Acta Sci. Math., 48 (1985), 201—213

Empirical kernel transforms of parameter-estimated empirical processes

LAJOS HORVÁTH

In honour of Professor Károly Tandori on his 60th birthday

1. Introduction. Let $d \geqq 1$ be an integer and let $X_1, X_2, \ldots$ be a sequence of independent d-dimensional random vectors with common distribution function $F(x)$, $x \in R^d$. We assume that a parametric family of d-variate distribution functions is given,

$$\mathscr{F} = \{F(x, \theta): x \in R^d;\ \theta \in \Theta \subset R^p\},$$

and the common distribution function of the $X_1, X_2, \ldots$ belongs to this family, i.e., there is a parameter $\theta_0 \in \Theta$ so that $F(x) = F(x; \theta_0) \equiv F_0(x)$. The true value θ_0 is unknown. Consider the estimated empirical process defined by

$$\beta_n(x) = n^{1/2}(F_n(x) - F(x;\ \theta_n)), \quad x \in R^d,$$

where F_n is the empirical distribution function of $X_1, \ldots, X_n$ and $\theta_n = (\theta_{n1}, \ldots, \theta_{np})$ is some estimator of θ_0 based on the random sample $X_1, \ldots, X_n$.

The weak convergence of the estimated empirical process was studied by several authors. We will use the general strong approximation result of BURKE *et al* [1] in this note. Introduce the notations $\theta = (\theta_1, \ldots, \theta_p)$ and

$$\nabla F(x;\ \theta^*) = \nabla_\theta F(x; \theta)|_{\theta=\theta^*} = \left(\frac{\partial}{\partial \theta_1} F(x;\ \theta), \ldots, \frac{\partial}{\partial \theta_p} F(x;\ \theta)\right)\bigg|_{\theta=\theta^*},$$

and let

$$\langle x, y\rangle = \sum_{j=1}^{p} x_j y_j, \quad x = (x_1, \ldots, x_p), \quad y = (y_1, \ldots, y_p),$$

stand for the inner product in R^p. Let $a^{\mathrm{T}} = (a_1, \ldots, a_p)^{\mathrm{T}}$ denote the column vector corresponding to the row vector $a = (a_1, \ldots, a_p)$. The norm of a vector $x = (x_1, \ldots, x_p)$ and a matrix $M = \{m_{ij}\}_{i,j=1}^{p}$ is defined by $\|x\| = \max\{|x_i|: 1 \leqq i \leqq p\}$ and

Received May 9, 1983.

$\|M\|=\max\{|m_{ij}|:\ 1\leqq i,j\leqq p\}$. If ξ_n converges to zero in probability we will use the notation $\xi_n \xrightarrow{P} 0$ $(n\to\infty)$. A Brownian bridge $B^{F_0}(x)$, $x\in R^d$, associated with the distribution function F_0 is a d-variate Gaussian random field such that $EB^{F_0}(x)=0$ and $EB^{F_0}(x)B^{F_0}(y)=F_0(x\wedge y)-F_0(x)F(y)$, where $x\wedge y=(\min(x_1, y_1), \dots, \min(x_d, y_d))$.

Theorem A (BURKE, M. CSÖRGŐ, S. CSÖRGŐ and RÉVÉSZ [1], and S. CSÖRGŐ [2]). *Suppose that the sequence θ_n satisfies the following conditions:*

(i)
$$n^{1/2}(\theta_n-\theta_0) = n^{-1/2}\sum_{j=1}^{n} l(X_j;\theta_0)+\varepsilon_n,$$

where $l(\cdot\,;\theta_n)$ is a measurable d-dimensional (row) vector-valued function and $\varepsilon_n \xrightarrow{P} 0$ $(n\to\infty)$.

(ii)
$$El(X_1;\theta_0) = 0.$$

(iii) $M(\theta_0)=El^T(X_1;\ \theta_0)l(X_1;\ \theta_0)$ *is a finite and 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) $d=1$; *Each component of the vector function $l(x;\ \theta_0)$ is of bounded variation on each finite interval.*

$d>1$; *The partial derivatives of each component of the vector function l, with order not exceeding d, exist almost everywhere (with respect to the d-dimensional Lebesgue measure) on R^d, and for any $u>0$*

$$\sup_{\|x\|\leqq u}\sum_{j=1}^{d}\sum_{\substack{j_1,\dots,j_d\geqq 0\\ j_1+\dots+j_d=j}}\left\|\frac{\partial^j l(x;\ \theta_0)}{\partial x_1^{j_1}\dots\partial x_d^{j_d}}\right\|<\infty.$$

If the underlying probability space is rich enough, one can define a sequence of d-dimensional Brownian bridges $\{B_n^{F_0}(x)\}$ associated with the distribution function F_0 such that

$$\sup_{x\in R^d}|\beta_n(x)-D_n(x;\ \theta_0)|\xrightarrow{P} 0\quad (n\to\infty),$$

where

$$D_n(x;\ \theta_0) = B_n^{F_0}(x)-\Big\langle \int_{R^d} l(y;\ \theta_0)\,dB_n^{F_0}(y),\ \nabla_\theta F(x;\ \theta_0)\Big\rangle$$

is a sequence of copies of the Durbin process.

