The Annals of Probability
1978, Vol. 7, No. 5, 790–810

APPROXIMATIONS OF THE EMPIRICAL PROCESS WHEN PARAMETERS ARE ESTIMATED

BY M. D. BURKE[1], M. CSÖRGŐ[2], S. CSÖRGŐ[3] AND P. RÉVÉSZ

University of Calgary, Carleton University, Szeged University and the Mathematical Institute, Budapest

Almost sure and in-probability representations of the empirical process by appropriate Gaussian processes are obtained when unknown parameters of the underlying distribution function are estimated. As to the method of estimation, we consider maximum likelihood and maximum likelihood-like estimators and construct the above-mentioned representations under a null hypothesis. Similar results are obtained also when using Durbin's more general class of estimators under a sequence of alternatives which converge to the null hypothesis. The resulting Gaussian processes depend, in general, on the true value of the unknown parameters.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed (i.i.d.) random variables with distribution function F. Let $F_n(x)$ denote the proportion of the X_i $(1 \leqslant i \leqslant n)$ which are less than or equal to x, $x \in R$. F_n is the empirical distribution function based on $X_1, X_2, \cdots, X_n$. The empirical process based on $X_1, X_2, \cdots, X_n$ is defined by

$$\alpha_n(x) = n^{\frac{1}{2}}[F_n(x) - F(x)], \qquad x \in R. \tag{1.1}$$

Beginning with Breiman (1968) and Brillinger (1969), there has been much work done on approximating α_n almost surely by sequences of Brownian bridges. Kiefer (1972) was the first to obtain a strong (almost sure) approximation of α_n in terms of a Gaussian process in both x and n. Csörgő and Révész (1975) obtained similar results with a new method for the multivariate empirical process when F is the uniform distribution function on the d-dimensional unit cube. We will quote the result with the best rates, obtained by Komlós, Major and Tusnády (1975). They formulated this result in the case when F is the uniform-(0, 1) distribution function, but their theorem can be extended easily to arbitrary F (cf. Remark 1 of S. Csörgő (1978)).

Received July 29, 1976.

[1]Research supported by a Canadian N. R. C. Scholarship at Carleton University and a University of Calgary research grant.

[2]Research partially supported by a Canadian N.R.C. Grant and a Killam Senior Research Scholarship at Carleton University.

[3]Work done while this author was on leave from Szeged University and a Visiting Scientist at Carleton University, supported by Canadian N. R. C. operating grants of D. A. Dawson and J. N. K. Rao.

AMS 1970 *subject classifications.* Primary 62E20, 62H15, 60F15, 60F05; secondary 62G99, 60G15.

Key words and phrases. Empirical processes, parametric estimation, strong approximations, Kiefer process, composite goodness-of-fit hypotheses.

A Kiefer process K, defined on $[0, 1] \times (0, \infty)$, is a separable Gaussian process with mean $EK(s, y) = 0$ and covariance function

$$EK(s, y_1)K(t, y_2) = \min(y_1, y_2)\{\min(s, t) - st\}.$$

For fixed $y > 0$, $y^{-\frac{1}{2}}K(s, y) =_{\mathfrak{D}} B(s)$, a Brownian Bridge defined on [0, 1]. We have

THEOREM A (Komlós, Major and Tusnády (1975)). *If the underlying probability space is rich enough, one can define a Brownian bridge* $\{B_n(s);\ 0 \leqslant s \leqslant 1\}$ *for each n and a Kiefer process* $\{K(s, y);\ 0 \leqslant s \leqslant 1,\ 0 < y < \infty\}$ *such that*

$$\sup_{-\infty<x<\infty}|\alpha_n(x) - B_n(F(x))| =_{\text{a.s.}} O\{n^{-\frac{1}{2}}\log n\}$$

and

$$\sup_{-\infty<x<\infty}|n^{\frac{1}{2}}\alpha_n(x) - K(F(x), n)| =_{\text{a.s.}} O\{\log^2 n\},$$

where α_n *is defined by* (1.1).

REMARK 1. By the phrase "if the underlying probability space is rich enough," we mean that an independent sequence of Wiener processes, which is independent of the originally given i.i.d. sequence $\{X_n\}$, can be constructed on the assumed probability space. Throughout this paper, it will be assumed that the underlying probability spaces are rich enough in this sense.

From a statistical point of view, Theorem A is useful to construct confidence intervals for an unknown distribution function F and also to construct goodness-of-fit tests for a completely specified F. Most goodness-of-fit problems arising in practice, however, do not usually specify F completely, and, instead of one specific F, we are frequently given a whole parametric family of distribution functions $\{F(x; \theta);\ \theta \in \Xi \subseteq R^p\}$. From a goodness-of-fit point of view, the unknown parameters θ are a nuisance (nuisance parameters), which render most goodness-of-fit null hypotheses composite ones. There are many ways of "getting rid of θ" so as to reduce composite goodness-of-fit hypotheses to simple ones. As far as the empirical process is concerned, one natural way of doing this is to "estimate out θ" by using some kind of a "good estimator" sequence $\{\hat{\theta}_n\}$, based on random samples $X_1, X_2, \cdots, X_n$ $(n = 1, 2, \cdots)$ on $F(x; \theta)$.

Concerning the classical Cramér-von Mises and Komogorov-Smirnov statistics, Darling (1955) and Kac, Kiefer and Wolfowitz (1955) investigated their asymptotic distributions when the unknown parameters of a *specified* distribution function were to be estimated first. Durbin (1973a) considered the more global question of weak convergence of the empirical process under a given sequence of alternative hypotheses when parameters of a continuous *unspecified* distribution function $F(x; \theta)$ are estimated from the data. The estimators themselves were to satisfy certain maximum likelihood-like conditions. Durbin (1973a) showed that, for such a general class of estimators, the estimated empirical process converges weakly to a Gaussian process, whose mean and covariance functions he also gave.

In this article, we are going to use the recently developed strong approximation methodology of Kiefer (1972), Csörgő-Révész (1975) and Komlós-Major-Tusnády (1975) to study the problem of obtaining asymptotic in-probability and almost sure representations, in terms of Gaussian processes in both x and n, of the empirical process when parameters are estimated. This approach has already been demonstrated in the preliminary drafts of Csörgő-Komlós-Major-Révész-Tusnády (1974) and Burke-Csörgő (1976). In this exposition we follow the same road (correcting also previous oversights while going along), but we also weaken substantially the regularity conditions under which these representations will hold. As to the type of estimation of the parameters $\theta \in R^p$ of $F(x; \theta)$, we follow Durbin (1973a) and, as mentioned above, in addition to his weak convergence, we obtain explicit representations of the limiting Gaussian process in a straightforward way.

In Section 3 we formulate and prove the above mentioned two-parameter representation theorems under the null hypotheses. Section 4 illustrates how a maximum likelihood estimation situation can fit into our methodology. In Section 5 the results of Section 3 are extended to also cover a sequence of alternatives. While going along we point out how the results of Durbin follow from ours. In Section 6 the in-probability representation result of Section 3 is extended to the estimated multivariate empirical process.

In connection with our work in Sections 5 and 6, we should also mention that convergence in distribution of the Cramér-von Mises functional of the multivariate empirical process under contiguous alternatives when parameters are estimated was also studied by Neuhaus (1974, 1976).

The theorems in this paper can also be stated using a sequence of Brownian bridges instead of the Kiefer process. However, the approach of working with Gaussian processes defined in terms of Kiefer processes gives one a process in both x and n, while a corresponding construction in terms of Brownian bridges B_n would only give a Gaussian process in x, for each n.

Using the strong approximation results of Csörgő and Révész (1978), the sample quantile process when underlying parameters are estimated can be handled similarly to the empirical process in this paper.

Using the strong approximation results of S. Csörgő (1978) for the empirical characteristic process, the parameter estimated empirical characteristic process is investigated by S. Csörgő (1979).

2. Notation.

(2.1) The transpose of a vector V will be denoted by V^t.

(2.2) The norm $\|\ \|$ on R^p is defined by $\|(y_1, \cdots, y_p)\| = \max_{1 \leq i \leq p}|y_i|$.

(2.3) For a function $g(x; \theta)$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$, $\nabla_\theta g(x; \theta_0)$ will denote the vector of partial derivatives $((\partial/\partial\theta_1)g(x; \theta), (\partial/\partial\theta_2)g(x; \theta), \cdots, (\partial/\partial\theta_p)g(x; \theta))$ evaluated at $\theta = \theta_0$. Also $\nabla_\theta^2 g(x; \theta_0)$ will denote the vector $((\partial^2/\partial\theta_1^2)g(x; \theta), \cdots, (\partial^2/\partial\theta_p^2)g(x; \theta))$ evaluated at $\theta = \theta_0$.

(2.4) The matrix $[(\partial^2/\partial\theta_i\,\partial\theta_j)g(x;\theta)]_{i,j}$ will be denoted by $g''_{\theta\theta}(x,\theta)$.

(2.5) For a matrix or vector $V=(v_{ij})$, let $|V|$ denote the matrix $(|v_{ij}|)$, let $\int V$ denote $(\int v_{ij})$, and let V^δ denote (v_{ij}^δ).

3. Approximations of the estimated empirical process. For an i.i.d. sequence $X_1, X_2, \cdots$ from a family of distribution functions $\{F(x;\theta);\ x\in R,\ \theta\in\Xi\subseteq R^p\}$, let $\{\hat\theta_n\}$ be a sequence of estimators of θ based on $X_1, X_2, \cdots, X_n$. Consider the estimated empirical process defined by

$$\hat\alpha_n(x)=n^{\frac{1}{2}}\left[F_n(x)-F(x;\hat\theta_n)\right], \tag{3.1}$$

where $x\in R$ and F_n is the empirical distribution function.

