ON THE EMPIRICAL PROCESS WHEN PARAMETERS ARE ESTIMATED

M. Csörgő, J. Komlós, P. Major, P. Révész, G. Tusnády

Ottawa, Budapest, Budapest, Budapest, Budapest

INTRODUCTION

Let $X_1, X_2, \dots$ be a sequence of i.i.d.r.v.'s with d.f. $P(X_i < x) = F(x)$ $(i = 1, 2, \dots; -\infty < x < +\infty)$ where $F(x)$ is continuous; further let $F_n(x)$ be the empirical d.f. based on the sample $X_1, X_2, \dots, X_n$ and finally let $\alpha_n(x) = \sqrt{(n)}\,(F_n(x) - F(x))$.

Several authors have proved that the process $\alpha_n(x)$ is near to a Gaussian Process (G.P.) (in some sense). The two most natural G.P.'s from this point of view are the following:

(i) Brownian Bridge (B.B.) $B(x)$ $(0 \leq x \leq 1)$ with covariance function $E\, B(x_1) \,.\, B(x_2) = \min(x_1, x_2) - x_1 \,.\, x_2$,

(ii) Kiefer Process (K.P.) $K(x, y)$ $(0 \leq x \leq 1;\ 0 \leq y < \infty)$ with covariance function $E\, K(x_1, y_1)\, K(x_2, y_2) = \min(y_1, y_2)\,[\min(x_1, x_2) - x_1 x_2]$.

The two strongest results stating that $\alpha_n(x)$ can be approximated by a B.B. resp. by a K.P. are the following:

Theorem A [1]. *If the underlying probability space is rich enough then one can define a sequence $\{B_n(x)\}$ of B.B's such that:*

$$\alpha_n(x) - B_n(F(x)) = \varepsilon_n(x)$$

where $\{\varepsilon_n(x)\}$ is a sequence of stochastic processes for which

$$\sup_x |\varepsilon_n(x)| = O\left(\frac{\log n}{\sqrt{n}}\right)$$

with probability 1.

THEOREM B [1]. *If the underlying probability space is rich enough then one can define a K.P. $K(x, y)$ such that:*

$$\sqrt{(n)}\, \alpha_n(x) - K(F(x), n) = \delta_n(x)$$

where $\{\delta_n(x)\}$ is a sequence of stochastic processes for which

$$\sup_x \delta_n(x) = O(\log^2 n)$$

*with probability 1 and**

$$E(\sup_x |\delta_n(x)|)^2 = O(\log^4 n)\,.$$

These results are extremely useful for goodness of fit statistical test, when $F(.)$ is supposed to be completely specified. In most cases, however, only the form of $F(.)$ is assumed, as the so-called null hypothesis H_0 while the possible parameters of $F(.)$ are not specified by it; i.e. we usually have a composite goodness of fit problem. This case was also investigated in many different papers. See for example [2] and [3] of Durbin, where a detailed reference list can also be found. In this paper we follow Durbin in many sense. However, we modify the process investigated by him by a scale transformation and in this way we get strong convergence instead of his weak convergence in a direct way.

First, we assume that $F(x; \theta)$ is a one-parameter family, θ belonging to an open interval $\mathscr{I}$ (possibly infinite) of the real line R^1 and we consider the maximum likelihood estimator (m.l.e.) $\hat{\theta}_n$ of θ_0 the true value of θ. Our main result states that the process

$$\beta_n(x) = n(F_n(x) - F(x; \hat{\theta}_n))$$

can be approximated by the following G.P.

$$G(x, n; \theta_0) = G(x, n) = K(F(x; \theta_0), n) - \frac{h(x; \theta_0)}{I(\theta_0)} \int \frac{\partial}{\partial\theta} \log f(x; \theta_0)\, d_x K(F(x; \theta_0), n)$$

(provided that $F(x; \theta)$ satisfies some natural conditions) where $K(.,.)$ is a K.P.,

$$h(x; \theta) = \frac{\partial F(x; \theta)}{\partial\theta}, \quad f(x; \theta) = \frac{\partial F(x; \theta)}{\partial x},$$

$$I(\theta) = \int \frac{\left(\frac{\partial f(x; \theta)}{\partial\theta}\right)^2}{f(x; \theta)}\, dx\,,$$

* This second result is not formulated explicitly in [1] but it follows easily from Theorem 4 of [1].

the Fisher information number. It is easy to check that $G(x, n)$ is a G.P. and its covariance function is

$$EG(x_1, n_1)\, G(x_2, n_2) =$$
$$= \min(n_1, n_2)\left[\min(F(x_1; \theta_0)\, F(x_2; \theta_0)) - F(x_1; \theta_0)\, F(x_2, \theta_0) - h(x_1; \theta_0)\, h(x_2; \theta_0\right].$$

So, in general the distribution of G and also that of $1/\sqrt{(n)}\, \sup_x G(x; n)$, depends on F and θ.

