi \sin \sqrt{\lambda}(1 - t)}{\sqrt{\lambda} \sin \sqrt{\lambda}}, & t > \xi; \\[2ex] \dfrac{\sin \sqrt{\lambda}t \sin \sqrt{\lambda}(1 - \xi)}{\sqrt{\lambda} \sin \sqrt{\lambda}}, & t < \xi. \end{cases}$$

Hence $v = h \int_0^1 G(\xi, t)\varphi''(\xi)\, d\xi$, so that (6.10) will give (6.8) after integration by parts.

The above results give various methods of obtaining $D(\lambda)$ and the ch.fn. of C^2, which is $[D(2it)]^{-1/2}$. The distribution can be obtained by the inversion formula

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}[D(2it)]^{-1/2}\, dt,$$

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

where[2] $U(x) = \int_0^x u(y)\, dy$ is the cdf of C^2, and $0 < \lambda_1 < \lambda_2 < \cdots$ are the (simple) zeros of $D(\lambda)$.

7. Properties of the C_n^2 test. A matter of central interest is whether the C_n^2 test has certain distribution-free properties. For example, it is important to know if it is parameter-free, that is, if the limiting distribution of C_n^2 does not

[2] The factor $(-1)^{k-1}$ is missing throughout Smirnov's [14] analysis.

depend on the parameter θ in $F(x; \theta)$. Clearly the test will be parameter-free if and only if $\varphi(u)$ given by (6.2) does not depend on θ; only in this case do we really have a usable test. This turns out to be the case when θ is a location or scale parameter.

THEOREM 7.1 *If $R(x)$ is a* cdf *and if $F(x; \theta) = R(x - \theta)$ or $F(x; \theta) = R(x/\theta)$, for $\theta > 0$, then $\varphi(u)$ is independent of θ and the C_n^2 test is parameter-free.*

To prove the theorem, we need only compute $\varphi(u)$ as given by (6.2). Let $r(x)$ be the density corresponding to $R(x)$. Then, when $F(x; \theta) = R(x - \theta)$, easy calculation shows

$$\varphi(u) = -\left(\int_{-\infty}^{\infty} [r'^2(x)/r(x)]\, dx\right)^{-1/2} r(R^{-1}(u)). \tag{7.1}$$

When $F(x; \theta) = R(x/\theta)$,

$$\varphi(u) = -\left(\int_{-\infty}^{\infty} x^2[r'^2(x)/r(x)]\, dx - 1\right)^{-1/2} R^{-1}(u)r(R^{-1}(u)), \tag{7.2}$$

where $R^{-1}(u)$ is the inverse of $R(x)$.

Another case of interest is $F(x; \theta) = [R(x)]^\theta$ for $\theta > 0$. Here easy computation shows

$$\varphi(u) = u \log u. \tag{7.3}$$

Likewise when $F(x; \theta) = 1 - [1 - R(x)]^\theta$ for $\theta > 0$, we get

$$\varphi(u) = -(1 - u) \log (1 - u).$$

In these latter two cases, $\varphi(u)$ does not even depend on R, so the test is distribution-free over this class of cdf's.

In general, however, $\varphi(u)$ does depend on θ. Consider, for example,

$$F(x; \theta) = \theta R^2(x) + (1 - \theta)R(x), \qquad 0 \leqq \theta \leqq 1.$$

For this, simple calculation gives

$$\varphi(u) = \left(\frac{\tanh^{-1}\theta}{\theta} - 1\right)^{-1/2} \left(u - \frac{-1 + \theta + \sqrt{(1 - \theta)^2 + 4u\theta}}{2\theta}\right).$$

Although not depending on R, this does depend on θ. At $\theta = 0$ we get $\varphi(u) = \sqrt{3}\, u(1 - u)$.