First we list the set of all conditions which will be used in this section. We emphasize that only subsets of it will be used at different stages in the sequel.

(i) $n^{\frac{1}{2}}(\hat\theta_n-\theta_0)=n^{-\frac{1}{2}}\Sigma_{j=1}^n l(X_j,\theta_0)+\varepsilon_{1n}$, where θ_0 is the theoretical value of θ, $l(\cdot,\theta_0)$ is a measurable p-dimensional vector valued function, and ε_{1n} converges to zero in a manner to be specified later on.

(ii) $El(X_j,\theta_0)=0$.

(iii) $M(\theta_0)=E\{l(X_j,\theta_0)^t l(X_j,\theta_0)\}$ is a finite nonnegative definite matrix.

(iv) The vector $\nabla_\theta F(x;\theta)$ is uniformly continuous in x and $\theta\in\Lambda$, where Λ is the closure of a given neighbourhood of θ_0.

(v) Each component of the vector function $l(x,\theta_0)$ is of bounded variation on each finite interval.

(vi) The vector $\nabla_\theta F(x,\theta_0)$ is uniformly bounded in x, and the vector $\nabla_\theta^2 F(x;\theta)$ is uniformly bounded in x and $\theta\in\Lambda$, where Λ is as in (iv).

(3.2)

(vii)
$$\lim_{s\searrow 0}(s\log\log 1/s)^{\frac{1}{2}}\|l(F^{-1}(s;\theta_0),\theta_0\|=0$$
and
$$\lim_{s\nearrow 1}((1-s)\log\log 1/(1-s))^{\frac{1}{2}}\|l(F^{-1}(s;\theta_0),\theta_0)\|=0,$$
where $F^{-1}(s;\theta_0)=\inf\{x: F(x;\theta_0)\geqslant s\}$.

(viii)
$$s\|(\partial/\partial s)l(F^{-1}(s;\theta_0),\theta_0)\|\leqslant C,\ 0<s<\tfrac{1}{2}$$
and
$$(1-s)\|(\partial/\partial s)l(F^{-1}(s;\theta_0),\theta_0)\|\leqslant C,\ \tfrac{1}{2}<s<1$$
for some positive constant C, where the vector of partial derivatives of the components of $l(F^{-1}(s;\theta_0),\theta_0)$ with respect to s, $(\partial/\partial s)l(F^{-1}(s;\theta_0),\theta_0)$, exists for all $s\in(0,1)$.

The estimated empirical process $\hat\alpha_n(x)$ of (3.1) will be approximated by the two-parameter Gaussian process

(3.3)
$$G(x,n)=n^{-\frac{1}{2}}K(F(x;\theta_0),n)-\left\{\int l(x,\theta_0)d_x n^{-\frac{1}{2}}K(F(x;\theta_0),n)\right\}\nabla_\theta F(x;\theta_0)^t,$$

where K is the Kiefer process of Theorem A. G has mean function $EG(x, n) = 0$ and covariance function

$$\begin{aligned} EG(x, n)G(y, m) = \min(n, m) \cdot (nm)^{-\frac{1}{2}}\{&F(\min(x, y); \theta_0) \\ &- F(x; \theta_0)F(y; \theta_0) - J(x) \\ &\cdot \nabla_\theta F(y; \theta_0)' - J(y) \cdot \nabla_\theta F(x; \theta_0)' \\ &+ \nabla_\theta F(x; \theta_0) \cdot M(\theta_0) \cdot \nabla_\theta F(y; \theta_0)'\}, \end{aligned} \tag{3.4}$$

where $M(\theta_0)$ is defined by (3.2) (iii) and

$$J(x) = \int_{-\infty}^{x} l(z, \theta_0) d_z F(z; \theta_0).$$

(Here, of course, $F(\min(x, y); \theta_0) = \min(F(x; \theta_0), F(y; \theta_0))$, but for the sake of later reference (the multivariate case, cf. Section 6) we use the former form.) Since $M(\theta_0)$ is nonnegative definite, there is a nonsingular matrix $D(\theta_0)$ such that

$$D(\theta_0)' M(\theta_0) D(\theta_0) = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.5}$$

where I is the identity matrix and rank I = rank $M(\theta_0)$. Hence $G(x, n)$ of (3.3) can be written as

$$G(x, n) = n^{-\frac{1}{2}} K(F(x; \theta_0), n) - n^{-\frac{1}{2}} W(n) \cdot D^{-1}(\theta_0) \cdot \nabla_\theta F(x; \theta_0)', \tag{3.6}$$

where $W(n) = \int l(x, \theta_0) d_x K(F(x; \theta_0), n) \cdot D(\theta_0)$ is a vector-valued Wiener process with covariance structure: $\min(n, m)$ multiplied by (3.5).

Clearly we have for each n that

$$G(x, n) =_{\mathfrak{D}} D(x) = B(F(x; \theta_0)) - \{\int l(x, \theta_0) d_x B(F(x; \theta_0))\} \nabla_\theta F(x; \theta_0)', \tag{3.7}$$

where $=_{\mathfrak{D}}$ stands for the equality of all finite-dimensional distributions and $B(x)$ is a Brownian bridge. Thus $ED(x)D(y) = \{\ \}$, where $\{\ \}$ is the right-hand side factor in (3.4).

THEOREM 3.1. *Suppose that the sequence* $\{\hat{\theta}_n\}$ *satisfies* (3.2) (i), (ii), (iii), *and let*

$$\varepsilon_{2n} = \sup_{-\infty < x < \infty} |\hat{\alpha}_n(x) - G(x, n)|.$$

Then

(a) $\varepsilon_{2n} \to_p 0$, *if conditions* (3.2) (iv), (v) *hold and* $\varepsilon_{1n} \to_p 0$;

(b) $\varepsilon_{2n} \to_{\text{a.s.}} 0$, *if conditions* (3.2) (vi)–(viii) *hold and* $\varepsilon_{1n} \to_{\text{a.s.}} 0$;

(c) $\varepsilon_{2n} =_{\text{a.s.}} O\{\max(h(n), n^{-\varepsilon})\}$ *for some* $\varepsilon > 0$, *if conditions* (3.2) (vi)–(viii) *hold and* $\varepsilon_{1n} =_{\text{a.s.}} O\{h(n)\}$, $h(n) > 0$, $h(n) \to 0$.

REMARK 2. Durbin's (1973a) result (under his null hypotheses—Corollary 1 in Durbin (1973a)), i.e., $\hat{\alpha}_n(F^{-1}(\cdot, \hat{\theta}_n)) \to_{\mathfrak{D}} D(F^{-1}(\cdot; \theta_0))$, follows from part (a) of Theorem 3.1, because of (3.7). Here $\to_{\mathfrak{D}}$ denotes weak convergence in the function space $D[0, 1]$. (This will also be the case in Section 5 under his sequences of alternatives. See also Remark 7 in Section 5 concerning Durbin's original setup.) We should point out that Durbin used conditions (3.2) (i)–(iv), with $\varepsilon_{1n} \to_p 0$, to prove this weak convergence, but not (v). This slight regularity condition (3.2) (v) (satisfied, sure enough, in each practical situation) is the only price we pay for obtaining our in-probability representation of the limiting Gaussian process in both x and n. Nevertheless, if one still would like to get rid of this condition, then the use of Theorem A is still advantageous. As the proof of part (a) will show, we have (without (v))

$$\sup_x|\hat{\alpha}_n(x) - Y_n(F(x;\theta_0))| \to_p 0, \tag{3.8}$$

where

$$Y_n(s) = n^{-\frac{1}{2}}K(s,n) - \left\{n^{-\frac{1}{2}}\Sigma_{j=1}^n l(X_j,\theta_0)\right\} \nabla_\theta F\left(F^{-1}(s;\theta_0);\theta_0\right)^t.$$

In this way we could save a tightness-proof, since the tightness of $\{Y_n\}$ reduces to the a.s. continuity of the Kiefer process. But one still has to prove the convergence of the finite-dimensional distributions of Y_n to those of $D(\cdot)$ in (3.7), which is, on one hand, again easier than for $\hat{\alpha}_n$, but, on the other hand, is essentially a repetition of the proof of Lemma 3 in Durbin (1973a). We should also note, however, that, unlike in Durbin (1973a), the continuity of $F(x,\theta)$ in x is not used in Theorem 3.1. Conditions (3.2) (iv) and (vi) can be satisfied without the continuity of F (example: the binomial distribution).

REMARK 3. Conditions (3.2) (vi)–(viii) are the extra ones used to obtain our a.s. representation (in case of part (c) with a rate sequence) of the limiting Gaussian process in both x and n. These latter conditions are weaker than those in the preliminary drafts of Csörgő et al. (1974) and Burke and Csörgő (1976). Commonly used distributions such as the normal and exponential would satisfy these conditions when maximum likelihood estimators are employed. Indeed, in the latter case $l(x;\theta_0)$ of (3.2) (i) turns out to be proportional to $\nabla_\theta \log f(x;\theta_0)$ (cf. Theorem 4.1 and Remark 5), with $f(x;\theta) = (d/dx)F(x;\theta)$, and our extra requirements (3.2) (vi)–(viii) hold true for all those density functions whose tail behaviour in the sense of these requirements is similar to that of the exponential density.

Introduce the following

$$\varepsilon_{3n}(s) = n^{\frac{1}{2}}\left[F_n\left(F^{-1}(s;\theta_0)\right) - s\right] - n^{-\frac{1}{2}}K(s,n), \tag{3.9}$$

where K is the Kiefer process of Theorem A. We have

$$\varepsilon_{3n}(F(x;\theta_0)) = n^{\frac{1}{2}}\left[F_n(x) - F(x;\theta_0)\right] - n^{-\frac{1}{2}}K(F(x;\theta_0),n).$$

Our proof of Theorem 3.1 hinges on the following two lemmas.