This means that the above result cannot be applied to test the composite hypothesis $H_0 : F = F(x; \theta)\ \theta \in \mathscr{I}$.

To avoid this problem, Durbin proposed that instead of $\hat{\theta}_n$ one should use $\tilde{\theta}_n$ to estimate θ_0, where $\tilde{\theta}_n$ is the m.l.e. of θ_0 based only on any randomly picked half of the sample $X_1, X_2, \ldots, X_n$. He remarked, [3], that the process $\sqrt{(n)}\,(F_n(x) - F(x; \tilde{\theta}_n))$ converges weakly to the usual B.B. with a completely specified F.

This result will be also reproduced via our strong convergence method.

Another, perhaps even more useful idea to avoid the problem of θ still appearing in the limiting process is the following: we estimate the process $G(x, n)$ by

$$\hat{G}(x, n) = K(F(x; \hat{\theta}_n), n) - \frac{h(x; \hat{\theta}_n)}{I(\hat{\theta}_n)} \int \frac{\partial}{\partial \theta} \log f(x; \hat{\theta}_n)\, d_x K(F(x; \hat{\theta}_n), n)$$

namely it can be seen that

$$\sup_x |\hat{G}(x, n) - G(x, n)| = O(n^{\varepsilon})$$

with probability 1, for some $0 < \varepsilon < \frac{1}{2}$. This means that $\hat{G}(x, n)$ is just as good an approximation of $\beta_n(x)$ as $G(x, n)$ was.

Now let $M(\theta, \alpha)$ $(0 < \alpha < 1;\ \theta \in \mathscr{I})$ be the number for which

$$P\left\{\frac{1}{\sqrt{n}} \sup_x |G(x, n; \theta)| > M(\theta, \alpha)\right\} = \alpha\,.$$

It is easy to check that $M(\theta, \alpha)$ is uniquely determined in this way and it is continuous in θ and α, this implies:

$$\alpha = \lim_{n \to \infty} P\left\{\frac{1}{\sqrt{n}} \sup_x |G(x, n; \hat{\theta}_n)| > M(\hat{\theta}_n, \alpha)\right\} =$$
$$= \lim_{n \to \infty} P\{\sqrt{(n)} \sup_x |F_n(x) - F(x; \hat{\theta}_n)| > M(\hat{\theta}_n, \alpha)\}\,.$$

Then one can propose the following test of level α: reject the composite hypothesis $H_0 : F = F(x; \theta),\ \theta \in \mathscr{I}$ if $\hat{\theta}_n = \theta_1$ and $1/\sqrt{(n)} \sup_x |F_n(x) - F(x; \hat{\theta}_n)| > M(\theta_1, \alpha)$.

1. APPROXIMATION OF $\beta_n(x)$ BY $G(x, n)$

From now on the following conditions will be assumed:

C.1. $F(x; \theta)$ is absolutely continuous for all $\theta \in \mathscr{I}$ and the density function $f(x; \theta) = \partial/\partial x\, F(x; \theta)$ is continuous on $R^1 \times \mathscr{I}$,

C.2.

$$\int |f(x; \theta_1) - f(x; \theta_2)|\, dx > 0 \quad \text{if} \quad \theta_1 \neq \theta_2 (\theta_1 \in \mathscr{I},\ \theta_2 \in \mathscr{I}),$$

C.3.

$$0 < I(\theta) < \infty,$$

and $I(\theta)$ is continuous on $\mathscr{I}$,

C.4. there exists a $p \geqq 0$ such that $\sup_{\theta \in \mathscr{I}} (1 + |\theta|)^{-p} I(\theta) < \infty$,

C.5. for all $\delta > 0$ and $\theta \in \mathscr{I}$ we have

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\theta - \varepsilon}^{\theta + \varepsilon} \int_{\mathfrak{A}} \frac{\left(\frac{\partial}{\partial \theta} f(x; \theta) \right)^2}{f(x; \theta)}\, dx\, d\theta$$

where

$$\mathfrak{A} = \left\{ x : \left| \log \frac{f(x; \theta + \varepsilon)}{f(x; \theta)} \right| > \delta \right\},$$

C.6. there exists a $\delta > 0$ such that

$$\sup_{\theta} |\theta - \theta_0|^{\delta} \int \sqrt{[f(x; \theta) f(x; \theta_0)]}\, dx < \infty,$$

C.7.

$$\frac{\partial}{\partial \theta} \int f(x; \theta)\, dx = \int \frac{\partial}{\partial \theta} f(x; \theta)\, dx = \frac{\partial^2}{\partial \theta^2} \int f(x; \theta)\, dx = \int \frac{\partial^2}{\partial \theta^2} f(x; \theta)\, dx = 0,$$

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{t} f(x; \theta)\, dx = \int_{-\infty}^{t} \frac{\partial}{\partial \theta} f(x; \theta)\, dx$$