In Section 6 we have given various methods of obtaining the Fredholm determinant $D(\lambda)$ for the kernel $\rho(u, v) = \min (u, v) - uv - \varphi(u)\varphi(v)$. The eigenvalues, λ_k for $k = 1, 2, \cdots$, of this determinant are positive since ρ is a correlation function; that is, the case of one negative root λ_0 of $D(\lambda)$ discussed at the end of the proof of Theorem 6.2 does not occur. For the limiting semi-invariants of C_n^2 we then have [6]

$$K_j = 2^{j-1}(j - 1)!\, \sum 1/\lambda_n^j, \qquad j = 1, 2, \cdots, \tag{7.4}$$

or $K_j = 2^{j-1}(j-1)! \int_0^1 \rho_j(u, u)\, du$, where $\rho_j(u, v)$ is the jth iterate of the kernel $\rho(u, v)$. Thus the mean K_1 and variance K_2 of C^2 are, for example,

$$\mu = K_1 = \int_0^1 \rho(u, u)\, du = \int_0^1 [s(1-s) - \varphi^2(s)]\, ds = 1/6 - \int_0^1 \varphi^2(u)\, du, \tag{7.5}$$

$$\begin{aligned} d^2 = K_2 &= 2 \int_0^1 \int_0^1 [\min(u, v) - uv - \varphi(u)\varphi(v)]^2\, du\, dv \\ &= 1/45 - 8 \int_0^1 \varphi(t)(1-t) \int_0^t s\varphi(s)\, ds\, dt + 2(\mu - 1/6)^2. \end{aligned} \tag{7.6}$$

These expressions may be compared with the corresponding expressions for the case when there is no parameter to be estimated, that is, the W_n^2 test of the hypothesis (1.2). From the remark at the end of Theorem 6.2 we conclude that $\lambda_k \geqq \pi^2 k^2$, so that from (7.4) it follows that *all the semi-invariants (in particular the mean and variance) for* C^2 *are less than the corresponding semi-invariants for* W^2.

Expressing now W^2 and C^2 as a weighted sum of χ^2 variables, we have

$$W^2 = \sum G_j^2/\pi^2 j^2, \qquad C^2 = \sum G_j^2/\lambda_j,$$

where $G_1, G_2, \cdots$ are independent Gaussian random variables with mean 0 and variance 1. All the weights $1/\lambda_j$ in C^2 are less than the corresponding weights $1/\pi^2 j^2$ in W^2; this is the analogue to the reduction of the number of "degrees of freedom" in the usual χ^2 test. It is accounted for by the fact that there is additional freedom in fitting the empirical and theoretical cdf's because of the estimated parameter θ.

In particular, by (7.5) the mean of C^2 is less than the mean for W^2, which is $1/6$, by the factor $\int_0^1 \varphi^2(u)\, du$. This dimensionless quantity gives an idea of the "diameter" of the family $F(x; \theta)$ with respect to θ.

8. Illustrative examples.

A. Location parameter. Let $F(x; \theta)$ be of the form

$$F(x; \theta) = R(x - \theta), \qquad f(x; \theta) = r(x - \theta),$$

for a cdf $R(x)$ whose density is $r(x) = R'(x)$. Then denoting

$$\sigma = \left(\int_{-\infty}^{\infty} [r'^2(x)/r(x)]\, dx \right)^{-1/2}$$

we have by (7.1), $\varphi(u) = -\sigma r(R^{-1}(u))$. Since

$$\varphi(R(y)) = -\sigma r(y), \qquad \varphi'(R(y)) = -\sigma r'(y)/r(y),$$

we get by (7.5), (7.6), and (6.8)

$$\mu = 1/6 - \sigma^2 \int_{-\infty}^{\infty} r^3(x)\, dx,$$

$$d^2 = 1/45 - 8\sigma^2 \int_{-\infty}^{\infty} [1 - R(y)]r^2(y)\, dy \int_{-\infty}^{\infty} R(x)r^2(x)\, dx + 2(\mu - 1/6)^2$$

$$D(\lambda) = (\sin \sqrt{\lambda})/\sqrt{\lambda} - 2\sigma^2 \int_{-\infty}^{\infty} \cos [\sqrt{\lambda}(1 - R(y))]r'(y)\, dy \int_{-\infty}^{y} \cos [\sqrt{\lambda}\, R(x)]r'(x)\, dx.$$

Unfortunately, even in simple cases it does not generally appear possible to evaluate $D(\lambda)$ further, let alone invert the ch.fn. $[D(2it)]^{-1/2}$ of C^2. This seems to be true for the normal distribution, that is, $r(x) = (1/\sqrt{2\pi}) \exp(-\frac{1}{2}x^2)$.