LEMMA 3.1. *Suppose that the vector function $l(x,\theta_0)$ satisfies conditions* (3.2) (iii) *and* (v). *Then, as $n \to \infty$,*

$$L_n = \int l(x,\theta_0)d_x\varepsilon_{3n}(F(x;\theta_0)) \to_p 0.$$

PROOF OF LEMMA 3.1. Let $T_j(x)$ denote the total variation of the jth component $l_j(\cdot,\theta_0)$ of $l(\cdot,\theta_0)$ on the interval $[-x,x]$, $j = 1,\cdots,p$, and let $T(x) = (T_1(x),\cdots,T_p(x))$. Clearly we can choose a sequence of positive numbers u_n tending so slowly to infinity that $\|T(u_n)\|n^{-\frac{1}{2}}\log^2 n \to 0$. (If $\|T(n)\|$ is bounded, then any $u_n \to \infty$ sequence will suffice, while if $\|T(n)\| \nearrow \infty$, then we take $u_n = T^{-1}(v_n)$, where $v_n \nearrow \infty$ so that $v_n = o\{n^{\frac{1}{2}}/\log^2 n\}$, and $T^{-1}(y) = \inf\{x : \|T(x)\| \geqslant y\}$). With this u_n then, consider

$$L_n = \int_{|x|>u_n} l(x, \theta_0) d_x n^{\frac{1}{2}}[F_n(x) - F(x; \theta_0)] - \int_{|x|>u_n} l(x, \theta_0) d_x n^{-\frac{1}{2}} K(F(x; \theta_0), n)$$
$$+ \int_{|x|\leqslant u_n} l(x, \theta_0) d_x \varepsilon_{3n}(F(x; \theta_0)) = L_{1n} - L_{2n} + L_{3n}.$$

Integrating by parts and using Theorem A one obtains

$$\|L_{3n}\| \leqslant \|\int_{-u_n}^{u_n} \varepsilon_{3n}(F(x; \theta_0))\, dl(x, \theta_0)\| + \|[\varepsilon_{3n}(F(x; \theta_0)) l(x, \theta_0)]_{x=-u_n}^{u_n}\|$$
$$=_{\text{a.s.}} O\{n^{-\frac{1}{2}}\log^2 n\} \|T(u_n)\| \to 0.$$

If the components in L_{1n} and L_{2n} are denoted respectively by $L_{1n}^{(j)}$ and $L_{2n}^{(j)}$, $j = 1, \cdots, p$, then we have $EL_{1n}^{(j)} = EL_{2n}^{(j)} = 0$ and

$$(3.10) \quad E(L_{1n}^{(j)})^2 = E(L_{2n}^{(j)})^2 = \int_{|x|>u_n} l_j^2(x, \theta_0)\, dF(x; \theta_0)$$
$$- \left(\int_{x\leqslant -u_n} l_j(x, \theta_0)\, dF(x; \theta_0)\right)^2 - \left(\int_{x\geqslant u_n} l_j(x, \theta_0)\, dF(x; \theta_0)\right)^2.$$

Whence, by the Chebishev inequality, with $\varepsilon > 0$,

$$P\{\|L_{1n}\| + \|L_{2n}\| > 2\varepsilon\} \leqslant \frac{2}{\varepsilon^2} \Sigma_{j=1}^{p} \int_{|x|>u_n} l_j^2(x, \theta_0)\, dF(x; \theta_0),$$

and this latter bound tends to zero by condition (3.2) (iii), since $u_n \to \infty$. □

LEMMA 3.2. *Suppose that the vector function* $l(F^{-1}(s; \theta_0), \theta_0)$ *satisfies conditions* (3.2) (vii) *and* (viii). *Then, as* $n \to \infty$,

$$L_n = \int_0^1 l(F^{-1}(s; \theta_0), \theta_0)\, d\varepsilon_{3n}(s) =_{\text{a.s.}} O\{n^{-\varepsilon}\},$$

for some $\varepsilon > 0$, *where* $\varepsilon_{3n}(s)$ *is again that of* (3.9).

PROOF OF LEMMA 3.2. We have

$$L_n = \int_0^1 \varepsilon_{3n}(s)(\partial/\partial s) l(F^{-1}(s; \theta_0), \theta_0)\, ds.$$

This latter equality is correct provided the function $\varepsilon_{3n}(s) l(F^{-1}(s; \theta_0), \theta_0)$ is almost surely the zero vector at $s = 0$ and $s = 1$. This, in turn, is true by condition (3.2) (vii) and by the fact that the Kiefer process $K(s, n)$ and the empirical process $n^{\frac{1}{2}}[F_n(F^{-1}(s; \theta_0)) - s]$ behave like $(s \log\log 1/s)^{\frac{1}{2}}$ and $((1-s)\log\log 1/(1-s))^{\frac{1}{2}}$ as $s \searrow 0$ and $s \nearrow 1$, respectively.

Consider now

$$L_n = \int_0^{n^{-\frac{1}{3}}} + \int_{n^{-\frac{1}{3}}}^{\frac{1}{2}} + \int_{\frac{1}{2}}^{1-n^{-\frac{1}{3}}} + \int_{1-n^{-\frac{1}{3}}}^{1} = L_{1n}^* + L_{2n}^* + L_{3n}^* + L_{4n}^*.$$

By Theorem A and the first part of (3.2) (viii) we have almost surely

$$\|L_{2n}^*\| \leqslant O\{n^{-\frac{1}{2}}\log^2 n\} \int_{n^{-\frac{1}{3}}}^{\frac{1}{2}} \|(\partial/\partial s) l(F^{-1}(s; \theta_0), \theta_0)\|\, ds$$
$$\leqslant O\{n^{-\frac{1}{2}}\log^2 n\} \int_{n^{-\frac{1}{3}}}^{\frac{1}{2}} s^{-1}\, ds$$
$$= O\{n^{-\frac{1}{2}}\log^3 n\}.$$

Also,

$$\|L_{1n}^*\| \leqslant \int_0^{n^{-\frac{1}{3}}} |\alpha_n(F^{-1}(s;\theta_0))| \cdot \|(\partial/\partial s) l(F^{-1}(s;\theta_0),\theta_0)\|\, ds$$
$$+ \int_0^{n^{-\frac{1}{3}}} |n^{-\frac{1}{2}} K(s,n)| \cdot \|(\partial/\partial s) l(F^{-1}(s;\theta_0),\theta_0)\|\, ds.$$

Since by (ii) of Theorem 3.1 in Csáki (1975) (cf. also (1.9) in Csáki (1977))

$$\sup_{0<s<n^{-\frac{1}{3}}} |\alpha_n(F^{-1}(s;\theta_0))(s(1-s))^{-\frac{1}{2}}| =_{\text{a.s.}} O\{\log^2 n\}, \tag{3.11}$$

and by (1.15.1) in Csörgő and Révész (1979),

$$\limsup_{n\to\infty} \sup_{0<s<1} |K(s,n)[4ns(1-s)\log\log(n/(s(1-s)))]^{-\frac{1}{2}}| =_{\text{a.s.}} 1, \tag{3.12}$$

we have by the first part of (3.2) (viii)

$$\|L_{1n}^*\| \leqslant O\{\log^2 n\} \int_0^{n^{-\frac{1}{3}}} (\log\log(n/s))^{\frac{1}{2}} s^{-\frac{1}{2}}\, ds, \quad \text{a.s.},$$
$$=_{\text{a.s.}} O\{n^{-\frac{1}{7}}\}.$$

The terms L_{3n}^* and L_{4n}^* are estimated similarly and hence the lemma. □

REMARK 4. The fact that the proof of Lemma 3.2 gives a rate of convergence $O\{n^{-\frac{1}{7}}\}$ is only incidental. This rate is dependent on the way the stochastic integral was broken into parts. For any specific $O\{h(n)\}$ rate of ε_{1n} in (c) of Theorem 3.1, one should try to make that of Lemma 3.2 better for the sake of ε_{2n} in (c).

PROOF OF THEOREM 3.1. Using the one-term Taylor expansion of F with respect to θ_0 we obtain

$$\hat{\alpha}_n(x) = n^{\frac{1}{2}}[F_n(x) - F(x;\theta_0)] - n^{\frac{1}{2}}[F(x;\hat{\theta}_n) - F(x;\theta_0)]$$
$$= n^{-\frac{1}{2}} K(F(x;\theta_0), n) - n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)\nabla_\theta F(x;\theta_n^*)^t + \varepsilon_{3n}(F(x;\theta_0)) \tag{3.13}$$
$$= n^{-\frac{1}{2}} K(F(x;\theta_0)) - n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)\nabla_\theta F(x;\theta_0)^t + \varepsilon_{3n}(F(x;\theta_0) + \varepsilon_{4n}(x)$$

where ε_{3n} is defined by (3.9), and by Theorem A

$$\sup_x |\varepsilon_{3n}(F(x;\theta_0))| =_{\text{a.s.}} O\{n^{-\frac{1}{2}}\log^2 n\}, \tag{3.14}$$

while $\|\theta_n^* - \theta_0\| \leqslant \|\hat{\theta}_n - \theta_0\|$ and

$$\varepsilon_{4n}(x) = n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)(\nabla_\theta F(x;\theta_0) - \nabla_\theta F(x;\theta_n^*))^t.$$

It follows from (3.2) (i), (ii) and (iii) that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is asymptotically a normal vector, and thus $\|\hat{\theta}_n - \theta_0\| \to_p 0$. Hence, using also (3.2) (iv), we have

$$\sup_x \|\varepsilon_{4n}(x)\| \to_p 0. \tag{3.15}$$

Also, by conditions (3.2) (i) and (ii)

$$
\begin{aligned}
n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) &= n^{-\frac{1}{2}}\Sigma_{j=1}^n l(X_j, \theta_0) + \varepsilon_{1n} \\
&= \int l(x, \theta_0) d_x n^{\frac{1}{2}} F_n(x) + \varepsilon_{1n} \\
&= \int l(x, \theta_0) d_x n^{\frac{1}{2}}[F_n(x) - F(x; \theta_0)] + \varepsilon_{1n} \\
&= \int l(x, \theta_0) d_x n^{-\frac{1}{2}} K(F(x; \theta_0), n) + L_n + \varepsilon_{1n},
\end{aligned} \tag{3.16}
$$

where L_n is of Lemma 3.1. Since the vector $\nabla_\theta F(x; \theta_0)$ is uniformly bounded in x by (3.2) (iv), part (a) of the theorem follows from (3.14), (3.15) and (3.16).