The limiting Gaussian process of this theorem depends, in general, not only on F but also on θ_0, the true, unknown value of parameter. On the other hand, the distributions of the functionals of $D_1(x;\ \theta_0)$ (supremum functional, square integral functional) as functions of θ_0 are unknown. According to the references below, Bolshev asked whether there is a kernel k such that the random variable

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(x, y)\, dD_1(x;\, \theta_0) D_1(y;\, \theta_0), \quad d = 1, \quad x, y \in R^1$$

have a prescribed distribution. This problem was investigated by HMALADZE [6], [7] and DZAPARIDZE and NIKULIN [5]. The methods of Dzaparidze and Nikulin are based on the orthogonal expansion of the limiting Gaussian process. In the case $d=1$ and when only shift and scale parameters are estimated they proposed statistics whose limit distributions are independent of the unknown parameters and depend only on F, but these limit distributions are usually complicated and therefore it would be hard to compute percentage points for these statistics. Using the martingale property of $\beta_n(x)$ if $d=1$, Hmaladze proved some weak convergence results in L_2 sense. Analogous results were obtained earlier by NEUHAUS [10], [11]. He proved the weak convergence of $\beta_n(x)$, $x \in [0, 1]^d$, $d \geqq 1$, in supremum metric under contiguous alternatives.

The above question was generalized by S. CSÖRGŐ [2], who introduced

$$\int_{R^d} k(x, y)\, d\beta_n(x), \quad x \in R^d, \quad y \in R^q,$$

a kernel transform of the parameter-estimated empirical process. Here $q \geqq 1$ is an arbitrary integer. Assuming some regularity conditions on k, he proved that

$$\sup_{y \in R^q} \left| \int_{R^d} k(x, y)\, d\beta_n(x) - \int_{R^d} k(x, y)\, dD_n(x;\, \theta_0) \right| \xrightarrow{P} 0, \quad n \to \infty.$$

Unfortunately, he cannot choose a kernel k and a functional h on the space of continuous functions on R^q such that the random variable

$$h\left(\int_{R^d} k(x, y)\, dD_1(x;\, \theta_0) \right)$$

has a distribution not depending on θ_0.

In this note we are interested in a sequence of kernel type transformations of $\beta_n(x)$, where the kernel will also depend on the sample. We are able to choose a sequence of kernels $\{k^N(x, y;\, \theta_0)\}$ such that

$$\int_{R^d} k^N(x, y) \beta_n(x)\, dx, \quad y \in I^q, \tag{1.1}$$

converges weakly to a q-dimensional standard Wiener process or to the standard Brownian bridge on I^q (or, for that matter, to any prescribed Gaussian process) if N and $n(N)$ go to infinity, where I^q is the unit cube of R^q. The transformations (1.1) depend, in general, on the unknown parameter θ_0, therefore we will prove, that there is a sequence of random kernels $\{k^n_{N,m}(x, y)\}$ based on the sample $X_1, \ldots, X_n$ such that

$$\int_{R^d} k^n_{N,m}(x, y)\,\beta_n(x)\,dx, \quad y\in I^q,$$

converges weakly to a prescribed process, if $N, m(N), n(N, m)$ go to infinity. Our methods may be extended to some more general parametric models. Section 2 presents a general result, where the role of $\beta_n(x)$ is played by an arbitrary sequence of processes $X_n(x)$. Section 3 will then specify the result for $\beta_n(x)$.

2. Main theorem. The reproducing kernel Hilbert space $H(R)$ generated by a covariance function $R(t, s)$ plays a fundamental role in our note. Suppose that $X(t)$, $t\in I^d$, is a centered Gaussian process having continuous paths on I^d a.s. and continuous covariance function:

$$EX(t) = 0, \quad EX(t)X(s) = R(t, s), \quad t, s\in I^d.$$

It is well known, that the space $\mathscr{C}$ of all continuous functions $I^d \to R$ with the topology induced by the supremum norm is a separable Banach space and the collection $\mathscr{C}^*$ of all linear and bounded functionals on $\mathscr{C}$ can be identified with the space of all (regular) measures ν on the Borel subsets of I^d. If ν^+ and ν^- denote the Hahn decomposition of ν, then $\|\nu\| = \nu^+(I^d)+\nu^-(I^d)$ is a norm on $\mathscr{C}^*$. Sato [12] has shown, that for some complete orthonormal sequence (CONS) $\{e_i(t), i\geqq 1\}$ in $H(R)$ one can write

$$e_i(t) = \int_{I^d} R(t, s)\,\nu_i(ds),$$

where $\nu_i\in\mathscr{C}^*$ and

(2.1) $$\nu_i = \mu_i/\sigma_i, \quad \mu_i\in\mathscr{C}^*,$$

(2.2) $$\|\mu_i\| = 1,$$

(2.3) $$\sigma_i = \Big[\int_{I^d}\int_{I^d} R(t, s)\,\nu_i(ds)\,\nu_i(dt)\Big]^{1/2} > 0,$$

(2.4) $$\int_{I^d}\int_{I^d} R(t, s)\,\nu_i(ds)\,\nu_j(dt) = 0, \quad i\neq j.$$

Mangano [9] proved that

(2.5) $$\int_{I^d} e_i(t)\,\nu_j(dt) = \delta_{ij},$$

where $\delta_{ii}=1$, $\delta_{ij}=0$, $i\neq j$. The following lemma is a simple variant of Lemma 2.2 in [9].