As a special case, let $F(x; \theta)$ be the Cauchy distribution $F(x; \theta) = R(x - \theta)$. Then

$$R(x) = \tfrac{1}{2} + (1/\pi) \tan^{-1} x = u;$$

$$r(x) = \frac{1}{\pi}\frac{1}{1 + x^2}, \qquad x = -\cot \pi u = R^{-1}(u)$$

$$\sigma^2 = 2, \qquad \varphi(u) = (-\sqrt{2}/\pi) \sin^2 \pi u.$$

Consequently $\mu = 1/6 - 1/4\pi^2$, and

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

Another case in which $D(\lambda)$ can be evaluated explicitly is $R(x) = \frac{1}{2} + \frac{1}{2} \tanh x$, in which case $\varphi(u) = \sqrt{3}\, u(1 - u)$.

B. Scale parameter. Let $F(x; \theta) = R(x/\theta)$ for $\theta > 0$, and $r(x) = R'(x)$. Put

$$\sigma = \left(\int_{-\infty}^{\infty} x^2[r'^2(x)/r(x)]\, dx\right)^{-1/2}$$

so that, by (7.2), $\varphi(u) = -\sigma R^{-1}(u)\, r(R^{-1}(u))$. Since

$$\varphi(R(y)) = -\sigma y r(y), \qquad \varphi'(R(y)) = -\sigma(y\, r'(y)/r(y) + 1),$$

we obtain by (7.5), (7.6), and (6.8)

$$\mu = 1/6 - \sigma^2 \int_{-\infty}^{\infty} y^2\, r^3(y)\, dy,$$

$$d^2 = 1/45 - 8\sigma^2 \int_{-\infty}^{\infty} [1 - R(y)]\, y\, r^2(y) \int_{-\infty}^{y} R(x)\, x\, r^2(x)\, dx\, dy + 2(\mu - 1/6)^2,$$

$$D(\lambda) = (\sin \sqrt{\lambda})/\sqrt{\lambda} - 2\sigma^2 \int_{-\infty}^{\infty} \cos \sqrt{\lambda}\, [1 - R(y)]\, [y\, r'(y) + r(y)] \int_{-\infty}^{y} \cos \sqrt{\lambda} R(x)[x r'(x) + r(x)]\, dx\, dy.$$

Consider again the case of the Cauchy distribution, where

$$R(x) = \tfrac{1}{2} + \frac{1}{\pi}\tan^{-1} x = u, \qquad r(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \qquad x = -\cot \pi u.$$

The above formulas give

$$\sigma = \sqrt{2}, \qquad \varphi(u) = (1/\sqrt{2\pi}) \sin 2\pi u,$$

$$\mu = 1/6 - \frac{1}{4\pi^2}, \qquad D(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}}\frac{1}{1-\lambda/4\pi^2}.$$

In this case $D(\lambda)$ has zeros at $\lambda = \pi^2 j^2$, except for $j = 2$. In the representation $C^2 = \sum^* G_j^2/\pi^2 j^2$, where $G_1, G_2, \cdots$ are independent Gaussian random variables with mean 0 and variance 1, the star on the summation means to exclude the term $j = 2$. Thus C^2 and W^2 differ only in that W^2 has this term included. This is a genuine case of a loss of a "degree of freedom", that is, one less term in the sum of the squares.

