The Annals of Statistics
1973, Vol. 1, No. 2, 279–290

WEAK CONVERGENCE OF THE SAMPLE DISTRIBUTION FUNCTION WHEN PARAMETERS ARE ESTIMATED

BY J. DURBIN

London School of Economics and Political Science

The weak convergence of the sample df is studied under a given sequence of alternative hypotheses when parameters are estimated from the data. For a general class of estimators it is shown that the sample df, when normalised, converges weakly to a specified normal process. The results are specialised to the case of efficient estimation.

1. Introduction. Suppose we have a sample of independent observations $x_1, \cdots, x_n$ from a distribution with continuous df $F(x, \theta)$ depending on a vector θ of p parameters. Let $F_{n0}(t)$ denote the proportion of the values $x_1, \cdots, x_n$ for which $F(x_i, \theta_0) \leqq t$, $0 \leqq t \leqq 1$. It is well known that when $\theta = \theta_0$ the sample process $n^{\frac{1}{2}}\{F_{n0}(t) - t\}$ converges weakly to the tied-down Brownian motion process. In this paper we extend this result to the case where a set of nuisance parameters is estimated from the sample and where the remaining parameters take values specified by a given sequence of alternative hypotheses.

Let $\theta = [\theta_1', \theta_2']'$ where θ_1 is a vector of p_1 parameters specified as equal to θ_{10} on the null hypothesis H_0 and where θ_2 is a vector of p_2 parameters whose value is unknown and is to be estimated from the data. Let $\hat{\theta}_{2n}$ be an estimator of θ_2 from a class to be specified later, let $\hat{\theta}_n = [\theta_{10}', \hat{\theta}_{2n}']'$ and let $\hat{F}_n(t)$ denote the estimated sample df, i.e. the proportion of the values $x_1, \cdots, x_n$ for which $F(x_i, \hat{\theta}_n) \leqq t$, $0 \leqq t \leqq 1$.

We shall study the weak convergence of the estimated sample process $\hat{y}_n(t) = n^{\frac{1}{2}}\{\hat{F}_n(t) - t\}$ under the sequence of alternative hypotheses $H_n: \theta_1 = \theta_{1n} = \theta_{10} + \gamma n^{-\frac{1}{2}}$, $n = m, m+1, \cdots$ where γ is a given vector and m is a given positive integer. Under conditions to be stated, Theorem 1 proves that $\hat{y}_n$ converges weakly to a normal process with mean function (4) and covariance function (5) below. Theorem 2 specialises this result to the case of efficient estimation.

These results are needed for the derivation of the asymptotic distributions of statistics proposed for tests of goodness of fit of composite hypotheses and for studying the asymptotic powers of these tests against alternative hypotheses of interest. For example, the results will be used elsewhere for studying the construction and performance of tests of normality and exponentiality based on statistics of Kolmogorov–Smirnov and Cramér–von Mises types.

The relation of the results obtained to previous work by Chernoff and Lehmann (1954), Darling (1955), Kac, Kiefer and Wolfowitz (1955), Roy (1956), Barton

Received March 1971; revised July 12, 1972.

[1] Visiting Fellow in the Department of Statistics, Institute of Advanced Studies, Australian National University when this paper was written.

(1956), Watson (1957, 1958, 1959), Chibisov (1965) and Moore (1971) is discussed in Section 6.

2. Preliminaries. The exposition will be based freely on the techniques and results of Billingsley (1968), referred to henceforth as **B**. We shall be concerned with the convergence of sequences of constant and random elements of various metric spaces. Where possible without ambiguity, a real or vector-valued function $z(t)$, $0 \leqq t \leqq 1$, will be denoted simply by z. The special functions $z(t) = t$, $z(t) = \hat{t}_n(t)$ and $z(t) = 0$, $0 \leqq t \leqq 1$, will however be denoted in bold type by $\mathbf{t}$, $\hat{\mathbf{t}}_n$ and $\mathbf{0}$.

Our basic spaces will be the space C of continuous functions on [0, 1], the space D of right-continuous functions with left-hand limits on [0, 1] and Euclidean spaces E^k for various k. On C we use the uniform metric c defined for a pair of elements x and y by

$$c(x, y) = \sup_{0 \leqq t \leqq 1} |x(t) - y(t)| ,$$

on D we use the Skorokhod metric $d = d(x, y)$ defined on page 111 of **B** and on E^k we use the metric defined for a pair of elements $x = [x_1, \cdots, x_k]'$, $y = [y_1, \cdots, y_k]'$ by $e(x, y) = \max_{i=1,\cdots,k} |x_i - y_i|$, $k = 1, 2, \cdots$. This metric is also used for matrices by writing the elements of a matrix as a vector. If $x = \{[x_1(t), \cdots, x_k(t)]' \; 0 \leqq t \leqq 1\}$ is a vector-valued function where each $x_i \in C$ we use the metric $\max_{i=1,\cdots,k} \{\sup_{0 \leqq t \leqq 1} |x_i(t) - y_i(t)|\}$.

The terms weak convergence and convergence in distribution are used in the following sense. Let X_n be a measurable mapping from a probability space $(\Omega, \mathscr{B}, \mu)$ to a metric space (S, ρ) for $n = 0, 1, 2, \cdots$. If $\lim_{n\to\infty} E\{f(X_n)\} = E\{f(X_0)\}$ for all bounded real functions f which are continuous in the metric ρ we say that X_n converges weakly or converges in distribution to X_0 and we write $X_n \to_{\mathscr{D}} X_0$. The notation $\to_P$ indicates convergence in probability. The σ-field of (S, ρ) under consideration is that generated by its open sets.

When x is a scalar or scalar-valued function and α is a vector $[\alpha_1, \cdots, \alpha_k]'$ we write $\partial x/\partial\alpha$ for the vector or vector-valued function $[\partial x/\partial\alpha_1, \cdots, \partial x/\partial\alpha_k]'$.