Lemma 2.1. *Let $R(t, s)$ and $G(t, s)$ be continuous covariance functions of two centered Gaussian processes with a.s. continuous paths on I^d. Let N be a positive integer* *and* $\{e_i(t),\ i\geqq 1,\ t\in I^d\}$ *be a CONS in* $H(R)$ *generated by measures* $\{\nu_i,\ i\geqq 1\}$. *If*

$$\sup_{(t,s)\in I^{2d}} |G(t,s)-R(t,s)| = \Delta \leqq 1/K_1, \tag{2.6}$$

then there exists an orthonormal set $\{f_i,\ 1\leqq i\leqq N\}$ *of functions in* $H(G)$ *generated by the measures* $\{\varkappa_i,\ 1\leqq i\leqq N\}$ *such that*

$$\|\varkappa_i-\nu_i\| \leqq \Delta K_2, \quad 1\leqq i\leqq N \tag{2.7}$$

and

$$\sup_{t\in I^d} |e_i(t)-f_i(t)| \leqq \Delta K_3, \quad 1\leqq i\leqq N, \tag{2.8}$$

where K_1, K_2, K_3 *are suitably chosen polynomials of* $N, M, \|\nu_i\|,\ 1\leqq i\leqq N$, *with positive coefficients and* $M=\max_{(t,s)\in I^{2d}} |R(t,s)|$.

Proof. The proof follows from the construction of Mangano. He constructed measures $\varkappa_i$, $1\leqq i\leqq N$, which are linear combinations of the measures ν_i, $1\leqq i\leqq N$. It is not too difficult to check that the measures and functions $f_i(t)$, $1\leqq i\leqq N$, constructed by Mangano satisfy (2.7) and (2.8) with suitably chosen functions K_1, K_2, K_3.

Let $\{X_n(t),\ t\in I^d\}$ be a sequence of stochastic processes such that

$$\sup_{t\in I^d} |X_n(t)-Y_n(t)| \xrightarrow{P} 0 \quad (n\to\infty), \tag{2.9}$$

where $\{Y_n(t),\ t\in I^d\}$ is a sequence of copies of a Gaussian process $\{Y(t;\theta),\ t\in I^d\}$ depending on a parameter θ. We suppose, that the process $Y(t;\theta)$ has continuous paths on I^d a.s., its covariance function is continuous for every $\theta\in\Theta$, where $\Theta\subset R^p$ is a compact parameter set, and

$$\sup_{(t,s)\in I^{2d}} |R(t,s;\theta)-R(t,s;\theta^*)| \to 0, \quad \text{if} \quad \theta\to\theta^*. \tag{2.10}$$

It is well known from functional analysis, that the kernel function $R(t,s;\theta)$ has a sequence of eigenvalues $\{\lambda_i(\theta),\ i\geqq 1\}$ and eigenfunctions $\{\varphi_i(t;\theta),\ i\geqq 1\}$, that is

$$\lambda_i(\theta)\varphi_i(t;\theta) = \int_{I^d} R(s,t;\theta)\varphi_i(s;\theta)\,ds,$$

$$\int_{I^d} \varphi_i(t;\theta)\varphi_j(t;\theta)\,dt = \delta_{ij}, \quad \lambda_i(\theta)>0.$$

The sequence of eigenvalues and eigenfunctions determines a CONS in $H(R(\theta))$:

$$e_i(t;\theta) = (\lambda_i(\theta))^{-1/2}\varphi_i(t;\theta). \tag{2.11}$$

It follows from (2.11) that in this case

$$\nu_i(ds) = (\lambda_i(\theta))^{-3/2}\varphi_i(s;\theta)\,ds, \tag{2.12}$$

so we have

$$\|v_i(\theta)\| \leqq (\lambda_i(\theta))^{-3/2} M(\theta), \tag{2.13}$$

where $M(\theta)=\sup_{(t,s)\in I^{2d}} |R(t, s; \theta)|$.

Let N be given. Then the polynomials K_1, K_2 and K_3 of Lemma 2.1 which depend only on N, $M(\theta)$ and $\|v_i(\theta)\|$, $1\leqq i\leqq N$, are continuous positive functions of θ. Let $0<\varepsilon<2 \inf_{\theta\in\Theta} \{1/(K_1 \min (1/K_2, 1/K_3))\}$ and define $\delta(\theta)>0$ for every $\theta\in\Theta$ in the following way: if $|\theta^*-\theta|<\delta(\theta)$, then the inequality

$$\sup_{(t,s)\in I^{2d}} |R(t, s; \theta)-R(t, s; \theta^*)| \leqq \frac{\varepsilon}{2} \operatorname{mix}\left(\frac{1}{K_2}, \frac{1}{K_3}\right) \leqq \frac{1}{K_1} \tag{2.14}$$

holds. The existence of $\delta(\theta)$ follows from (2.10). Let $A(\varepsilon, \theta)$ denote the open ball with centre θ and radius $\delta(\theta)$. Then the union $\bigcup_{\theta\in\Theta} A(\varepsilon, \theta)$ covers the set Θ. Using the compactness of Θ, we have that a finite sequence $A_1(\varepsilon, \theta_1), \ldots, A_l(\varepsilon, \theta_l)$ also covers Θ. If $\theta\in A_i(\varepsilon, \theta_i)$, $1\leqq i\leqq l$, then we can define with Mangano's method an orthogonal set $\{f_{j,i}(\theta), 1\leqq j\leqq N\}$ in $H(R(\theta))$ generated by the measures $\{\varkappa_{j,i}(\theta), 1\leqq j\leqq N\}$. As we said in the proof of Lemma 2.1, these functions and measures are linear combinations of $\{e_{j,i}, 1\leqq j\leqq N\}$ and $\{v_{j,i}, 1\leqq j\leqq N\}$, where $\{e_{j,i}, 1\leqq j\leqq N\}$ is an orthonormal set in $H(R(\theta_i))$. If the measures $\{v_{j,i}, 1\leqq j\leqq N\}$ are generated by the eigenfunctions $\{\varphi_{j,i}, 1\leqq j\leqq N\}$ of $R(t, s; \theta_i)$ then $\varkappa_{j,i}$ can be written in the form