C. Exponential parameter. Suppose, as before, that $F(x; \theta) = [R(x)]^\theta$ for $\theta > 0$. Then, as in (7.3), $\varphi(u) = u \log u$ and

$$\mu = 1/6 - 2/27 \cong .092593, \qquad d^2 = 1/45 - 88/1125 + 16/324 + 8/729 \cong .0043566,$$

$$D(\lambda) = (\sin\sqrt{\lambda})/\sqrt{\lambda} - 2\int_0^1 (1 + \log t)\cos\sqrt{\lambda}\,(1-t)$$
$$\cdot \int_0^t (1 + \log s)\cos\sqrt{\lambda}\, s\, ds\, dt.$$

9. Further considerations.

A. Incorporation of a weight function. If, instead of the C_n^2 statistic defined by (1.4), we use a weight function $\psi(x) \geqq 0$ as in [6], and define a new test statistic C_n^{*2} by

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

we can carry through a theory paralleling that in the preceding pages. Here $\psi(x)$ must have a bounded first derivative in $0 \leqq x \leqq 1$, and we should be led to the evaluation of the distribution of $C^{*2} = \int Y^2(u)\psi(u)\, du$, where $Y(u)$ is the process defined in Theorem 4.1. In this case we will need to use the results for $D_1(\lambda)$ given in Theorems 6.2 and 6.3, using

$$\rho_1(s, t) = k(s, t)\sqrt{\psi(s)\psi(t)}, \qquad \varphi_1(s) = \varphi(s)\sqrt{\psi(s)},$$

where $k(s, t)$ is given by (3.8) and $\varphi(s)$ is as in (4.4). A method for evaluating the function $d_1(\lambda)$ in Theorems 6.2 and 6.3 is given in [6] for $\rho_1(s, t)$ of this form. Then the ch.fn. for C^{*2} is $[D_1(2it)]^{-1/2}$.

B. A Kolmogoroff statistic. In place of C_n^2 we can consider

$$K_n = \sup_{-\infty<x<\infty} \sqrt{n}\,|F_n(x) - F(x; \hat{\theta}_n)|,$$

following the test proposed by Kolmogoroff (cf. [6]). The limiting distribution of K_n is the same as the distribution of

$$K = \sup_{0<t<1} |Y(t)|,$$

where $Y(t)$ is the process of Theorem 4.1; that is, $Y(u)$ is Gaussian with mean 0 and covariance $\rho(s, t) = \min(s, t) - st - \varphi(s)\varphi(t)$. Unfortunately $Y(u)$ is not Markoffian and the usual methods of determining the distribution of K by diffusion theory will not work. However, in the representation of $Y(u)$ by (4.8), the processes $Y(u)$ and $Z(u)$ are jointly Markoffian and a two dimensional diffusion equation could be used.

Grenander and Rosenblatt [13] were led to an absorption probability for a Gaussian process with covariance $\min(u, v) + g(u)g(v)$ in the estimation of the spectrum of a stationary process. In their case the covariance is the sum of two positive definite kernels, rather than difference, as in $\rho(s, t)$, which simplifies matters considerably in calculating distributions.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] H. Cramér, "On the composition of elementary errors," II *Skand. Aktuarietids.*, Vol. 11 (1928), pp. 141–180.

[3] W. G. Cochran, "The χ^2 test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 315–345.

[4] N. Smirnov, "Sur la distribution de ω^2," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 449–452.

[5] R. von Mises, *Wahrscheinlichkeitsrechnung*, Deuticke, Leipzig and Wien (1931).

[6] 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.

[7] Z. W. Birnbaum, "Distribution free tests of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 1–8.

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

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

[10] M. Kac and A. J. F. Siegert, "An explicit representation of a stationary Gaussian process," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 438–442.

[11] R. Fortet, "Random functions from a Poisson process," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 373–385.

[12] E. T. Whitaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, 1927.

[13] U. Grenander and M. Rosenblatt, "Statistical spectral analysis of time series arising from stationary stochastic processes," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 537–558. (See also "On spectral analysis of stationary time series," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 38 (1952) pp. 519–521.)

[14] N. V. Smirnov, "On the distribution of the ω^2 criterion," (Russian), *Rec. Math.*, Vol. 2 (1937), pp. 973–993.