3. Asymptotic distribution of $\hat{y}_n$ for general $\hat{\theta}_{2n}$. Suppose that the vector θ_2 of nuisance parameters is estimated by $\hat{\theta}_{2n}$ and that we wish to consider the construction and performance of tests of the null hypothesis

$$H_0 : \theta_1' = \theta_{10} \tag{1}$$

based on the estimated sample process

$$\hat{y}_n(t) = n^{\frac{1}{2}}\{\hat{F}_n(t) - t\} , \qquad 0 \leqq t \leqq 1 , \tag{2}$$

where $\hat{F}_n(t)$ is the proportion of $x_1, \cdots, x_n$ for which $F(x_i, \hat{\theta}_n) \leqq t$ with $\hat{\theta}_n = [\theta_{10}', \hat{\theta}_{2n}']'$. Our objective is to study the limiting distribution of $\hat{y}_n$ under a sequence of alternative hypotheses $\{H_n\}$ defined as follows. Let θ_{20} be a specified value of θ_2, let ν denote the closure of a given neighbourhood of $\theta_0 = [\theta_{10}', \theta_{20}']'$ and let $m = \min\{k : [\theta_{10}' + \gamma n^{-\frac{1}{2}}, \theta_2']' \in \nu$ for all $n \geqq k > 2\}$ where γ is a given constant vector. Then take

(3) $$H_n : \theta = \theta_n = \begin{bmatrix} \theta_{1n} \\ \theta_{20} \end{bmatrix} \quad \text{with} \quad \theta_{1n} = \theta_{10} + \gamma n^{-\frac{1}{2}}, \qquad n = m, m+1, \cdots .$$

The limiting distribution of $\hat{y}_n$ on H_0 with $\theta_2 = \theta_{20}$ can be obtained by putting $\gamma = 0$.

We assume that $\hat{\theta}_{2n}$ satisfies the assumptions

(A1) $$n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) = \frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^{n} l(x_i, \theta_n) + A\gamma + \varepsilon_{1n}$$

where l is measurable and, for a random observation x,

(i) $$E\{l(x, \theta_n) \mid \theta = \theta_n\} = 0 , \qquad n \geqq m ,$$

(ii) $E\{l(x, \theta_n)l(x, \theta_n)' \mid \theta = \theta_n\} = L(\theta_n)$, a finite nonnegative-definite matrix for each $n \geqq m$ which converges to a finite nonnegative-definite matrix $L = L(\theta_0)$ as $n \to \infty$, and where

(iii) A is a given finite matrix of order $(p_2 \times p_1)$,

(iv) $\varepsilon_{1n} \to_P 0$.

Estimators satisfying these assumptions exist, e.g. where F is $N(\theta_2, \theta_1)$, $\theta_{20} = 0$, $l(x_i, \theta_n) = x_i$, $A = 0$ and $\varepsilon_{1n} = 0$. At first sight it may appear strange that we have chosen to make the estimating function $l(x, \theta_n)$ depend on θ_n since in practice the estimator $\hat{\theta}_{2n}$ would not, as a function of the observations, depend on the alternative hypothesis. However, it would in any case be necessary to consider the mean of $n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20})$ on H_n and this would generally, to the first order, take the form $A\gamma$. Given this, it is mathematically convenient to separate off the part with zero mean, as is done in (A1). It should be emphasized that this formulation has been adopted to facilitate consideration of distributions under H_n. If null-hypothesis distributions only are of interest one takes $\theta_n = \theta_0$ and $\gamma = 0$ in (A1) and the point does not arise.

The following assumptions will be made in addition to (A1).

(A2) (i) $F(x, \theta)$ is continuous in x for all $\theta \in \nu$.

(ii) The vector-valued function $g(t, \theta)$ defined by (6) exists and is continuous in (t, θ) for all $\theta \in \nu$ and all $0 \leqq t \leqq 1$.

Let $x(t, \theta) = \inf\{x : F(x, \theta) = t\}$ be the inverse of the transformation $t = F(x, \theta)$. Our main result is

THEOREM 1. *On Assumptions* (A1), (A2) *and under the sequence of alternatives* $\{H_n\}$, $\hat{y}_n$ *converges weakly to the normal process* $z(t)$, $0 \leqq t \leqq 1$, *in D with mean function*

(4) $$E\{z(t)\} = \gamma'\{g_1(t) - A'g_2(t)\}$$

and covariance function

(5) $$C\{z(t_1), z(t_2)\} \\ = \min(t_1, t_2) - t_1 t_2 - h(t_1)'g_2(t_2) - h(t_2)'g_2(t_1) + g_2(t_1)'Lg_2(t_2)$$

where

$$(6)\qquad \begin{bmatrix} g_1(t) \\ g_2(t) \end{bmatrix} = g(t) = g(t, \theta_0) \qquad \textit{with} \quad g(t, \theta) = \frac{\partial F(x, \theta)}{\partial \theta}\bigg|_{x=x(t,\theta)}$$

and

$$(7)\qquad h(t) = h(t, \theta_0) \qquad \textit{with} \quad h(t, \theta) = \int_{-\infty}^{x(t,0)} l(x, \theta)\, dF(x, \theta)\,.$$

From this theorem we deduce two corollaries, the first of which gives the limiting distribution under H_0 and the second of which gives the limiting distribution when the estimator $\hat{\theta}_{2n}$ is "superefficient" (L is a null matrix, i.e. a matrix of zeros) or θ_{20} is known exactly ($l(x_i, \theta_n) = 0$, A is a null matrix and $\varepsilon_{1n} = 0$).

COROLLARY 1. *Under the null hypothesis H_0, but otherwise under the conditions of Theorem 1, $\hat{y}_n \rightarrow_{\mathscr{L}} z$ where z is distributed as in the statement of Theorem 1 with $\gamma = 0$.*

COROLLARY 2. *When $n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) \rightarrow_P A\gamma$ and Conditions (A2) are satisfied, $\hat{y}_n \rightarrow_{\mathscr{L}} z_0$ where z_0 is the normal process in D with mean function (4) and covariance function*

$$(8)\qquad C\{z_0(t_1), z_0(t_2)\} = \min(t_1, t_2) - t_1 t_2\,;$$

A may be taken to be a null matrix.