$$\varkappa_{j,i}(dt) = c_{j,i}\varphi_{j,i}(t)\, dt, \tag{2.15}$$

where the $c_{j,i}$, $1\leqq i\leqq l$, $1\leqq j\leqq N$, are constants. It follows from the definition of A_i and from Lemma 2.1, that if $\theta, \theta^*\in A_i$, then

$$\sup_{t\in I^d} |f_{j,i}(t; \theta)-f_{j,i}(t; \theta^*)| < \varepsilon, \quad 1 \leqq j \leqq N, \tag{2.16}$$

$$\|\varkappa_{j,i}(\theta)-\varkappa_{j,i}(\theta^*)\| < \varepsilon, \quad 1 \leqq j \leqq N. \tag{2.17}$$

The covariance function of the limiting process $\{Y(t; \theta_0), t\in I^d\}$ depends on an unknown parameter θ_0, which will be estimated with a sequence of random variables θ_n, such that

$$|\theta_n-\theta_0| \xrightarrow{P} 0 \quad (n\to\infty). \tag{2.18}$$

Let $\varepsilon=2^{-m}$ and define the following random functions and measures: if $\theta_n\in A_i(2^{-m}; \theta_i)$, then $\{f^n_{j,i}, 1\leqq j\leqq N\}$ denotes the corresponding orthonormal sequence and $\{\varkappa^n_{j,i}, 1\leqq j\leqq N\}$ denotes the measures corresponding to $\{v_{j,i}(\theta_i), 1\leqq j\leqq N\}$.

Jain and Kallianpur [8] proved that if $\{Y(t), t\in I^d\}$ is a Gaussian process with continuous sample paths a.s., having mean zero and continuous covariance $R(t, s)$, then the sums

$$\sum_{j=1}^{N} \xi_j \varphi_j(t) \tag{2.19}$$

converge uniformly in $t \in I^d$ to $Y(t)$ a.s. as $N \to \infty$, where $\{\varphi_j, j \geqq 1\}$ is a CONS in $H(R)$ and $\{\xi_j, j \geqq 1\}$ is a suitable sequence of independent standard normal random variables. On the other hand, in case $d=1$; if G is a covariance such that there exists a Gaussian process with this covariance and with almost all sample paths continuous, $\{\xi_j^*, j \geqq 1\}$ is a sequence of independent standard normal variables, and if $\{\psi_j, j \geqq 1\}$ is a CONS in $H(G)$, then the sums

$$\sum_{j=1}^{N} \xi_j^* \psi_j(t) \tag{2.20}$$

converge uniformly a.s. in $t \in [0, 1]$ to a Gaussian process whose covariance is G and almost all of whose sample paths are continuous, as $N \to \infty$ (Theorem 2 in [8]). If $d > 1$, then some further conditions on G are needed to retain this statement.

We will assume in this section, that Z is a Gaussian process with almost all sample paths continuous,

$$EZ(t) = 0, \quad EZ(t)Z(s) = G(t, s), \quad t, s \in I^q, \; q \geqq 1,$$

and for every sequence of standard normal random variables $\{\xi_j^*, j \geqq 1\}$ there is a centered Gaussian process Z^* having continuous sample paths a.s. and covariance G such that

$$\sup_{t \in I^q} \Big| \sum_{j=1}^{N} \xi_j^* \psi_j(t) - Z^*(t) \Big| \xrightarrow{P} 0 \quad (N \to \infty), \tag{2.21}$$

where $\{\psi_j(t), j \geqq 1\}$ is a CONS in $H(G)$.

Let $X_n^*(s)$ denote the empirical kernel transform of $X_n(t)$

$$X_n^*(s) = \sum_{j=1}^{N} \int_{I^d} X_n(t) \psi_j(s) \varkappa_{j,i}^n(dt), \quad \text{if} \quad \theta_n \in A_i(2^{-m}, \theta_i) \; s \in I^q. \tag{2.22}$$

If the sequence $\{v_{j,i}, j \geqq 1\}$ is generated the eigenfunctions $\{\varphi_{j,i}, j \geqq 1\}$ of $R(\cdot, \theta_i)$, then the transform can be written in the form

$$X_n^*(s) = \int_{I^d} X_n(t) k_{N,m}^n(t, s)\, dt, \quad s \in I^q, \tag{2.23}$$

where

$$k_{N,m}^n = \sum_{j=1}^{N} c_{j,i}^n \varphi_{j,i}(t) \psi_j(s), \quad \text{if} \quad \theta_n \in A_i(2^{-m}, \theta_i), \tag{2.24}$$

is a random kernel function.

Theorem 2.2. *If the underlying probability space is rich enough, then we can define a sequence $\{Z_n(s)\}$ of copies of $Z(s)$ such that we have*

$$\sup_{s\in I^q} |X_n^*(s)-Z_n(s)| \xrightarrow{P} 0,$$

if N, $m(N)$ and $n(N, m)$ go to infinity.