To prove parts (b) and (c) we use the two-term Taylor expansion of F with respect to θ_0 in the second term of the first row in (3.13). Remembering Notation (2.5) and applying also (3.16), we obtain

$$
\begin{aligned}
\hat{\alpha}_n(x) &= n^{-\frac{1}{2}} K(F(x; \theta_0), n) - n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)\nabla_\theta F(x; \theta_0)^t \\
&\quad - \tfrac{1}{2} n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)^2 \nabla_\theta^2 F(x, \theta_n^*)^t + \varepsilon_{3n}(F(x; \theta_0)) \\
&= G(x, n) + (L_n + \varepsilon_{1n})\nabla_\theta F(x; \theta_0)^t \\
&\quad - \tfrac{1}{2} n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)^2 \nabla_\theta^2 F(x, \theta_n^*)^t + \varepsilon_{3n}(F(x; \theta_0)),
\end{aligned}
$$

where $\|\theta_n^* - \theta_0\| \leqslant \|\hat{\theta}_n - \theta_0\|$. If $\varepsilon_{1n} \to_{\text{a.s.}} 0$, then it follows from (3.2) (i) that $\theta_n^* \to_{\text{a.s.}} \theta_0$. Hence the vector $\nabla_\theta^2 F(x; \theta_n^*)$ is almost surely uniformly bounded in x and n, by (3.2) (vi). Because of (3.2) (iii) the law of iterated logarithm can be applied componentwise for the partial sum sequence in (3.2) (i). Whence we get $\|n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)^2\| =_{\text{a.s.}} O\{n^{-\frac{1}{2}}\log\log n\}$, that is

$$\sup_x |n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)^2 \nabla_\theta^2 F(x, \theta_n^*)^t| =_{\text{a.s.}} O\{n^{-\frac{1}{2}}\log\log n\}.$$

Thus, if (3.2) (i), (ii), (iii) and (vi) hold and $\varepsilon_{1n} \to_{\text{a.s.}} 0$, then

$$\hat{\alpha}_n(x) - G(x, n) =_{\text{a.s.}} \varepsilon_{5n}(x) + O\{n^{-\frac{1}{2}}\log^2 n\},$$

where $\varepsilon_{5n}(x) = (L_n + \varepsilon_{1n})\nabla_\theta F(x; \theta_0)^t$. If (3.2) (vii), (viii) hold, then by Lemma 3.2 $L_n =_{\text{a.s.}} O\{n^{-\varepsilon}\}$. Whence by (3.2) (vi)

$$\sup_x |\varepsilon_{5n}(x)| \to_{\text{a.s.}} 0.$$

If, in addition, $\varepsilon_{1n} =_{\text{a.s.}} O\{h(n)\}$, $h(n) > 0$, $h(n) \to 0$, then

$$\sup_x |\varepsilon_{5n}(x)| =_{\text{a.s.}} O\{\max(h(n), n^{-\varepsilon})\}. \qquad \square$$

The limiting Gaussian process G of Theorem 3.1 depends, in general, not only on F but also on θ_0, the true theoretical value of θ. Thus, in general, Theorem 3.1 cannot be used to test the composite hypothesis

$$H_0 : F \in \{F(x; \theta) : \theta \in \Xi \subseteq R^p\}.$$

In order to give an approximate solution to the latter problem, we define the process $\hat{G}(x, n)$ by

(3.17) $$\hat{G}(x, n) = n^{-\frac{1}{2}}K(F(x; \hat{\theta}_n), n) - n^{-\frac{1}{2}}W(n) \cdot D^{-1}(\hat{\theta}_n) \cdot \nabla_\theta F(x; \hat{\theta}_n)^t,$$

where $M(\theta)$ (cf. (3.2) (iii)) is assumed to exist and is nonnegative definite for $\theta \in \Lambda$, and $D(\theta)^t M(\theta) D(\theta)$ is assumed to satisfy (3.5). For $W(n)$ see (3.6). We have

THEOREM 3.2. *Suppose that the partial derivatives* $(\partial/\partial\theta_j)D^{-1}(\theta)\nabla_\theta F(x; \theta)^t$, $1 \leq j \leq p$, *exist and are uniformly bounded on* $R \times \Lambda$. *Then, under the conditions* (3.2) (i), (ii), (iii) *and* (vi)

(3.18) $$\sup_{-\infty<x<\infty}|\hat{G}(x, n) - G(x, n)| = \varepsilon_{6n},$$

where $\varepsilon_{6n} \to_p 0$ *if* $\varepsilon_{1n} \to_p 0$, *and* $\varepsilon_{6n} =_{\text{a.s.}} O(n^{-\delta})$ *for some* $\delta > 0$, *if* $\varepsilon_{1n} \to_{\text{a.s.}} 0$ *as* $n \to \infty$. *Consequently,*

$$\sup_{-\infty<x<\infty}|\hat{\alpha}(x) - \hat{G}(x, n)| = \varepsilon_{2n}^*$$

where ε_{2n}^* *converges to zero like* ε_{2n} *of* (a) *or* (b) *or* (c) *in Theorem* 3.1, $\hat{G}$ *is defined by* (3.17) *and* ε_{1n} *by* (3.2) (i).

PROOF. Assume $\varepsilon_{1n} \to_{\text{a.s.}} 0$, for it will be clear from the proof where to make the obvious changes to arrive at the conclusion of Theorem 3.1 in the case when $\varepsilon_{1n} \to_p 0$. We have

$$\hat{G}(x, n) - G(x, n) = n^{-\frac{1}{2}}\{K(F(x; \hat{\theta}_n), n) - K(F(x; \theta_0), n)\}$$
$$- n^{-\frac{1}{2}}W(n)\{D^{-1}(\hat{\theta}_n)\nabla_\theta F(x; \hat{\theta}_n)^t - D^{-1}(\theta_0)\nabla_\theta F(x; \theta_0)^t\}.$$

Csörgő and Révész (1979, Chapter 1) have shown:

(3.19) $$\lim_{n\to\infty}\sup_{0\leq t\leq 1}\sup_{0\leq s\leq h_n}\gamma_n|K(t + s, n) - K(t, n)| =_{\text{a.s.}} 1,$$

where $\gamma_n = (2nh_n \log h_n^{-1})^{-\frac{1}{2}}$ and h_n is a sequence of positive numbers satisfying $(\log h_n^{-1})(\log\log n)^{-1} \nearrow \infty$.

On letting $h_n = n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}$, we have $\gamma_n = \{2(n \log\log n)^{\frac{1}{2}}\log[n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}]\}^{-\frac{1}{2}}$ and hence using Taylor's theorem,

$$n^{-\frac{1}{2}}\sup_x|K[F(x; \hat{\theta}_n), n] - K[F(x; \theta_0), n]|$$
$$= n^{-\frac{1}{2}}\sup_x\Big|K\Big[F(x; \theta_0) + (\hat{\theta}_n - \theta_0)\nabla_\theta F(x; \theta_0)^t$$
(3.20) $$+ \tfrac{1}{2}(\hat{\theta}_n - \theta_0)^2\nabla_\theta^2 F(x; \theta_n^*)^t, n\Big]$$
$$- K[F(x; \theta_0), n]\Big|$$
$$=_{\text{a.s.}} O\{n^{-\frac{1}{4}}(\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}\},$$

where $\|\theta_n^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$. The latter equality of (3.20) holds by condition (3.2) (vi) and the fact that $\|\hat{\theta}_n - \theta_0\| =_{\text{a.s.}} O\{n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}\}$, if $\varepsilon_{1n} \to_{\text{a.s.}} 0$.