The proof of Theorem 1 will be based on the following five lemmas.

LEMMA 1. *Let $F_n(t)$ be the proportion of $x_1, \cdots, x_n$ for which $F(x_i, \theta_n) \leqq t$, let*

$$(9)\qquad y_n(t) = n^{\frac{1}{2}}\{F_n(t) - t\}\,, \qquad 0 \leqq t \leqq 1\,,$$

and let $\hat{t}_n(t) = F(x(t, \hat{\theta}_n), \theta_n)$. Then

$$(10)\qquad y_n(\hat{t}_n(t)) = y_n(t) + \varepsilon_{2n}(t)\,, \qquad 0 \leqq t \leqq 1\,,$$

where $\varepsilon_{2n} \rightarrow_P \mathbf{0}$.

PROOF. Let $x_a(t) = x(t, \theta_a)$ and $\hat{t}(t) = F(x_a(t), \theta_b)$ for $\theta_a, \theta_b \in \nu$ and $0 \leqq t \leqq 1$. Then

$$\sup_t |\hat{t}(t) - t| = \sup_{x_a} |F(x_a, \theta_b) - F(x_a, \theta_a)|$$
$$= \sup_{x_a} \left| (\theta_b - \theta_a)' \frac{\partial F(x, \theta)}{\partial \theta}\bigg|_{\substack{x = x_a \\ \theta = \theta^*}} \right|$$

where θ^* is between θ_a and θ_b. By (A2) (ii) every component of $|\partial F(x, \theta)/\partial \theta| < M$ for all x and $\theta \in \nu$ for some M since $\partial F/\partial \theta = g(t, \theta)$ and this is continuous on the compact set $\theta \in \nu$, $0 \leqq t \leqq 1$. Hence $\hat{t}(t)$ converges uniformly to t as $\theta_b \rightarrow \theta_0$ and $\theta_a \rightarrow \theta_0$. Thus $\sup_t |\hat{t}_n(t) - t| \rightarrow 0$ as $\hat{\theta}_n \rightarrow \theta_0$ and $\theta_n \rightarrow \theta_0$. Using a slight modification of the usual proof of the multivariate central limit for independent and identically distributed random vectors (e.g. as in Breiman (1968) page 238) to allow for the fact that the variance matrix of $n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20})$ depends on n we deduce from Assumptions (A1) that $n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) \rightarrow_{\mathscr{L}} N(A\gamma, L)$. Thus $\hat{\theta}_{2n} \rightarrow_P \theta_{20}$ so $\hat{\theta}_n \rightarrow_P \theta_0$. Consequently $\sup_t |\hat{t}_n(t) - t| \rightarrow_P 0$, i.e. $\hat{\mathbf{t}}_n \rightarrow_P \mathbf{t}$.

Now $y_n \to_{\mathscr{L}} y$ where y is the tied-down Brownian motion process, i.e. the zero-mean normal process in D with covariance function $E[y(t_1)y(t_2)] = \min(t_1, t_2) - t_1 t_2$ (B, page 141). Since $y_n \to_{\mathscr{L}} y$, $\hat{\mathbf{t}}_n \to_P \mathbf{t}$ and (D, d) is separable, Theorem 4.4 of B implies that $(y_n, \hat{\mathbf{t}}_n) \to_{\mathscr{L}} (y, \mathbf{t})$ where $(y_n, \hat{\mathbf{t}}_n)$ and $(y, \mathbf{t})$ are random elements in the product space $D \times D$. Moreover, the mapping from $(y_n, \hat{\mathbf{t}}_n)$ in $D \times D$ to ε_{2n} in D is continuous in the product topology determined by metric c at the point $(y_c, \mathbf{t})$ where y_c is any continuous element of D. This is proved as follows. Given ε, choose δ so that $\delta < \varepsilon/3$ and so that $c(\hat{\mathbf{t}}_n, \mathbf{t}) < \delta$ implies $\sup_t |y_c(\hat{t}_n(t)) - y_c(t)| < \varepsilon/3$. Then for any y_n satisfying $c(y_n, y_c) < \delta$ we have

$$\begin{aligned}\sup_t |\varepsilon_{2n}(t)| &= \sup_t |y_n(\hat{t}_n(t)) - y_n(t)| \\ &\leqq \sup_t [|y_n(\hat{t}_n(t)) - y_c(\hat{t}_n(t))| + |y_c(\hat{t}_n(t)) - y_c(t)| + |y_n(t) - y_c(t)|] \\ &\leqq 2\sup_t |y_n(t) - y_c(t)| + \sup_t |y_c(\hat{t}_n(t)) - y_c(t)| \\ &\leqq 2\delta + \varepsilon/3 < \varepsilon \qquad \text{which proves continuity.}\end{aligned}$$

Since y is continuous with probability one and $\mathbf{t}$ is continuous, and since convergence in metric d to a continuous element is uniform, the mapping $(y_n, \hat{\mathbf{t}}_n) \to \varepsilon_{2n}$ from $D \times D$ to D is continuous with probability one in the product topology determined by metric d with respect to the distribution of $(y, \mathbf{t})$. Hence by Theorem 5.1, Corollary 1, of B, $\varepsilon_{2n} \to_{\mathscr{L}} y - y = \mathbf{0}$ and this in turn implies $\varepsilon_{2n} \to_P 0$, which proves the lemma. This argument has been based on the treatment of random change of time on pages 144–145 of B where further discussion of measurability considerations can be found.