Proof. Let ε and δ be arbitrary positive constants. The distribution of $\sup_{t\in I^d} |Y_n(t)|$ is independent of n, so there is a constant M_1 such that we have

$$P\{\sup_{t\in I^d} |Y_n(t)| > M_1\} \leqq \delta/8. \tag{2.25}$$

Using condition (2.21) we have that

$$P\Big\{\sup_{s\in I^q} \Big| \sum_{j=N+1}^{\infty} \xi_j \psi_j(s)\Big| > \varepsilon/3\Big\} < \delta/4, \tag{2.26}$$

if $N \geqq N_0$ for every sequence of independent standard normal random variables. Set

$$M_2 = \max_{1\leqq j\leqq N} \sup_{s\in I^q} |\psi_j(s)|.$$

Let $m=m(N)$ be so large that

$$2^{-m} < \varepsilon/(3NM_1M_2). \tag{2.27}$$

The sequence θ_n goes to θ_0 in probability, therefore there is a paramater subset $A_i(2^{-m}, \theta_i)$ such that

$$P\{\theta_n \in A_i(2^{-m}, \theta_i)\} > 1-\delta/2, \tag{2.28}$$

if $n \geqq n_1(N, m)$.

The transformed process X_n^* can be decomposed as

$$X_n^*(s) = \int_{I^d} (X_n(t)-Y_n(t)) \sum_{j=1}^{N} \psi_j(s) \varkappa_{j,i}^n(dt) +$$

$$+ \int_{I^d} Y_n(t) \sum_{j=1}^{N} \psi_j(s)\big(\varkappa_{j,i}^n(dt) - \varkappa_{j,i}(\theta_0)(dt)\big) + \int_{I^d} Y_n(t) \sum_{j=1}^{N} \psi_j(s) \varkappa_{j,i}(\theta_0)(dt) =$$

$$= a_{1n}(s) + a_{2n}(s) + a_{3n}(s),$$

say. We assume that $\theta_0 \in A_i(2^{-m}, \theta_i)$. Using (2.9) we have that

$$P\{\sup_{s\in I^q} |a_{1n}(s)| > \varepsilon/3,\ \theta_n \in A_i(2^{-m}, \theta_i)\} \leqq$$

$$\leqq P\{\sup_{t\in I^d} |X_n(t)-Y_n(t)| NM_2(\max_{1\leqq j\leqq N} \|v_{j,i}\| + 2^{-m}) > \varepsilon/3,\ \theta_n \in A_i(2^{-m}, \theta_i)\} \leqq \delta/8,$$

if $n \geqq n_0 = \max\big(n_1(N, m), n_2(N, m)\big)$. The second term also goes to zero in probability, because it follows from (2.25), and (2.27) that

$$P\{\sup_{s\in I^q}|a_{2n}(s)|>\varepsilon/3,\ \theta_n\in A_i(2^{-m},\theta_i)\}\leqq$$
$$\leqq P\{M_1M_2N\max_{1\leqq j\leqq N}(\|\varkappa^n_{j,i}-\varkappa_{j,i}(\theta_0)\|)>\varepsilon/3\}=0.$$

The orthonormal set $\{f_{j,i}(\theta_0),\ 1\leqq j\leqq N\}$ corresponding to the measures $\{\varkappa_{j,i}(\theta_0),\ 1\leqq j\leqq N\}$ can be completed to a CONS $\{f_{j,i}(\theta_0),\ j\geqq 1\}$ in $H(R(\theta_0))$. The sequence $\sum_{j=1}^{M}\xi^n_{j,i}f_{j,i}(t)$ converges uniformly in $t\in I^d$ to $Y_n(t)$, as $M\to\infty$, a.s. with a suitably chosen sequence of independent standard normal variables $\{\xi^n_{j,i},\ j\geqq 1\}$ (see Theorem 1 of JAIN and KALLIANPUR [8]). So by (2.5), $a_{3n}(s)$ can be decomposed as a finite sum

$$a_{3n}(s)=\sum_{j=1}^{N}\xi^n_{j,i}\psi_j(s).$$

Using the condition (2.21), the partial sums

$$\sum_{j=1}^{N}\xi^n_{j,i}\psi_j(s)$$

converge (as $N\to\infty$) uniformly in $s\in I^q$ to a separable Gaussian process denoted by $Z_n(s)$. On the other hand, we have that

$$\Big\{\Big(Z_n(s),\ \sum_{j=1}^{N}\xi^n_{j,i}\psi_j(s)\Big),\ s\in I^q\Big\}\overset{\mathscr{D}}{=}\Big\{\Big(Z(s),\ \sum_{j=1}^{N}\xi_j\psi_j(s)\Big),\ s\in I^q\Big\},$$

where

$$Z(s)=\sum_{j=1}^{\infty}\xi_j\psi_j(s)\quad\text{a.s.},$$

and $\overset{\mathscr{D}}{=}$ denotes equality in distribution. So it follows form (2.26) that

$$P\{\sup_{s\in I^q}|a_{3n}(s)-Z_n(s)|>\varepsilon/3\}\leqq\delta/4.$$

Summing up, we proved that if $N\geqq N_0(\varepsilon,\delta)$, $m\geqq m_0(N,\varepsilon,\delta)$ and $n\geqq n_0(N,m,\varepsilon,\delta)$, then

$$P\{\sup_{s\in I^q}|X^*_n(s)-Z_n(s)|>\varepsilon\}\leqq$$
$$\leqq P\{\sup_{s\in I^q}|X^*_n(s)-Z_n(s)|>\varepsilon,\ \theta_n\in A_i(2^{-m},\theta_i)\}+P\{\theta_n\notin A_i(2^{-m},\theta_i)\}<\delta,$$

which is the desired conclusion.