On letting $Q(x, \theta_n^*)$ be the matrix of partial derivatives of $D^{-1}(\theta) \cdot \nabla_\theta F(x; \theta)^t$ evaluated at $\theta = \theta_n^*$, we have

$$n^{-\frac{1}{2}} W(n) \cdot \left[D^{-1}(\hat{\theta}_n) \cdot \nabla_\theta F(x; \hat{\theta}_n)^t - D^{-1}(\theta_0) \cdot \nabla_\theta F(x; \theta_0)^t \right]$$

$$= n^{-\frac{1}{2}} W(n) \left[(\hat{\theta}_n - \theta_0) \cdot Q(x, \theta_n^*) \right]^t$$

$$=_{\text{a.s.}} O\left\{ n^{-\frac{1}{2}} \log \log n \right\},$$

by the law of the iterated logarithm for the Wiener process $W(n)$ and for the partial sum sequence of (3.2) (i), and the uniform boundedness of Q on $R \times \Lambda$, where $\|\theta_n^* - \theta_0\| \leqslant \|\hat{\theta}_n - \theta_0\|$. This, together with (3.20), implies $\varepsilon_{6n} =_{\text{a.s.}} O(n^{-\delta})$ for some $\delta > 0$ if $\varepsilon_{1n} \to_{\text{a.s.}} 0$, and hence the theorem. □

4. The maximum likelihood case. Maximum likelihood estimators often satisfy (3.2) (i) with $l(x, \theta_0) = \nabla_\theta \log f(x, \theta_0) \cdot I^{-1}(\theta_0)$, where f is the density function of F and $I^{-1}(\theta_0)$ is the inverse of the Fisher information matrix:

$$I(\theta_0) = E(\nabla_\theta \log f(X_1; \theta_0))^t \cdot (\nabla_\theta \log f(X_1; \theta_0)),$$

and $\varepsilon_{1n} \to_{\text{a.s.}} 0$ or $\varepsilon_{1n} \to_p 0$. For illuminating comments regarding this matter (the familiar Cramér-type conditions) we refer to Section 4 in Durbin (1973a). In particular, we find ourselves in agreement with his suggestion that for any particular problem a maximum likelihood or other putative efficient estimator should be first constructed and then the validity of (3.2) (i) should be checked directly. Nevertheless, we are going to illustrate here that, under certain regularity conditions, maximum likelihood estimators have a sum representation with $\varepsilon_{1n} \to_{\text{a.s.}} O(n^{-\varepsilon})$, for some $\varepsilon > 0$, in (3.2) (i). Extending the technique for θ one-dimensional by Ibragimov and Has'minskiĭ (1973b), suppose that a vector θ is estimated by $\{\hat{\theta}_n\}$, a sequence of maximum likelihood estimators. Assume the following conditions:

(i) $F(x; \theta)$ has a density function $f(x; \theta)$,
(ii) The parameter set Ξ is an open cube (bounded or unbounded) in R^p,
(iii) The density $f(x; \theta)$ is measurable with respect to $X \times \xi$, where X is the collection of Borel subsets of R and ξ is a σ-algebra of measurable subsets of Ξ,
(iv) If $\theta \neq \theta'$, then $\int |f(x; \theta) - f(x; \theta')| \, dx > 0$,
(v) There is a $\delta > 0$, such that

$$\sup_\theta \|\theta - \theta_0\|^\delta \int \left[f(x; \theta) f(x; \theta_0) \right]^{\frac{1}{2}} dx < \infty,$$

(4.1) where θ_0 is the theoretical true value of θ.
(vi) All second partial derivatives of the function $g(x, \theta) = \log f(x; \theta)$ with respect to the components of θ exist.
(vii) First and second partial differentiation of $\int f(x; \theta) \, dx$, with respect to the components of θ, can be taken under the integral sign.

(viii) The matrix $E|g''_{\theta\theta}(X_1, \theta)|^{1+\delta}$ and the vector $E|\nabla_\theta g(X_1, \theta)|^{2+\delta}$ have bounded components, for some $\delta > 0$, on compact subsets of Ξ (cf. Section 2 for notation).

(ix) There exist a function $H(x)$ and $\beta > 0$ such that $\| g''_{\theta\theta}(x, \theta_2) - g''_{\theta\theta}(x, \theta_1)\| < H(x)\|\theta_2 - \theta_1\|^\beta$, and $EH(X_1)$ exists for each $\theta \in \Xi$.

We note that (4.1) (i) to (v) correspond to conditions I and III of Ibragimov and Has'minskiĭ (1972). We have the following generalization of Theorem 1 in Ibragimov and Has'minskiĭ (1973b).

THEOREM 4.1. *If the conditions* (4.1) *are satisfied, then for some* $\varepsilon > 0$

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \cdot I(\theta_0) - n^{-\frac{1}{2}}\Sigma_{j=1}^n \nabla_\theta g(X_j, \theta_0) =_{\text{a.s.}} O(n^{-\varepsilon}),$$

where $\nabla_\theta g(x, \theta_0) = \nabla_\theta \log f(x; \theta_0)$.

REMARK 5. In view of the above theorem, under the conditions (4.1) and (3.2) (vi) to (viii), the conclusion (c) of Theorem 3.1 holds with $l(x, \theta_0) = \nabla_\theta g(x, \theta_0) \cdot I^{-1}(\theta_0)$ and $\varepsilon_{2n} =_{\text{a.s.}} O\{n^{-\varepsilon}\}$ for some $\varepsilon > 0$. In this case, the covariance of G is given by

$$\begin{aligned} EG(x, n)G(y, m) = \min(n, m) \cdot (nm)^{-\frac{1}{2}}\{&\min(F(x; \theta_0), F(y; \theta_0)) \\ &- F(x; \theta_0)F(y, \theta_0) - \nabla_\theta F(x; \theta_0) \\ &\cdot I^{-1}(\theta_0) \cdot \nabla_\theta F(y; \theta_0)'\}. \end{aligned} \tag{4.2}$$

Theorem 3.2 can be similarly adapted to Theorem 4.1.

PROOF OF THEOREM 4.1. Using the identity $\Sigma_{j=1}^n \nabla_\theta g(X_j, \hat{\theta}_n) = 0$ and the vector form of Taylor's theorem,

$$\begin{aligned} S_n &\equiv n^{-\frac{1}{2}}\Sigma_{j=1}^n \nabla_\theta g(X_j, \theta_0) - n^{-\frac{1}{2}}\Sigma_{j=1}^n[\nabla_\theta g(X_j, \hat{\theta}_n)] \\ &= n^{-\frac{1}{2}}(\theta_0 - \hat{\theta}_n) \cdot \Sigma_{j=1}^n g''_{\theta\theta}(X_j, \theta_{jn}), \end{aligned}$$

where $\|\theta_{jn} - \theta_0\| \leqslant \|\hat{\theta}_n - \theta_0\|$. By adding and subtracting $E[g''_{\theta\theta}(X_j, \theta_0)]$ and $g''_{\theta\theta}(X_j, \theta_0)$, and since $E[g''_{\theta\theta}(X_j, \theta_0)] = -I(\theta_0)$, we obtain

$$\begin{aligned} S_n = n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \cdot I(\theta_0) &+ n^{-\frac{1}{2}}(\theta_0 - \hat{\theta}_n)\Sigma_{j=1}^n\{g''_{\theta\theta}(X_j, \theta_0) - E[g''_{\theta\theta}(X_j, \theta_0)]\} \\ &+ n^{-\frac{1}{2}}(\theta_0 - \hat{\theta}_n)\Sigma_{j=1}^n\{g''_{\theta\theta}(X_j, \theta_{jn}) - g''_{\theta\theta}(X_j, \theta_0)\}. \end{aligned}$$

Denote the second and third summands in the above expression by J_2 and J_3, respectively. Then we may write

$$J_2 = n^{\frac{1}{2}-2\varepsilon}(\theta_0 - \hat{\theta}_n) \cdot n^{2\varepsilon-1}\Sigma_{j=1}^n\{g''_{\theta\theta}(X_j, \theta_0) - Eg''_{\theta\theta}(X_j, \theta_0)\}.$$

By (4.1) (viii), each component of $g''_{\theta\theta}(X_j, \theta_0)$ possesses an absolute moment of order $1 + \delta$ for some $\delta > 0$, which implies that for $2\varepsilon < \delta(1 + \delta)^{-1}$,

$$n^{2\varepsilon-1}\Sigma_{j=1}^n\{g''_{\theta\theta}(X_j, \theta_0) - Eg''_{\theta\theta}(X_j, \theta_0)\} \to_{\text{a.s.}} 0,$$

as $n \to \infty$, in view of Marcinkiewicz's Theorem (cf. Loève (1963), 4^0 of Section 16.4).

It follows from Ibragimov and Has'minskiĭ (1973a, page 86) that for any $\eta > 0$, there exists an n_0 such that for $n \geqslant n_0$, $E\|n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)\|^\eta$ exists and is bounded in n. Hence, by Markov's inequality (cf. (2.11.1) of Rényi (1970)),

$$P\left\{n^{\frac{1}{2}-2\varepsilon}\|\hat{\theta}_n - \theta_0\| > n^{-\varepsilon}\right\}$$
$$= P\left\{n^{\frac{1}{2}}\|\hat{\theta}_n - \theta_0\| > n^{\varepsilon}\right\}$$
$$\leqslant n^{-k\varepsilon}E\|n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)\|^k.$$

By the Borel-Cantelli Lemma, on choosing $k > \varepsilon^{-1}$,

$$P\left\{\limsup_{n\to\infty} n^{\frac{1}{2}-2\varepsilon}\|\hat{\theta}_n - \theta_0\| > n^{-\varepsilon}\right\} = 0,$$

that is, $J_2 =_{\text{a.s.}} O\{n^{-\varepsilon}\}$.