LEMMA 2. *Let*

$$z_n(t) = y_n(t) + \gamma'\{g_1(t) - A'g_2(t)\} - w_n'g_2(t), \qquad 0 \leqq t \leqq 1,\ n \geqq m, \tag{11}$$

where

$$w_n = \frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^n l(x_i, \theta_n). \tag{12}$$

Then

$$\hat{y}_n = z_n + \varepsilon_{3n} \qquad \textit{where} \quad \varepsilon_{3n} \to_P \mathbf{0}.$$

PROOF. Let $x_0(t) = x(t, \hat{\theta}_n)$, $0 \leqq t \leqq 1$. Then $\hat{F}_n(t)$ is the proportion of $x_1, \cdots, x_n \leqq x_0(t)$. Now $\hat{t}_n(t) = F(x(t, \hat{\theta}_n), \theta_n) = F(x_0(t), \theta_n)$. Thus $F_n(\hat{t}_n(t))$ is the proportion of $x_1, \cdots, x_n \leqq x_0(t)$. Consequently $\hat{F}_n(t) = F_n(\hat{t}_n)$, on suppressing the argument t of $\hat{t}_n(t)$, so that

$$\begin{aligned}\hat{y}_n(t) &= n^{\frac{1}{2}}\{\hat{F}_n(t) - t\} \\ &= n^{\frac{1}{2}}\{F_n(\hat{t}_n) - \hat{t}_n\} + n^{\frac{1}{2}}(\hat{t}_n - t), \qquad \text{i.e.} \\ \hat{y}_n(t) &= y_n(\hat{t}_n) + n^{\frac{1}{2}}(\hat{t}_n - t).\end{aligned} \tag{13}$$

Suppressing the argument t of $x_0(t)$, for $\hat{\theta}_n \in \nu$ we have

$$\begin{aligned}n^{\frac{1}{2}}(\hat{t}_n - t) &= n^{\frac{1}{2}}\{F(x_0, \theta_n) - F(x_0, \hat{\theta}_n)\} \\ &= n^{\frac{1}{2}}(\theta_n - \hat{\theta}_n)' \frac{\partial F(x, \theta)}{\partial \theta}\bigg|_{\substack{x = x_0 \\ \theta = \theta_n^*}}\end{aligned} \tag{14}$$

where θ_n^* is between θ_n and $\hat{\theta}_n$. Let $t_n^* = F(x_0, \theta_n^*)$. Then

$$\frac{\partial F(x, \theta)}{\partial \theta}\bigg|_{\substack{x=x_0 \\ \theta=\theta_n^*}} = \frac{\partial F(x, \theta)}{\partial \theta}\bigg|_{\substack{x=x(t_n^*, \theta_n^*) \\ \theta=\theta_n^*}} = g(t_n^*, \theta_n^*) .$$

By the proof of Lemma 1, on taking $\theta_a = \hat{\theta}_n$ and $\theta_b = \theta_n^*$ it follows that $t_n^* = F\{x(t, \hat{\theta}_n), \hat{\theta}_n^*\}$ converges to $F\{x(t, \theta_0), \theta_0\} = t$ uniformly in t as $\hat{\theta}_n \to \theta_0$ and $\theta_n^* \to \theta_0$. But $(\hat{\theta}_n, \theta_n^*) \to_P (\theta_0, \theta_0)$. Hence on writing $\mathbf{t}_n^*$ for $(t_n^*, 0 \leqq t \leqq 1)$ we have $\mathbf{t}_n^* \to_P \mathbf{t}$.

For a typical component function $g_a(t_n^*, \theta_n^*)$ of $g(t_n^*, \theta_n^*)$ we have

$$\sup_t |g_a(t_n^*, \theta_n^*) - g_a(t, \theta_0)| \leqq \sup_t |g_a(t_n^*, \theta_n^*) - g_a(t_n^*, \theta_0)| + \sup_t |g_a(t_n^*, \theta_0) - g_a(t, \theta_0)| .$$

Now Assumption (A2) (ii) implies that $g(t, \theta)$ is continuous in θ at $\theta = \theta_0$ uniformly in t so $\sup_t |g_a(t_n^*, \theta_n^*) - g_a(t_n^*, \theta_0)| = \sup_t g_a(t, \theta_n^*) - g_a(t, \theta_0)| \to_P 0$ as $\theta_n^* \to_P \theta_0$. Assumption (A2) (ii) also implies that $g(t, \theta)$ is uniformly continuous in t at $\theta = \theta_0$ so $\sup_t |g_a(t_n^*, \theta_0) - g_a(t, \theta_0)| \to_P 0$ as $\mathbf{t}_n^* \to_P \mathbf{t}$. Since $\theta_n^* \to_P \theta_0$ and $\mathbf{t}_n^* \to_P \mathbf{t}$ it follows that $\sup_t |g_a(t_n^*, \theta_n^*) - g_a(t, \theta_0)| \to_P 0$ and hence that

$$g(t_n^*, \theta_n^*) = g(t) + \varepsilon_{4n}(t) \tag{15}$$

where $\varepsilon_{4n} \to_P \mathbf{0}$.

We also have

$$n^{\frac{1}{2}}(\theta_n - \hat{\theta}_n) = n^{\frac{1}{2}} \begin{bmatrix} \theta_{1n} - \theta_{10} \\ \theta_{20} - \hat{\theta}_{2n} \end{bmatrix} = \begin{bmatrix} \gamma \\ -n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) \end{bmatrix} \quad \text{for } n \geqq m \tag{16}$$

where

$$n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) = w_n + A\gamma + \varepsilon_{1n}$$

from (A1) and (12).