3. Applications. So far $q\geqq 1$ was arbitrary, and from now on we choose $q=1$ since univariate limit processes are more convenient to handle. First we study the estimated empirical process when $d=1$. Let $F^{-1}(t;\theta)$ denote the inverse function to $F(t;\theta)$,

$$F^{-1}(t;\theta)=\inf\{s:\ F(s;\theta)\geqq t\}.$$

It follows from Theorem A, that

$$\sup_{0 \leqq t \leqq 1} |\beta_n(F^{-1}(t;\, \theta_n)) - D_n(F^{-1}(t;\, \theta_0);\, \theta_0)| \xrightarrow{P} 0 \quad (n \to \infty)$$

and

$$ED_n(F^{-1}(t;\, \theta_0);\, \theta_0) = 0,$$

$$ED_n(F^{-1}(t;\, \theta_0);\, \theta_0) D_n(F^{-1}(s;\, \theta_0);\, \theta_0) = R_1(t, s;\, \theta_0) = t \wedge s - ts -$$
$$-J(F^{-1}(t;\, \theta_0);\, \theta_0) \nabla_\theta^T F(F^{-1}(s;\, \theta_0);\, \theta_0) - J(F^{-1}(s;\, \theta_0);\, \theta_0) \nabla_\theta^T F(F^{-1}(t;\, \theta_0);\, \theta_0) +$$
$$+ \nabla_\theta F(F^{-1}(t;\, \theta_0);\, \theta_0) M(\theta_0) \nabla_\theta^T F(F^{-1}(s;\, \theta_0);\, \theta_0),$$

where

$$J(t;\, \theta_0) = \int_{-\infty}^{t} l(u;\, \theta_0)\, dF(u;\, \theta_0).$$

The processes $D_n(F^{-1}(t;\theta_0);\theta_0)$ have continuous sample functions a.s. if $\nabla_\theta F(F^{-1}(t;\theta_0);\theta_0)$ and $J(F^{-1}(t;\theta_0))$ are continuous functions of t. The covariance function $R_1(t, s;\, \theta^*)$ will be continuous in θ^* if $M(\theta^*)$, $J(F^{-1}(t;\theta^*);\theta^*)$ and $\nabla_\theta F(F^{-1}(t;\theta^*);\theta^*)$ are continuous functions of $\theta^* \in \Theta$. The random function $R_1(t, s;\, \theta_n)$ is an estimate of the covariance function of the limit process. So we can define β_n^*, the empirical transform of $\beta_n(F^{-1}(t;\, \theta_n))$ as it was defined by (2.22) or (2.23). The sample $X_1, \ldots, X_n$ from a distribution belonging to the parametric family $\mathcal{F}$ determine only the random measures (and functions) in the definition of the empirical transform, so we can choose the eigenfunctions $\{\psi_j, j \geqq 1\}$ of the limit process without restriction. For example, if

$$\psi_k(s) = (\sqrt{2}/k\pi) \sin k\pi s, \quad 0 \leqq s \leqq 1, \tag{3.1}$$

then the limit process will be the Brownian bridge. If

$$\psi_1(s) = s, \tag{3.2}$$

$$\psi_{k+1}(s) = (\sqrt{2}/k\pi) \sin k\pi s, \quad 0 \leqq s \leqq 1 \tag{3.3}$$

then $\beta_n^*(s)$, $0 \leqq s \leqq 1$, will converge weakly to the Wiener process.

Theorem 3.1. *We suppose, that the conditions* (i), (ii), (v) *of Theorem A are satisfied and*

(iii)* $M(\theta^*) = El^{\mathrm{T}}(X^*;\, \theta^*) l(X^*;\, \theta^*)$ *is a finite, nonnegative definite matrix and* $M(\theta^*)$ *is continuous in* $\theta^* \in \Theta$, *where* $P(X^* < t) = F(t;\, \theta^*)$,

(iv)* $J(F^{-1}(t;\theta^*);\theta^*)$ *and* $\nabla_\theta F(F^{-1}(t;\theta^*);\theta^*)$ *are uniformly continuous in* t, $0 \leqq t \leqq 1$, *and* $\theta^* \in \Theta$, *where* $\Theta \subset R^p$ *is a compact parameter set and the true value* θ_0 *is an interior point of* Θ.

Then we can define a sequence $\{Z_n(s)\}$ *of copies of* $Z(s)$ *on the probability space of Theorem 2.2 such that we have*

$$\sup_{0 \leqq s \leqq 1} |\beta_n^*(s) - Z_n(s)| \xrightarrow{P} 0,$$

if N, m, n *go to infinity.*

Proof. It follows from the conditions of the theorem that $R_1(t, s; \theta)$, $0 \leqq t, s \leqq 1$, $\theta \in \Theta$, satisfies (2.10) and $|\theta_n - \theta_0| \xrightarrow{P} 0$, as $n \to \infty$. The processes $D_n(F^{-1}(t; \theta); \theta)$ have continuous sample path functions a.s., so this theorem is a consequence of Theorem 2.2.