By (4.1) (vi) to (ix), we obtain $\|J_3\| < n^{-\frac{1}{2}}\|\hat{\theta}_n - \theta_0\|^{1+\beta}\Sigma_{j=1}^n H(X_j)$. By the SLLN, $\Sigma_{j=1}^n n^{-1}H(X_j) \to_{\text{a.s.}} EH(X_1) < \infty$. In an argument similar to that used for J_2, we obtain for $2\varepsilon < \beta$,

$$P\left\{n^{\frac{1}{2}}\|\hat{\theta}_n - \theta_0\|^{1+\beta} > n^{-\varepsilon}\right\} \leqslant n^{-k\varepsilon}E\left\{n^{\frac{1}{2}}\|\hat{\theta}_n - \theta_0\|^{(1+\beta)k}\right\}.$$

On choosing k so that $k > \varepsilon^{-1}$, $J_3 =_{\text{a.s.}} O\{n^{-\varepsilon}\}$ and hence the theorem. □

Let $n = 2m$ and

$$\bar{\alpha}_n(x) = n^{\frac{1}{2}}\left[F_n(x) - F(x; \bar{\theta}_n)\right],$$

where $\{\bar{\theta}_n\}$ is a sequence of maximum likelihood estimators based on a randomly chosen half of $X_1, X_2, \cdots, X_n$, satisfying (3.2) (i), i.e.,

$$m^{\frac{1}{2}}(\bar{\theta}_n - \theta_0) = m^{-\frac{1}{2}}\Sigma_{j=1}^m l(X_{k_j}, \theta_0) + \varepsilon_{1n},$$

with $l(x, \theta) = \nabla_\theta \log f(x, \theta)I^{-1}(\theta)$, and where F_n is based on the full sample. Durbin (1973b) points out that (under (3.2) (ii)–(iv)) $\bar{\alpha}_n(F^{-1}(\cdot; \bar{\theta}_n))$ converges weakly to a Brownian bridge (cf. also Durbin (1976)). This line of thinking was initiated by K. C. Rao (1972), and the result was also treated in Csörgő et al. (1974) and Burke and Csörgő (1976) in the manner of Theorem 3.1. Consider

$$\varepsilon_{7n} = \sup_{-\infty<x<\infty}|\bar{\alpha}_n(x) - n^{-\frac{1}{2}}K(F(x; \theta_0), n)|,$$

where K is the Kiefer process of Theorem A. The proofs in the latter two papers show that the problem of handling ε_{7n} rests in estimating $n^{\frac{1}{2}}(F(x; \theta_0) - F(x; \bar{\theta}_n)$. The latter, in turn, can now be handled by the method of proof of Theorem 3.1. Hence we have

THEOREM 4.2.
(a) *If conditions* (3.2) (ii)–(v) *hold and* $\varepsilon_{1n} \to_p 0$, *then* $\varepsilon_{7n} \to_p 0$;
(b) *If conditions* (3.2) (ii), (iii), (vi)–(viii) *hold and* $\varepsilon_{1n} \to_{\text{a.s.}} 0$, *then* $\varepsilon_{7n} \to_{\text{a.s.}} 0$,

(c) *If conditions* (3.2) (ii), (iii), (vi)–(viii) *hold and* $\varepsilon_{1n} = O\{h(n)\}$, $h(n) > 0$, $h(n) \to 0$, *then* $\varepsilon_{7n} = O\{\max(h(n), n^{-\varepsilon})\}$ *for some* $\varepsilon > 0$.

Putting together the conditions of Theorem 4.1 and part (c) here, we get $\varepsilon_{7n} = O\{n^{-\varepsilon}\}$ for some $\varepsilon > 0$.

5. Approximations under a sequence of alternatives. Suppose that the distribution function of the i.i.d. sequence is $F(x; \beta, \theta)$, where β is a p_1-dimensional vector of parameters which is assumed to be known, and θ is a p_2-dimensional vector of unknown parameters which is estimated by $\{\hat{\theta}_n\}$, based on $X_1, X_2, \cdots, X_n$. Consider the null hypothesis

(5.1) $$H_0 : (\beta, \theta) = (\beta_0, \theta_0),$$

where θ_0 stands for the theoretical true value of θ. Let

(5.2) $$\hat{\alpha}_n(x) = n^{\frac{1}{2}}\left[F_n(x) - F(x; \beta_0, \hat{\theta}_n)\right], x \in R,$$

where F_n is the empirical distribution function. In addition to H_0, we also wish to study $\hat{\alpha}_n$ under a sequence of alternatives $\{H_n\}$ defined as follows:

Let $\{\beta_n\}$ be a sequence of p_1-dimensional (nonrandom) vectors satisfying the condition

(5.3) $$\beta_n = \beta_0 + \gamma \cdot n^{-\frac{1}{2}},$$

where γ is a given constant vector. Let Λ_1 denote the closure of a given neighbourhood of β_0 and let $m = \min\{k; \beta_n \in \Lambda_1, \text{ for all } n \geqslant k > 2\}$. Then, consider

(5.4) $$H_n : (\beta, \theta) = (\beta_n, \theta_0),$$

for $n = m, m + 1, \cdots$ where β_n satisfies (5.3). If we choose $\beta_n = \beta_0$ for all n, i.e., $\gamma = 0$, then H_n and H_0 are identical.

Again we list all the conditions whose appropriate subcollections will be used in this section.

(5.5)

(i) Under H_n:
$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = n^{-\frac{1}{2}}\Sigma_{j=1}^n l(X_j, \beta_0, \theta_0) + A\gamma^t + \varepsilon_{8n},$$
where A is a given finite matrix of order $p_2 \times p_1$, l is a measurable p_2-dimensional vector valued function, and ε_{8n} converges to zero in a manner to be specified.

(ii) $E\{l(X_j, \beta_0, \theta_0)|H_n\} = 0$ for $n = 0$ and $n \geqslant m$.

(iii) $E\{l(X_j, \beta_0, \theta_0)^t l(X_j, \beta_0, \theta_0)|H_n\} = M(\beta_n, \theta_0)$, a finite nonnegative definite matrix for each $n \geqslant m$ which converges to a finite nonnegative matrix $M = M(\beta_0, \theta_0)$ as $n \to \infty$.

(iv) The vector $\nabla_\beta F(x; \beta, \theta_0)$ is uniformly continuous in x and $\beta \in \Lambda_1$, and the vector $\nabla_\theta F(x; \beta_0, \theta)$ is uniformly continuous in x and $\theta \in \Lambda_2$, where Λ_2 is the closure of a given neighbourhood of θ_0.

(v) Each component of $l(x, \beta_0, \theta_0)$ is of bounded variation on each finite interval.

(vi) The vectors $\nabla_\beta F(x; \beta_0, \theta_0)$, $\nabla_\theta F(x; \beta_0, \theta_0)$ are uniformly bounded in x, while the vector $\nabla^2_\beta F(x; \beta, \theta_0)$ is uniformly bounded in x and $\beta \in \Lambda_1$, and the vector $\nabla^2_\theta F(x; \beta_0, \theta)$ is uniformly bounded in x and $\theta \in \Lambda_2$.

(vii) Condition (3.2) (vii) holds for the vector $l(F^{-1}(s; \beta_0, \theta_0), \beta_0, \theta_0)$, where $F^{-1}(s; \beta, \theta) = \inf\{x : F(x; \beta, \theta) \geqslant s\}$.

(viii) Condition (3.2) (viii) holds for the vector $l(F^{-1}(s; \beta_0, \theta_0), \beta_0, \theta_0)$.

The estimated empirical process $\hat{\alpha}_n(x)$ of (5.2), under the sequence of alternatives $\{H_n\}$ of (5.4), will be estimated by the two-parameter Gaussian process

$$Z(x, n) = G(x, n) - A\gamma^t \nabla_\theta F(x; \beta_0, \theta_0)^t + \gamma \nabla_\beta F(x; \beta_0, \theta_0)^t, \tag{5.6}$$

with

$$G(x, n) = n^{-\frac{1}{2}}K(F(x; \beta_0, \theta_0), n)$$
$$- \left\{\int l(x, \beta_0, \theta_0) d_x n^{-\frac{1}{2}} K(F(x; \beta_0, \theta_0), n)\right\} \nabla_\theta F(x; \beta_0, \theta_0)^t.$$

This process $G(x, n)$ is the same process as defined by (3.3). The mean of Z is

$$EZ(x, n) = -A\gamma^t \nabla_\theta F(x; \beta_0, \hat{\theta}_n)^t + \gamma \nabla_\beta F(x; \beta_0, \hat{\theta}_n)^t,$$

and its covariance is given by (3.4), with the obvious changes in notation.

As in Section 3, on letting

$$\hat{Z}(x, n) = \hat{G}(x, n) - A\gamma^t \nabla_\theta F(x; \beta_0, \hat{\theta}_n)^t + \gamma \nabla_\beta F(x; \beta_0, \hat{\theta}_n)^t,$$

where $\hat{G}$ is defined by (3.17) (with the notation suitably modified), the results corresponding to Theorem 3.2 continue to hold under the sequence of alternatives $\{H_n\}$.

THEOREM 5.1. *Suppose that conditions* (5.5) (i)–(iii) *hold, and let*

$$\varepsilon_{9n} = \sup_{x \in R} |\hat{\alpha}_n(x) - Z(x, n)|.$$

Then, under the sequence of alternatives $\{H_n\}$,

(a) $\varepsilon_{9n} \to_p 0$, *if conditions* (5.5) (iv), (v) *hold and* $\varepsilon_{8n} \to_p 0$;

(b) $\varepsilon_{9n} \to_{\text{a.s.}} 0$, *if conditions* (5.5) (vi)–(viii) *hold and* $\varepsilon_{8n} \to_{\text{a.s.}} 0$;

(c) $\varepsilon_{9n} =_{\text{a.s.}} 0\{\max(h(n), n^{-\varepsilon})\}$ *for some* $\varepsilon > 0$, *if conditions* (5.5) (vi)–(viii) *hold and* $\varepsilon_{8n} =_{\text{a.s.}} O\{h(n)\}$, $h(n) > 0$, $h(n) \to 0$.