Substituting from (16), (15), (14), and (10) in (13) we obtain

$$\begin{aligned} \hat{y}_n(t) &= y_n(t) + \varepsilon_{2n}(t) + [\gamma', -w_n' - \gamma'A' - \varepsilon_{1n}']\{g(t) + \varepsilon_{4n}(t)\} \\ &= z_n(t) + \varepsilon_{3n}(t) \end{aligned}$$

where

$$\varepsilon_{3n}(t) = \varepsilon_{2n}(t) - \varepsilon_{1n}' g_2(t) + \gamma' \varepsilon_{5n}(t) - (w_n' + \gamma' A')\varepsilon_{6n}(t) - \varepsilon_{1n}' \varepsilon_{6n}(t)$$

with $\varepsilon_{4n}(t) = [\varepsilon_{5n}(t)', \varepsilon_{6n}(t)']'$. Now if u_n is any element of D, $d(u_n, \mathbf{0}) = c(u_n, \mathbf{0}) = \sup_t |u_n(t)|$. Thus $u_n \to_P \mathbf{0}$ if and only if $\sup_t |u_n(t)| \to_P 0$. We use this fact to prove that $\varepsilon_{3n} \to_P \mathbf{0}$. By Lemma 1 we have $\varepsilon_{2n} \to_P \mathbf{0}$. Since $g_2(t)$ is continuous for $0 \leqq t \leqq 1$ and therefore bounded and $\varepsilon_{1n}' \to_P 0$ we have $\varepsilon_{1n}' g_2(t) \to_P \mathbf{0}$. Also $w_n \to_{\mathscr{D}} w$ where w is $N(0, L)$ and $\varepsilon_{6n} \to_P \mathbf{0}$; thus $w_n' \varepsilon_{6n} \to_P \mathbf{0}$. Similarly, each of the remaining terms of $\varepsilon_{3n} \to_P \mathbf{0}$. Thus ε_{3n} is the sum of a number of functions each of which $\to_P \mathbf{0}$. Consequently $\varepsilon_{3n} \to_P \mathbf{0}$.

The remaining step in the proof of Theorem 1 is to show that z_n converges weakly to the normal process with mean and covariance functions (4) and (5). It might appear at first sight that this could be done by proving that z_n is a continuous function of y_n and using Theorem 5.1 of B. However, it turns out that w_n, and hence z_n, is discontinuous in y_n in metric d, and indeed in metric c, for estimators important in practice such as the sample mean from a normal distribution. Consequently the proof cannot be based on a continuity argument. Darling's (1955) proof of convergence of the Cramér–von Mises statistic seems faulty at this point, notwithstanding the "auxiliary assumption" on page 9 of his paper. We shall therefore use instead the basic technique for proving weak convergence, i.e. first prove convergence of the finite-dimensional distributions and then prove tightness.

LEMMA 3. *For z_n defined by* (11) *and for all* $0 < t_1 < \cdots < t_k < 1$,

$$[z_n(t_1), \cdots, z_n(t_k)]' \to_P [z(t_1), \cdots, z(t_k)]'$$

where z is a normal process in D with mean and covariance functions (4) *and* (5).

PROOF. Let $d_{ij} = 1 - t_i$ if $F(x_j, \theta_n) \leqq t_i$ and $d_{ij} = -t_i$ if $F(x_j, \theta_n) > t_i$, $i = 1, \cdots, k$ and $j = 1, \cdots, n$. Then $y_n(t_i) = n^{-\frac{1}{2}} \sum_{j=1}^n d_{ij}$, $E(d_{ij}) = 0$, $E(d_{ij} d_{i'j}) = \min(t_i, t_{i'}) - t_i t_{i'}$ and d_{ij}, $d_{i'j'}$ are independent for $j \neq j'$. Let $c_{ij} = d_{ij} - \gamma'\{g_1(t_i) - A' g_2(t_i)\} - l(x_j, \theta_n) g_2(t_i)$. Then $z_n(t_i) = n^{-\frac{1}{2}} \sum_{j=1}^n c_{ij}$, $E(c_{ij}) = \gamma'\{g_1(t_i) - A' g_2(t_i)\}$ and

$$C(c_{ij}, c_{i'j}) = \min(t_i, t_{i'}) - t_i t_{i'} - h(t_i, \theta_n)' g_2(t_{i'}) - h(t_{i'}, \theta_n) g_2(t_i) + g_2(t_i)' L(\theta_n) g_2(t_{i'})$$

since $E\{d_{ij} l(x_j, \theta_n)\} = \int_{-\infty}^{x(t_i, \theta_n)} l(x, \theta_n)\, dF(x, \theta_n) = h(t_i, \theta_n)$; also c_{ij} and $c_{i'j'}$ are independent for $j \neq j'$. Thus $[z_n(t_1), \cdots, z_n(t_k)]'$ is the standardised sum of independent and identically distributed vectors with mean given by (4) and variance matrix converging to the form given by (5). The lemma then follows by a multivariate central-limit theorem of Lindberg–Lévy type obtained by slightly modifying the usual proof, e.g. that given on page 238 of Breiman (1968), to allow for the fact that the variance matrix of $[z_n(t_1), \cdots, z_n(t_k)]'$ depends on n but converges to a positive-definite limit.

Note that the existence of $h(t, \theta_n)$ for $0 < t < 1$ follows from the existence of $h(1, \theta_n) = E\{l(x, \theta_n)\}$ which we assumed to be zero in (A1) (cf. Apostol (1957) Theorems 9–21). The existence of a process z in D with the distribution specified follows from Lemma 4.

In the following lemma $P\{A\}$ denotes the probability of an event A.

LEMMA 4. *The sequence $\{z_n\}$ of random elements is tight.*

PROOF. Paraphrasing Theorem 15.5 of B, tightness follows if

(i) for each positive η there exists an a such that

$$P\{|z_n(0)| > a\} \leqq \eta\,, \qquad n \geqq m\,,$$

where m is defined between (2) and (3),

(ii) for each positive ε and η, there exist a δ, with $0 < \delta < 1$, and an integer

n_0 such that

$$P\{w(z_n, \delta) \geqq \varepsilon\} \leqq \eta\,, \qquad n \geqq n_0\,,$$

where $w(x, \delta)$ is the modulus of continuity defined by

$$w(x, \delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|\,, \qquad 0 < \delta < 1\,.$$

The proof of (i) is immediate since $z_n(0) = y_n(0) + \gamma' g_1(0) + w_n' g_2(0)$ where $y_n(0) = 0$ with probability one, $g_1(0) = g_2(0) = 0$, and w_n has a finite variance matrix by (A1) (ii).