The most important special case of this theorem is when we estimate shift and location parameters only, i.e., the parametric family can be written in the form

$$\mathscr{F}_s = \left\{F\left(\frac{t-\theta_0^1}{\theta_0^2}\right), \ -\infty < \theta_0^1 < \infty, \theta_0^2 > 0, \ t \in R^1\right\}.$$

The covariance function of the limit process for the shift and location estimated empirical process was computed by DARLING [3] and DURBIN [4] (cf. [5]). They proved that the covariance function of $D_n(F^{-1}(t; \theta_0^1, \theta_0^2); \theta_0^1, \theta_0^2)$ does not depend on (θ_0^1, θ_0^2):

$$ED_n(F^{-1}(t; \theta_0^1, \theta_0^2); \theta_0^1, \theta_0^2) D_n(F^{-1}(s; \theta_0^1, \theta_0^2); \theta_0^1, \theta_0^2) = R_2(t, s) =$$
$$= t \wedge s - ts - [I_{11} I_{22} - I_{12}^2]^{-1} [I_{22} w_1(t) w_1(s) +$$
$$+ I_{11} w_2(t) w_2(s) - I_{12}(w_1(t) w_2(s) + w_2(t) w_1(s))],$$

where

$$w_1(t) = f(F^{-1}(t)), \quad w_2(t) = F^{-1}(t) f(F^{-1}(t)),$$

$$I_{11} = \int_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)}\right]^2 f(x) dx, \quad I_{12} = - \int_{-\infty}^{\infty} x \left[\frac{f'(x)}{f(x)}\right]^2 f(x) dx,$$

$$I_{22} = \int_{-\infty}^{\infty} x^2 \left[\frac{f'(x)}{f(x)}\right]^2 f(x) dx - 1,$$

and f, f' are the first and second derivatives of F assumed to exists. In this special case we do not have to estimate R_2 from the sample, so the transformation of $\beta_n(F^{-1}(t; \theta_{n1}, \theta_{n2}))$ will be non-random. If $\{\varphi_j^*, j \geqq 1\}$ is a CONS in $H(R_2)$ generated by the measures $\{v_j^*, j \geqq 1\}$ then the transformation of $\beta_n(F^{-1}(t; \theta_n^1, \theta_n^2))$

$$\beta_n^*(s) = \int_0^1 \beta_n(F^{-1}(t; \theta_{n1}, \theta_{n2})) \sum_{j=1}^N \psi_j(s) v_j^*(dt)$$

is also non-random, and

$$\sup_{0 \leqq s \leqq 1} |\beta_n^*(s) - Z_n(s)| \xrightarrow{P} 0 \quad (n \to \infty),$$

where $\{Z_n\}$ is a sequence of copies of Z.

Finally we study the general case, when d is an arbitrary positive integer. The transformation of the parameter estimated empirical process into the unit interval was very simple in the one dimensional case, but in the general case it is a bit more complicated. Let $F_j(x_j; \theta)$ denote the j^{th} marginal distribution of $F(x; \theta)$, $x = (x_1, \ldots, x_d)$. There is a d-variate distribution function $H(x; \theta)$, all the univariate marginals of which being uniformly distributed on [0, 1], such that

$$F(x;\theta) = H\big(F_1(x_1;\theta), \dots, F_d(x_d;\theta);\theta\big), \quad x = (x_1, \dots, x_d).$$

Let F_j^{-1} denote the left-continuous inverse of F_j and define the following function

$$\bar{F}(t;\theta) = H\big(F_1^{-1}(t_1;\theta), \dots, F_d^{-1}(t_d;\theta);\theta\big), \quad t = (t_1, \dots, t_d).$$

So the process $\beta_n(\bar{F}(t;\theta_n))$ is defined on I^d, and it follows from Theorem A, that

$$\sup_{t\in I^d} \big|\beta_n(\bar{F}(t;\theta_n)) - D_n(\bar{F}(t;\theta_0);\theta_0)\big| \xrightarrow{P} 0 \quad (n\to\infty).$$

The covariance function of $D_n(\bar{F}(t;\theta_0);\theta_0))$ can be computed from the representation of the Durbin process. We have, for $t=(t_1, \dots, t_d)$, $s=(s_1, \dots, s_d)$, that

$$ED_n\big(\bar{F}(t;\theta_0);\theta_0\big) = 0,$$

$$ED_n\big(\bar{F}(t;\theta_0);\theta_0\big)\, D_n\big(\bar{F}(s;\theta_0);\theta_0\big) = R_3(t,s;\theta_0) =$$
$$= F\big(\bar{F}(t\wedge s;\theta_0);\theta_0\big) - F\big(\bar{F}(t;\theta_0);\theta_0\big)F\big(\bar{F}(s;\theta_0);\theta_0\big) -$$
$$- J\big(\bar{F}(t;\theta_0);\theta_0\big)\nabla_\theta^T F\big(\bar{F}(s;\theta_0);\theta_0\big) - J\big(\bar{F}(s;\theta_0);\theta_0\big)\nabla_\theta^T F\big(\bar{F}(t;\theta_0);\theta_0\big) +$$
$$+ \nabla_\theta F\big(\bar{F}(t;\theta_0);\theta_0\big)M(\theta_0)\nabla_\theta^T F\big(\bar{F}(s;\theta_0);\theta_0\big).$$

The following theorem is a generalization of Theorem 3.1 for arbitrary d. Let β_n^* denote the empirical kernel-transformed empirical process defined by (2.22) or (2.23).