REMARK 6. Here the whole content (modified to the present situation) of Remarks 2 and 3 in Section 3 can be repeated. Specifically, Durbin (1973a) proved the weak convergence of $\hat{\alpha}_n(F^{-1}(\cdot; \beta_0, \hat{\theta}_n))$, under $\{H_n\}$, to a process that can be represented by putting a Brownian bridge $B(\cdot)$ into the definition (5.6) of Z in place of $n^{-\frac{1}{2}}K(\cdot, n)$. He used conditions (5.5) (i)–(iv) (with $\varepsilon_{8n} \to_p 0$) to prove this (he requires (5.5) (iv) in a slightly stronger form than ours), and the extra condition that $F(x; \beta, \theta)$ is continuous in x for all (β, θ) in some neighbourhood of (β_0, θ_0). If we want to prove this weak convergence (but not the two-parameter representation in (a)) without condition (5.5) (v), then the method proposed in Remark 2 works again with

$$Y_n(s) = n^{-\frac{1}{2}}K(s, n) - \left\{n^{-\frac{1}{2}}\Sigma_{j=1}^n l(X_j, \beta_0, \theta_0) + A\gamma^t\right\} \nabla_\theta F\left(F^{-1}(s; \beta_0, \theta_0); \beta_0, \theta_0\right)^t$$
$$+ \gamma \nabla_\beta F\left(F^{-1}(s; \beta_0, \theta_0); \beta_0, \theta_0\right)^t.$$

REMARK 7. Durbin (1973a) proves the weak convergence of

$$\hat{\delta}_n(s) = n^{\frac{1}{2}}\left[\hat{F}_n(s) - s\right], 0 \leqslant s \leqslant 1,$$

where $\hat{F}_n(s)$ is the proportion of $F(X_1; \beta_0, \hat{\theta}_n), \cdots, F(X_n; \beta_0, \hat{\theta}_n)$ which satisfy $F(X_j; \beta_0, \hat{\theta}_n) \leqslant s$. The processes $\hat{\delta}_n$ and $\hat{\alpha}_n$ are asymptotically equivalent. In order to see this we first note that $\hat{\delta}_n(s) = \hat{\alpha}_n(F^{-1}(s; \beta_0, \hat{\theta}_n))$, where now $F^{-1}(s; \beta_0, \theta) = \sup\{x : F(x; \beta_0, \theta) \leqslant s\}$. Secondly, if we assume that we have already carried out the program of the last sentence of Remark 6, one can easily prove

$$\sup_x|\hat{\alpha}_n(x) - Y_n\left(F(x; \beta_0,\hat{\theta}_n)\right)| \to_p 0, \qquad \text{as } n \to \infty.$$

Hence on putting $x = F^{-1}(s; \beta_0, \hat{\theta}_n)$, we have

$$\sup_{0\leqslant s\leqslant 1}|\hat{\delta}_n(s) - Y_n(s)| \to_p 0, \qquad \text{as } n \to \infty,$$

which establishes asymptotic equivalence.

PROOF OF THEOREM 5.1. By adding and subtracting, we have under H_n

$$\hat{\alpha}_n(x) = Q_{1n}(x) + Q_{2n}(x) - Q_{3n}(x),$$

where Q_{1n}, Q_{2n} and Q_{3n} are defined respectively in the first rows of (5.7), (5.9) and (5.10).

For the first term we have

$$\begin{aligned} Q_{1n}(x) &= n^{\frac{1}{2}}\left[F_n(x) - F(x; \beta_n, \theta_0)\right] \\ &= n^{-\frac{1}{2}}K(F(x; \beta_n, \theta_0), n) + \varepsilon_{3n}(F(x; \beta_n, \theta_0)) \qquad (5.7) \\ &= n^{-\frac{1}{2}}K(F(x; \beta_0, \theta_0), n) + \varepsilon_{10n}(x) + \varepsilon_{3n}(F(x; \beta_n, \theta_0)), \end{aligned}$$

where (cf. (3.14)) by Theorem A

$$\sup_x|\varepsilon_{3n}(F(x; \beta_n, \theta_0))| =_{\text{a.s.}} O\left\{n^{-\frac{1}{2}}\log^2 n\right\},$$

and by the modulus of continuity (3.19) of the Kiefer process,

$$\sup_x|\varepsilon_{10n}(x)| =_{\text{a.s.}} O\{n^{-\delta}\}, \qquad (5.8)$$

for any δ satisfying $0 < \delta < \frac{1}{4}$.

For the second term, when proving (a),

$$\begin{aligned} Q_{2n}(x) &= n^{\frac{1}{2}}\left[F(x; \beta_n, \theta_0) - F(x; \beta_0, \theta_0)\right] \\ &= n^{\frac{1}{2}}(\beta_n - \beta_0)\nabla_\beta F(x; \beta_n^*, \theta_0)^t \qquad (5.9) \\ &= \gamma\nabla_\beta F(x; \beta_0, \theta_0)^t + \varepsilon_{11n}(x), \end{aligned}$$

where $\|\beta_n^* - \beta_0\| \leq \|n^{-\frac{1}{2}}\gamma\|$, and hence by condition (5.5) (iv)

$$\sup_x|\varepsilon_{11n}(x)| \to 0.$$

When proving (b) and (c),

$$Q_{2n}(x) = \gamma\nabla_\beta F(x;\beta_0,\theta_0)' + \varepsilon_{12n}(x),$$

where, by condition (5.5) (vi),

$$\sup_x|\varepsilon_{12n}(x)| = \sup_x|n^{\frac{1}{2}}(\gamma^2/n)\nabla_\beta^2 F(x;\beta_n^*,\theta_0)'| = O\{n^{-\frac{1}{2}}\}.$$

For the third term we can repeat, under H_n, the proof of Theorem 3.1 to get

$$\begin{aligned} Q_{3n}(x) &= n^{\frac{1}{2}}\big[F(x;\beta_0,\hat{\theta}_n) - F(x;\beta_0,\theta_0)\big] \\ &= \big\{\textstyle\int l(x;\beta_0,\theta_0)d_x n^{-\frac{1}{2}}K(F(x;\beta_0,\theta_0),n) + A\gamma'\big\}\nabla_\theta F(x;\beta_0,\theta_0)' \\ &\quad + \varepsilon_{13n}(x), \end{aligned} \tag{5.10}$$

where $\sup_x|\varepsilon_{13n}(x)|$ behaves like ε_{9n} in (a), (b) or (c) in the formulation of the theorem. Combining this behaviour with that of the second term, we have the theorem since the first term was shown to behave always well. □

REMARK 8. If we assume that the function l possesses not only a finite second (cf. (5.5) (iii)) but a finite absolute moment of order r, $r > 2$, then we can proceed the following way. Let $D(\beta_n,\theta_0)$ be the nonsingular matrix for which $D(\beta_n,\theta_0)'\cdot M(\beta_n,\theta_0)D(\beta_n,\theta_0)$ satisfies (3.5). Then, by Komlós-Major-Tusnády (1976) and Major (1976),

$$\|\Sigma_{j=1}^n l(X_j,\beta_0,\theta_0)D(\beta_n,\theta_0)^{-1} - W(n)\| =_{\text{a.s.}} o(n^{1/r}),$$

where $W(n)$ is a vector-valued Wiener process (cf. 3.6). If the underlying probability space is still richer (if necessary), then there exists a Kiefer process $\tilde{K}$ such that $W(n) = \int l(x,\beta_0,\theta_0)d_x\tilde{K}(F(x;\beta_0,\theta_0),n)D(\beta_n,\theta_0)$. Let $\varepsilon_{14n} = \sup_x|\hat{\alpha}_n(x) - \tilde{Z}(x,n)|$, where $\tilde{Z}(x,n)$ has the same form as $Z(x,n)$ in (5.6) with $\tilde{G}(x,n)$ in place of $G(x,n)$, where

$$\begin{aligned} \tilde{G}(x,n) &= n^{-\frac{1}{2}}K(F(x;\beta_0,\theta_0),n) \\ &\quad - \big\{\textstyle\int l(x;\beta_0,\theta_0)d_x n^{-\frac{1}{2}}\tilde{K}(F(x;\beta_0,\theta_0),n)\big\}\nabla_\theta F(x;\beta_0,\theta_0)'. \end{aligned} \tag{5.11}$$

The above proof shows the following. Under the rth moment condition and (only) (5.5) (i)–(iv) we have $\varepsilon_{14n} \to_p 0$, if $\varepsilon_{8n} \to_p 0$, while under the rth moment condition and (only) (5.5) (i)–(iii), (vi) we have $\varepsilon_{14n} \to_{\text{a.s.}} 0$, if $\varepsilon_{8n} \to_{\text{a.s.}} 0$. Moreover, if in the latter case $\varepsilon_{8n} = O\{h(n)\}$, then $\varepsilon_{14n} =_{\text{a.s.}} O\{\max(h(n), n^{-\tau})\}$, where $\tau = \min(\delta, \lambda)$, with $\lambda = (r-2)/2r$, and δ is from (5.8). Naturally, the same type of "results" hold in the simpler setting of Section 3, i.e., under H_0, with $\tau = \lambda$. These "results" are entirely useless at the present stage, since we do not know anything about the joint distribution of K and $\tilde{K}$ in (5.11). Since $\varepsilon_{14n} \to 0$, it follows from Durbin's weak convergence theorem that $\tilde{G}(\cdot, n)$ converges weakly to $D(\cdot)$ of (3.7). The problem is how to replace $\tilde{K}$ by K in (5.11), so that we should not have to fall back to Durbin's weak convergence theorem in order to make sense out of $\tilde{G}(x, n)$. This was achieved in Lemmas 3.1 and 3.2 by imposing (very mild) extra restrictions on l. However, we conjecture that this should be possible without the latter restrictions. It appears that the proof of this conjecture (under the rth moment condition and only (5.5) (i)–(iii)) would require an extension of the proof of the Komlós-Major-Tusnády approximation to simultaneously approximate α_n and $\Sigma_{j=1}^{n} l(X_j)$.

6. An approximation of the multivariate empirical process when parameters are estimated. Here the role of Theorem A will be played by the following strong approximation result.