Putting $\gamma' = [\gamma_1, \cdots, \gamma_{p_1}]$, $g_1(t)' - g_2(t)'A = [g_{11}(t), \cdots, g_{1p_1}(t)]$, $w_n' = [w_{n1}, \cdots, w_{np_2}]$ and $g_2(t)' = [g_{21}(t), \cdots, g_{2p_2}(t)]$ we have

$$(17) \qquad |z_n(s) - z_n(t)| \leqq |y_n(s) - y_n(t)| + \sum_{i=1}^{p_1} |\gamma_i|\, |g_{1i}(s) - g_{1i}(t)| + \sum_{i=1}^{p_2} |w_{ni}|\, |g_{2i}(s) - g_{21}(t)|\,.$$

Following **B**, page 104, let $t_i' = F(x_i, \theta_n)$, $i = 1, \cdots, n$, $t_0' = 0$, $t_{n+1}' = 1$, and let $G_n(t)$ be, as a function of t ranging from 0 to 1, the distribution function corresponding to a uniform distribution of mass $(n + 1)^{-1}$ over each of the intervals (t_{i-1}', t_i'), $i = 1, \cdots, n + 1$. Let $Y_n(t) = n^{\frac{1}{2}}\{G_n(t) - t\}$, $0 \leqq t \leqq 1$. Then

$$(18) \qquad \sup_t |Y_n(t) - y_n(t)| \leqq \frac{1}{n^{\frac{1}{2}}}\,.$$

The tightness of the sequence $\{Y_n\}$ is demonstrated in the proof of Theorem 13.1 of **B**. By Theorem 8.2 of **B** this implies that for each positive ε' and η' there exist a δ', $0 < \delta' < 1$, and an integer n_0' such that

$$P\{w(Y_n, \delta') \geqq \varepsilon'\} \leqq \eta'\,, \qquad n \geqq n_0'\,.$$

Using (18) this implies that given ε, η there exist δ_1, n_1 such that

$$P\{|y_n(s) - y_n(t)| \geqq \varepsilon/3\} \leqq \eta/2$$

for $n \geqq n_1$ and all $|s - t| < \delta_1$.

Now $g_{1i}(t)$ is continuous by (A2) and therefore uniformly continuous, so given $\varepsilon > 0$, $\delta_2 > 0$ exists such that $|\gamma_i|\, |g_{1i}(s) - g_{1i}(t)| < \varepsilon/(3p_1)$ for $i = 1, \cdots, p_1$, $n \leqq n_2$ and all $|s - t| < \delta_2$, $0 \leqq s, t \leqq 1$. Similarly $g_{2i}(t)$ is uniformly continuous and $V(w_{ni})$ converges to the (ii)th element of L which is finite by (A1) (ii). Thus if $\eta > 0$ is given then n_3 and $\delta_3 > 0$ exist such that

$$P\{|w_{ni}|\, |g_{2i}(s) - g_{2i}(t)| \geqq \varepsilon/(3p_2)\} \leqq \eta/(2p_2) \qquad \text{for} \quad i = 1, \cdots, p_2,\ n \geqq n_3$$

and all $|s - t| < \delta_3$. Let $n_0 = \max(n_1, n_3)$ and $\delta = \min(\delta_1, \delta_2, \delta_3)$; the above statements then remain valid when n_1, n_3 are each replaced by n_0 and δ_1, δ_2, δ_3 are each replaced by δ.

Let E_0 denote the event $|z_n(s) - z_n(t)| \geqq \varepsilon$ and $E_1, \cdots, E_{p+1}$ the events

$$|y_n(s) - y_n(t)| \geqq \varepsilon/3\,,$$
$$|\gamma_1|\, |g_{11}(s) - g_{11}(t)| \geqq \varepsilon/(3p_1), \cdots, |w_{np2}|\, |g_{2p_2}(s) - g_{2p_2}(t)| \geqq \varepsilon/(3p_2)\,.$$

Let E_i^c denote the event complementary to E_i, $i = 0, \cdots, p + 1$. From (17)

$\bigcap_{i=1}^{p+1} E_i^c$ implies E_0^c so $P(E_0^c) \geqq P(\bigcap_{i=1}^{p+1} E_i^c) = 1 - P(\bigcup_{i=1}^{p+1} E_i)$. Hence $P(E_0) \leqq P(\bigcup_{i=1}^{p+1} E_i) \leqq \sum_{i=1}^{p+1} P(E_i) \leqq \eta/2 + p_2\eta/(2p_2) = \eta$ for $n \geqq n_0$ and all $|s - t| < \delta$. This proves (ii).

LEMMA 5. $z_n \rightarrow_{\mathscr{D}} z$.

PROOF. This follows immediately from Lemmas 3 and 4 using Theorem 15.1 of **B**.

PROOF OF THEOREM 1. Since $d(\hat{y}_n, z_n) \leqq c(\hat{y}_n, z_n) = \sup_t |\varepsilon_{3n}(t) = d(\varepsilon_{3n}, \mathbf{0}) \rightarrow_P 0$ by Lemma 2, it follows that $d(\hat{y}_n, z_n) \rightarrow_P 0$. Using Lemma 5, the result $\hat{y}_n \rightarrow_{\mathscr{D}} z$ then follows from Theorem 4.1 of **B**.

4. Efficient estimation. We now define the concept of an efficient estimator and apply to it the results of the previous section. We shall make the following further assumptions.