Theorem 3.2. *We suppose, that the conditions* (i), (ii), (v) *of Theorem A are satisfied and*

(iii)* $M(\theta^*) = El^T(X^*;\theta^*)l(X^*;\theta^*)$ *is a finite, nonnegative definite matrix and* $M(\theta^*)$ *is continuous in* $\theta^*\in\Theta$, *where* $P\{X^{*1}<t_1, \dots, X^{*d}<t_d\} = F(t;\theta^*)$, $t=(t_1, \dots, t_d)$, $X^* = (X^{*1}, \dots, X^{*d})$,

(iv)* $J\big(\bar{F}(t;\theta^*);\theta^*\big)$ *and* $\nabla_\theta F\big(\bar{F}(t;\theta^*);\theta^*\big)$ *are uniformly continuous in* $t\in I^d$ *and* $\theta^*\in\Theta$, *where* $\Theta\subset R^p$ *is a compact parameter set, and* θ_0 *is an interior point of* Θ.

Then we can define a sequence $\{Z_n(s), 0\leqq s\leqq 1\}$ *of copies of* $Z(s)$ *on the probability space of Theorem 2.2 such that we have*

$$\sup_{0\leqq s\leqq 1} |\beta_n^*(s) - Z_n(s)| \xrightarrow{P} 0,$$

if N, m, n *go to infinity.*

The proof of this theorem is very similar to the proof of Theorem 3.1, therefore it is omitted. These conditions are stronger than the conditions of Theorem A in order to guarantee the applicability of Theorem 3.1.

We proved only the existence of the empirical kernel transform with nice limiting properties, but so far we said nothing about the decomposition $A_i(\varepsilon;\theta_i)$, $i=1, \dots, l$, of the compact parameter set Θ and hence about the concrete choice of the $\varphi_{j,i}$ functions and the quantities $c_{j,i}$ in (2.24). We noticed in Section 2, that the eigenvalues and eigenfunctions of $R(t,s;\theta)$ determine a CONS in $H(R(\theta))$. Let

$\lambda_1(\theta), \ldots, \lambda_N(\theta)$ denote the first N largest eigenvalues of $R(\theta)$. We can choose a Θ^* neighbourhood of θ_0, such that there is a positive lower bound of $\lambda_j(\theta)$, $1 \leqq j \leqq N$, $\theta \in \Theta^*$. It follows from (2.12) and from the continuity of K_i, $i=1, 2, 3$, that we have

$$K_i(N, M(\theta), \|v_1(\theta)\|, \ldots, \|v_N(\theta)\|) \leqq L_i, \quad i = 1, 2, 3,$$

if $\theta \in \Theta^*$. Using (2.13) we see that every $\theta \in \Theta^*$ can be the centre of the balls $A_i(\varepsilon; \theta)$ and the radii of $A_i(\varepsilon; \theta)$ does not already depend on θ. Therefore an arbitrary devision of Θ^* will suffice if this devision is fine enough (for example, the common radius of the balls is small enough). But because $\theta_n \xrightarrow{P} \theta_0$, in practice we may assume that $\theta_n \in \Theta^*$, and hence the devision of $\Theta \setminus \Theta^*$ is completely arbitrary. This choice of the $A(\varepsilon; \theta_i)$ will be suitable for us, if we use the first N largest eigenvalues and the corresponding eigenfunctions of $R(\theta)$ to make the empirical kernel transformation.

Acknowledgement. I am grateful to Professor Sándor Csörgő for reading through the first draft to this note and for making a number of useful suggestions.

References

[1] M. D. BURKE, M. CSÖRGŐ, S. CSÖRGŐ and P. RÉVÉSZ, Approximations of the empirical processes when parameters are estimated. *Ann. Probab.*, **7** (1978), 790—810.

[2] S. CSÖRGŐ, Kernel-transformed empirical processes, *J. Multivar. Analysis*, **13** (1983), 511—533.

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

[4] J. DURBIN, *Distribution theory for tests based on the sample distribution function*, Regional conference series in applied math., No. 9, SIAM (Philadelphia, 1973).

[5] K. O. DZAPARIDZE and M. S. NIKULIN, The probability distributions of the Kolmogorov and ω^2 statistics for continuous distributions with shift and location parameters, *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, **85** (1976), 46—74. (Russian)

[6] E. V. HMALADZE, The use of ω^2 for testing parametric hypotheses, *Teorija Verojatn. Primen.*, **24** (1979), 280—297. (Russian)

[7] E. V. HMALADZE, Martingale approach in the theory of goodness-of-fit tests, *Teorija Verojatn. Primen.*, **26** (1981), 246—265. (Russian)

[8] N. C. JAIN and G. A. KALLIANPUR, A note on uniform convergence of stochastic processes, *Ann. Math. Statist.*, **41** (1970), 1360—1362.

[9] G. G. MANGANO, Sequential compactness of certain sequences of Gaussian random variables with values in $C[0, 1]$, *Ann. Probab.*, **4** (1976), 902—913.

[10] G. NEUHAUS, Asymptotic properties of the Cramér—von Mises statistics when parameters are estimated, in: *Proc. Prague Symp. Asymptotic Statist.* (3—6 Sept., 1973, ed. J. Hájek), Charles University (Prague, 1974); 257—297.

[11] G. NEUHAUS, Weak convergence under continuous alternatives of the empirical processes when parameters are estimated: the D_k approach, in: *Empirical Distributions and Processes*, Lect. Notes in Math. 566, Springer-Verlag (Berlin, 1976); 68—82.

[12] H. SATO, Gaussian measure on a Banach space and abstract Wiener measure, *Nagoya Math. J.*, **36** (1969), 65—81.

BOLYAI INSTITUTE
ARADI VÉRTANÚK TERE 1
6720 SZEGED, HUNGARY