THEOREM B (Philipp and Pinzur (1978)). *Let $X_1, X_2, \cdots$ be a sequence of independent k-dimensional random vectors with common distribution function F. Let $F_n(x)$, $x \in R^k$ denote the proportion of the X_i, $1 \leqslant i \leqslant n$, which satisfy $X_i \leqslant x$ in the usual partial ordering of R^k, and let $\alpha_n(x) = n^{\frac{1}{2}}[F_n(x) - F(x)]$, $x \in R^k$, be the multivariate empirical process. Then, if the underlying probability space is rich enough, one can construct a Gaussian process $K_F(x, n)$ such that*

$$\sup_{x \in R^k} |\alpha_n(x) - n^{-\frac{1}{2}} K_F(x, n)| =_{\text{a.s.}} O\{(\log n)^{-\lambda}\},$$

for some $\lambda > 0$. The process K_F has mean $EK_F(x, n) = 0$ and covariance function

$$EK_F(x, n)K_F(y, m) = (n \wedge m)\{F(x \wedge y) - F(x)F(y)\},$$

where $x \wedge y = (x_1 \wedge y_1, \cdots, x_k \wedge y_k)$, $x = (x_1, x_2, \cdots, x_k)$ and $y = (y_1, y_2, \cdots, y_k)$.

In this section the setting will be that of Secton 3, i.e., alternatives are not considered. If F belongs to some family $F(\cdot; \theta)$ of multivariate distribution functions, then the estimated empirical process is

$$\hat{\alpha}_n(x) = n^{\frac{1}{2}}\left[F_n(x) - F(x; \hat{\theta}_n)\right], x \in R^k,$$

where $\hat{\theta}_n$ is a sequence of estimators satisfying (3.2) (i) with $X_j \in R^k$. We will estimate this process by the Gaussian process ($x \in R^k$, and F_0 stands for the distribution function $F(\cdot; \theta_0)$)

$$G_{F_0}(x, n) = n^{-\frac{1}{2}} K_{F_0}(x, n) - \left\{\int_{R^k} l(x, \theta_0)\, dn^{-\frac{1}{2}} K_{F_0}(x, n)\right\} \nabla_\theta F(x; \theta_0)^t.$$

$EG_{F_0}(x, n) = 0$, and the covariance of G_{F_0} is given by the right-hand side of (3.4), where the domain of integration in $J(x)$ is now $\{y \in R^k : y \leqslant x\}$.

We will use the k-dimensional versions of conditions (3.2) (i)–(v), where in (v) the word "interval" is now understood as a rectangle in R^k, parallel with the coördinate axes.

Let $\alpha_n(x; \theta_0) = n^{\frac{1}{2}}[F_n(x) - F(x; \theta_0)]$, $x \in R^k$. The k-dimensional version of Lemma 3.1 states that if (3.2) (iii) and (v) are satisfied, then

$$L_n = \int_{R^k} l(x, \theta_0) d_x\left[\alpha_n(x; \theta_0) - n^{-\frac{1}{2}} K_{F_0}(x, n)\right] \to_p 0.$$

If in the proof of Lemma 3.1 $T_j(x)$ means the total variation of $l(\cdot, \theta_0)$ over the k-dimensional cube centred at the origin and having k-dimensional volume $(2x)^k$, and if u_n is chosen so that $\|T(u_n)\|(\log n)^{-\lambda} \to 0$, then the proof remains valid, provided that $|x| > u_n$ and $|x| \leqslant u_n$ are meant as $|x_1|, \cdots, |x_k| > u_n$ and $|x_1|, \cdots, |x_k| \leqslant u_n$ (and also the inequalities $x \leqslant -u_n$, $x \geqslant u_n$ are understood this way, i.e., componentwise). The very reason for this is that the Kiefer measure $n^{-\frac{1}{2}}K_{F_0}(A, n)$ of a k-dimensional rectangle A, parallel with the coordinate axes (defined by the usual inclusion-exclusion procedure), has independent values for disjoint rectangles A_1 and A_2. Hence the k-dimensional variant of formula (3.10) for computing the corresponding variance continues to hold. Thus the proof of part (a) of Theorem 3.1 goes through, and we have

THEOREM 6.1. *If conditions* (3.2) (i)–(v) *are satisfied, then*

$$\sup_{x \in R^k}|\hat{\alpha}_n(x) - G_{F_0}(x, n)| \to_p 0.$$

Let h be a functional on the space of real valued functions on R^k endowed with the supremum topology, and consider the Gaussian process $B_{F_0}(x)$, $x \in R^k$, with $EB_{F_0}(x) = 0$, $EB_{F_0}(x)B_{F_0}(y) = F(x \wedge y; \theta_0) - F(x; \theta_0)F(y; \theta_0)$. Let

$$D_{F_0}(x) = B_{F_0}(x) - \left\{\int_{R^k} l(x, \theta_0)\, dB_{F_0}(x)\right\} \nabla_\theta F(x; \theta_0)',$$

the covariance function of which is given by { } of (3.4) with $J(x)$ understood as noted above.

COROLLARY. *If h is continuous, then in each continuity point t of the limit distribution function*

$$\lim_{n\to\infty} P\{h(\hat{\alpha}_n(\cdot)) \leqslant t\} = P\{h(D_{F_0}(\cdot)) \leqslant t\}.$$

It would be perhaps also desirable to prove some analogues of parts (b) and (c) of Theorem 3.1 and to pursue the program of Section 5 for the estimated multivariate empirical process.

Acknowledgment. The authors wish to thank the Associate Editor and the referee for their helpful comments.

REFERENCES

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

BRILLINGER, D. R. (1969). An asymptotic representation of the sample df. *Bull. Amer. Math. Soc.* **75** 545–547.

BURKE, M. D. and CSÖRGŐ, M. (1976). Weak approximations of the empirical process when parameters are estimated. In *Empirical Distributions and Processes* (P. Gaenssler and P. Révész, eds.) *Lecture Notes in Mathematics* **566** 1–16, Springer-Verlag, Berlin.

CSÁKI, E. (1975). Some notes on the law of the iterated logarithm for empirical distribution functions. In *Limit Theorems of Probability Theory* (P. Révész, ed.). *Coll. Math. Soc. J. Bolyai* **11** 47–57. Keszthely (Hungary), 1974.

CSÁKI, E. (1977). The law of the iterated logarithm for normalized empirical distribution functions. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **38** 147–167.

CSÖRGŐ, M., KOMLÓS, J., MAJOR, P., RÉVÉSZ, P. and TUSNÁDY, G. (1974). On the empirical process when parameters are estimated. *Trans. Seventh Prague Conference* 1974, Academia, Prague 1977, 87–97.

CSÖRGŐ, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen type laws of invariance principle. II. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **31** 261–269.

CSÖRGŐ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882–894.

CSÖRGŐ, M. and Révész, P. (1979). *Strong Approximations in Probability and Statistics*. Book manuscript in progress.

CSÖRGŐ, S. (1978). Limit behaviour of the empirical characteristic function. Unpublished manuscript.

CSÖRGŐ, S. (1979). The empirical characteristic process when parameters are estimated. Manuscript in progress.

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

DURBIN, J. (1973a). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

DURBIN, J. (1973b). *Distribution Theory for Tests Based on the Sample Distribution Function*. Regional Conference Series on Appl. Math. **9**. SIAM, Philadelphia.

DURBIN, J. (1976). Kolmogorov-Smirnov tests when parameters are estimated. In *Empirical Distributions and Processes* (P. Gaenssler and P. Révész, ed.). *Lecture Notes in Mathematics* **566**, 33–44. Springer-Verlag, Berlin.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case I. Study of the likelihood ratio. *Theor. Probability Appl.* **17** 445–462.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1973a). Asymptotic behavior of some statistical estimators II. Limit theorems for the a posteriori density and Bayes' estimators. *Theor. Probability Appl.* **18** 76–91.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1973b). On the approximation of statistical estimators by sums of independent variables. *Soviet Math. Dokl.* **14** 883–887.

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.

KIEFER, J. (1972). Skorohod embedding of multivariate R.V.'s and the sample D.F. *Z. Wahrscheinlichkeitstheorie und verw, Gebiete* **24** 1–35.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample D.F. I. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **32** 111–131.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent R.V.'s, and the sample D.F. II. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **34** 33–58.

MAJOR, P. (1976). The approximation of partial sums of independent RV's. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **35** 213–220.

NEUHAUS, G. (1974). Asymptotic properties of the Cramér-von Mises statistic when parameters are estimated. In *Proc. Prague Symp. Asymptotic Statist.* 1973 (J. Hajek, ed.) **2** 257–297. Charles Univ. Prague.

NEUHAUS, G. (1976a). Asymptotic power properties of the Cramér-von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6** 95–110.

NEUHAUS, G. (1976b). Weak convergence under contiguous alternatives of the empirical process when parameters are estimated: the D_k approach. In *Empirical Distributions and Processes* (P. Gaenssler and P. Révész, eds.) *Lecture Notes in Mathematics* **566** 68–82, Springer-Verlag, Berlin.

LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand-Reinhold, Princeton.

PHILIPP, W. and PINZUR, L. (1978). An almost sure approximation for the multivariate empirical process. Manuscript in progress.

RAO, K. C. (1972). The Kolmogorov, Cramér-von Mises chi-square statistics for goodness-of-fit tests in the parametric case. Abstract: *IMS Bulletin*, **1** 87.

RÉNYI, A. (1970). *Foundations of Probability*, Holden Day, San Francisco.

M. D. BURKE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF CALGARY
CALGARY, ALBERTA T2N 1N4
CANADA

M. CSÖRGŐ
DEPARTMENT OF MATHEMATICS
CARLETON UNIVERSITY
COLONEL BY DRIVE
OTTAWA K1S 5B6
CANADA

S. Csörgő
Bolyai Institute
Szeged University
H-6726 Szeged
Aradi vértanúk tere 1
Hungary

P. Révész
Mathematical Institute of the
Hungarian Academy of Sciences
H-1053 Budapest
Reáltanoda u. 13-15
Hungary