(A3) (i) $F(x, \theta)$ has a density $f(x, \theta)$ such that for each n and almost all x, the vector $\partial \log f(x, \theta_n)/\partial\theta_2$ of partial derivatives of $\log f(x, \theta)$ with respect to the elements of θ_2, evaluated at $\theta = \theta_n$, exists and satisfies

$$E\left[\frac{\partial \log f(x, \theta_n)}{\partial \theta_2}\Big| \theta = \theta_n\right] = 0 .$$

(ii) For each n the $p_2 \times p_2$ matrix

$$\mathscr{I}(\theta_n) = E\left[\frac{\partial \log f(x, \theta_n)}{\partial \theta_2}\frac{\partial \log f(x, \theta_n)}{\partial \theta_2'}\Big| \theta = \theta_n\right]$$

is finite and positive definite and converges to the finite positive-definite matrix $\mathscr{I} = \mathscr{I}(\theta_0)$ as $n \rightarrow \infty$. The matrix

$$\mathscr{I}_{21} = E\left[\frac{\partial \log f(x, \theta)}{\partial \theta_2}\frac{\partial \log f(x, \theta)}{\partial \theta_1'}\Big| \theta = \theta_0\right]$$

is finite.

(iii)
$$\frac{\partial F(x, \theta)}{\partial \theta_2} = \int_{-\infty}^{x} \frac{\partial f(y, \theta)}{\partial \theta_2}\, dy$$

for all x when $\theta = \theta_0$.

When (A3) holds we define an *efficient estimator* of θ_{20} relative to the sequence of alternatives $\{H_n\}$ as an estimator $\hat{\theta}_{2n}$ such that

$$n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) = \frac{1}{n^{\frac{1}{2}}}\mathscr{I}^{-1}\sum_{i=1}^{n}\frac{\partial \log f(x_i, \theta_n)}{\partial \theta_2} + \mathscr{I}^{-1}\mathscr{I}_{21}\gamma + \varepsilon_{1n} \tag{19}$$

where $\varepsilon_{1n} \rightarrow_P 0$. Usually, (19) possesses the alternative form

$$n^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta_{20}) = \frac{1}{n^{\frac{1}{2}}}\mathscr{I}^{-1}\sum_{i=1}^{n}\frac{\partial \log f(x_i, \theta_0)}{\partial \theta_2} + \varepsilon_{1n}^{*} \tag{20}$$

where $\partial \log f(x, \theta_0)/\partial\theta_2$ is the same as $\partial \log f(x, \theta_n)/\partial\theta_2$ except that the derivatives are evaluated at $\theta = \theta_0$ instead of $\theta = \theta_n$ and where $\varepsilon_{1n}^{*} \rightarrow_P 0$. However, we shall not try to formulate conditions under which the transition from (19) to (20) is valid since we shall take (19) as the basic form. Of course, (19) and (20) are identical when H_0 is true so only (20) need be considered if null-hypothesis distributions only are under consideration.

Maximum-likelihood estimators often satisfy (19); indeed, when the familiar conditions of Cramér (1946) page 500 are suitably modified to deal with the behaviour of the estimator of θ_{20} under the sequence $\{H_n\}$ it can be proved that (19) holds. However, the Cramér-type conditions seem unduly restrictive. Moreover, the Cramér theorem and its analogue for the present problem refer not to the estimator for which the likelihood L attains its maximum but to solutions of the likelihood equation $\partial \log L/\partial\theta_2 = 0$ and there are complications when the solution of this equation is not unique. Other technical problems arise when alternative approaches to the study of the efficiency of maximum-likelihood estimation are employed. For these reasons I shall not attempt to formulate sufficient conditions under which the maximum-likelihood estimator of θ_{20} satisfies (19) but suggest instead that for any particular problem a maximum-likelihood or other putative efficient estimator is first constructed and then the validity of (19) is checked directly.

In the notation of Assumptions (A1) and Theorem 1 we have $l(x, \theta_n) = \mathscr{I}^{-1}\,\partial \log f(x, \theta_n)/\partial\theta_2$ and $A = \mathscr{I}^{-1}\mathscr{I}_{21}$. Also, (A3) (ii) implies $L = \mathscr{I}$ and (A3) (iii) implies $h(t) = \mathscr{I}^{-1}g_2(t)$. Substituting in Theorem 1 we deduce immediately

THEOREM 2. *Suppose that $F(x, \theta)$ satisfies Assumptions* (A2) *and* (A3) *and that $\hat{\theta}_{2n}$ satisfies* (19). *Then under the sequence $\{H_n\}$ defined by* (3), *$\hat{y}_n$ defined by* (2) *converges weakly to the normal process z in D with mean and covariance functions*

$$E\{z(t)\} = \gamma'\{g_1(t) - \mathscr{I}_{21}'\mathscr{I}^{-1}g_2(t)\} \tag{21}$$

$$C\{z(t_1), z(t_2)\} = \min(t_1, t_2) - t_1 t_2 - g_2(t_1)'\mathscr{I}^{-1}g_2(t_2)\,. \tag{22}$$

Corollaries 1 and 2 of Theorem 1 may be applied to this case by making the appropriate substitutions.

5. Dependence of the distribution of $\hat{y}_n$ on F. It is obvious from the form of the covariance function (5) that, unlike y_n, $\hat{y}_n$ is not distribution-free on H_0 since its asymptotic distribution depends on F. Worse, it is not even asymptotically parameter-free since this distribution depends in general on the value of θ_{20}. Fortunately, in many important cases the distribution of $\hat{y}_n$ does not depend on θ_{20}. These include cases where $F(x, \theta)$ is transformable into a form in which θ_2 is a location or scale parameter or is a vector of both and where an appropriate estimator $\hat{\theta}_{2n}$ is used. In such cases one can in principle compute asymptotic significance points for statistics based on $\hat{y}_n$ which are valid for all θ_{20}.

6. Previous work. The asymptotic distributions of test statistics based on the whole function $\hat{y}_n$ when parameters are estimated have previously been investigated by Darling (1955) and Kac, Kiefer and Wolfowitz (1955). In neither of these papers, however, is the weak convergence of $\hat{y}_n$ itself considered. In contrast, the finite-dimensional distributions of $\hat{y}_n$ have been extensively studied, invariably in connection with the behaviour of the chi-squared and other tests of fit based on grouped data.

Apart from the early work of Fisher, discussed by Watson (1958), the first treatment of a problem of this type was by Chernoff and Lehmann (1954). These authors considered the case of maximum-likelihood estimation and preassigned group boundaries. Their use of $\hat{\theta}_n$ was to compute the "expected" numbers in the cells. Thus, putting $t = F(x, \theta_n)$ and $\tilde{t}_n(t) = F(x, \hat{\theta}_n)$ they studied essentially the asymptotic finite-dimensional distributions of the process $y_n^c(t) = n^{\frac{1}{2}}\{F_n(t) - \tilde{t}_n(t)\}$, $0 \leqq t \leqq 1$. Now $y_n^c(t) = y_n(t) + n^{\frac{1}{2}}(t - \tilde{t}_n(t)) = y_n(t) + n^{\frac{1}{2}}\{F_n(x, \theta_n) - F(x, \hat{\theta}_n)\} = y_n(t) + n^{\frac{1}{2}}(\theta_n - \hat{\theta}_n)'\ g(t) + \varepsilon_n{}^*(t)$ where $\varepsilon_n{}^* \rightarrow_P \mathbf{0}$. From (13) and Lemma 1 we have $\hat{y}_n(t) = y_n(t) + n^{\frac{1}{2}}(\hat{t}_n(t) - t) = y_n(t) + n^{\frac{1}{2}}\{F(x(t, \hat{\theta}_n), \theta_n) - F(x(t, \hat{\theta}_n), \hat{\theta}_n)\} + \varepsilon_{2n}(t) = y_n(t) + n^{\frac{1}{2}}(\theta_n - \hat{\theta}_n)'g(t) + \varepsilon_n{}^{**}(t)$ where $\varepsilon_n{}^{**} \rightarrow_P \mathbf{0}$. It follows that under the conditions of Theorem 1, $\hat{y}_n$ and $y_n{}^c$ have the same distribution. This represents an extension of the Chernoff–Lehmann result in three respects, (i) the whole process is considered and not just the finite-dimensional distributions, (ii) the result holds under the sequence H_n and not just on H_0, (iii) inefficient as well as efficient estimators are considered.

A detailed treatment of the chi-squared test when the determination of class boundaries depends on the data was given by Roy (1956). Roy considered the effects of parameter estimation for a general class of estimators which, on H_0, satisfy Assumptions (A1) above. His results give essentially the finite-dimensional distributions of $\hat{y}_n$ on H_0. Regrettably, this work was never published, although Watson ((1958), reply to the discussion) and (1959) has provided a useful summary of it.

Independently of Roy and Watson (1957) obtained the asymptotic distribution on H_0 of the cell frequencies for the special case of tests of normality when the cell boundaries are based on maximum-likelihood estimates and extended this treatment subsequently (1958) to other distributions. Watson's results give the finite-dimensional distributions corresponding to Theorem 2 when $\gamma = 0$. In this and a third paper (1959) Watson discussed the implication of these results for the practical use of the chi-squared test.

Some related results were obtained, again independently, by Barton (1957) in a study of the effect of parameter estimation on a grouped form of Neyman's smooth test. He obtained a result (his Theorem II) analogous to that of Chernoff and Lehmann and a further result for the special case of estimating location and scale parameters (his Theorem III) analogous to that of Roy and Watson. Barton's work goes further to the extent that in his Theorem II he obtains the distribution of his statistics under a sequence of alternatives. The relation of Barton's results to the other work mentioned is discussed by Watson (1957, 1958).

For the case where no parameter estimation is involved Chibisov (1965) has proved that when the null hypothesis is Pr $(x \leqq t) = t$ and the sequence of alternatives is $\Pr(x \leqq t) = H_n^*(t)$, where $\lim_{n\to\infty} n^{\frac{1}{2}}\{H_n^*(t) - t\} = \delta(t)$, then under mild conditions on $\delta(t)$ and under the sequence of alternatives H_n^*, $\hat{y}_n \to_{\mathscr{D}} y + \delta$ where y is tied-down Brownian motion.

The results of Chernoff and Lehman, Roy and Watson have recently been extended to the multivariate case by Moore (1971).

Acknowledgment. I am grateful to Ronald Pyke and Patrick Billingsley for comments on earlier drafts leading to improvements in presentation.

REFERENCES

[1] APOSTOL, T. M. (1957). *Mathematical Analysis.* Addison-Wesley, Reading, Mass.

[2] BARTON, D. E. (1956). Neyman's ψ_k^2 test of goodness of fit when the null hypothesis is composite. *Skand. Aktuarietidskr.* **39** 216–245.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

[5] CHERNOFF, H. and LEHMANN, E. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25** 579–586.

[6] CHIBISOV, D. M. (1965). An investigation of the asymptotic power of the tests of fit. *Theor. Probability Appl.* **10** 421–437.

[7] CRAMÉR, H. (1964). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[8] DARLING, D. A. (1955). The Cramér–Smirnov test in the parametric case. *Ann. Math. Statist.* **26** 1–20.

[9] KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26** 189–211.

[10] MOORE, D. S. (1971a). A chi-square statistic with random cell boundaries. *Ann. Math. Statist* **42** 147–156.

[11] ROY, A. R. (1956). On chi-squared statistics with variable intervals. Unpublished technical report, Department of Statistics, Stanford Univ.

[12] WATSON, G. S. (1957). The chi-squared goodness-of-fit tests for normal distributions. *Biometrika* **44** 336–348.

[13] WATSON, G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc. Ser. B* **20** 44–61.

[14] WATSON, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics* **15** 440–468.

LONDON SCHOOL OF ECONOMICS
AND POLITICAL SCIENCE
HOUGHTON STREET
LONDON WC 2, ENGLAND