(the existence of the mentioned derivatives is assumed),

C.8. there exists a $\delta > 0$ and a $K > 0$ such that

$$\int \left| \frac{\partial^2}{\partial \theta^2} \log f(x; \theta_0) \right|^{1 + \delta} f(x; \theta_0)\, dx < K,$$

$$\int \left| \frac{\partial}{\partial \theta} \log f(x; \theta_0) \right|^{2 + \delta} f(x; \theta_0)\, dx < K,$$

C.9. there exists a function $k(x)$ and a $\gamma > 0$ such that

$$\left|\frac{\partial^2}{\partial\theta^2}(\log f(x;\theta_2) - \log f(x;\theta_1))\right| < k(x)\,|\theta_2 - \theta_1|^\gamma$$

and

$$\int k(x)\, f(x;\theta_0)\,\mathrm{d}x < \infty\,,$$

C.10. the derivatives

$$h(x;\theta) = \frac{\partial}{\partial\theta} F(x;\theta) \quad\text{and}\quad \frac{\partial^2}{\partial\theta^2} F(x;\theta)$$

exist and are bounded on $R^1 \times \mathcal{I}$,

C.11.

$$\int_{-\infty}^{+\infty} x^4\, \mathrm{d}F(x) \leqq K$$

for some $K > 0$,

C.12. $\partial/\partial\theta \log f(x;\theta)$ is bounded on $R^1 \times \mathcal{I}$ and absolutely continuous in x and

$$\frac{\partial^2}{\partial x\,\partial\theta} \log f(x;\theta)$$

exists and bounded on $\mathrm{R}^1 \times \mathcal{I}$.

In fact conditions C.1 – C.9 are the ones* which were used by Ibragimov and Hasminskii to prove

THEOREM C [4], [5]. *Suppose that* C.1 – C.9 *hold and let*

$$\sqrt{(n)}\,(\hat\theta_n - \theta_0) - \frac{1}{\sqrt{n}}\,\frac{1}{I(\theta_0)} \sum_{j=1}^{n} \frac{\partial}{\partial\theta} \log f(X_j;\theta_0) = \varrho_n$$

then

$$\varrho_n = \mathrm{O}(n^{-\varepsilon})$$

with probability 1 *and* $E\varrho_n^2 = \mathrm{O}(n^{-2\varepsilon})$ *for a suitable* $\varepsilon > 0$ *and*

$$E\sqrt{n}(\hat\theta_n - \theta_0)^k < A_k \quad (k = 1, 2, \ldots)$$

for a suitable $A_k > 0$ *if* n *is great enough.***

Now we can formulate our

---

* The second equation of C.7 was not used by them.

** In fact the relation $E\varrho_n^2 = \mathrm{O}(n^{-2\varepsilon})$ is not stated explicitly in [5] but the proof there implies it easily.

THEOREM 1. *Suppose that conditions C.1–C.12 are fulfilled. Then, if the underlying probability space is rich enough, one can construct a K.P.* $K(x, y)$ *such that*

$$\sup_x |G(x, n) - \beta_n(x)| = O(n^\varepsilon)$$

with probability 1, *for some* $0 < \varepsilon < \frac{1}{2}$, *where*

$$G(x, n) = G(x, n; \theta_0) =$$
$$= K(F(x; \theta_0), n) - \frac{h(x; \theta_0)}{I(\theta_0)} \int \frac{\partial}{\partial \theta} \log f(x; \theta_0) \, d_x K(F(x; \theta_0), n)$$

is a G.P. with the covariance function

$$EG(x_1, n_1)\, G(x_2, n_2) = \min(n_1, n_2) \left[\min(F(x_1; \theta_0)\, F(x_2; \theta_0)) - \right.$$
$$\left. - F(x_1; \theta_0)\, F(x_2; \theta_0) - h(x_1; \theta_0)\, h(x_2; \theta_0)\right].$$

Before proving this theorem we give a

LEMMA. *Under the conditions of our Theorem* 1 *we have*

$$L = \int \frac{\partial}{\partial \theta} \log f(x; \theta_0) \, d_x[n(F_n(x) - F(x; \theta_0)) - K(F(x; \theta_0), n)] = O(n^\varepsilon)$$

with probability 1 *for some* $0 < \varepsilon < \frac{1}{2}$ *where* $K(\cdot, \cdot)$ *is the K.P. defined in Theorem* B.

Proof. Clearly we have

$$L = \int_{-\infty}^{-\sqrt[4]{n}} + \int_{-\sqrt[4]{n}}^{+\sqrt[4]{n}} + \int_{+\sqrt[4]{n}}^{\infty}.$$

These integrals can be estimated as follows:

$$\int_{-\sqrt[4]{n}}^{+\sqrt[4]{n}} = - \int_{-\sqrt[4]{n}}^{+\sqrt[4]{n}} (\sqrt{(n)}\, \alpha_n(x) - K(F(x; \theta_0), n)) \frac{\partial^2}{\partial x\, \partial \theta} \log f(x; \theta_0) \, dx +$$
$$+ [\sqrt{(n)}\, \alpha_n(x) - K(F(x; \theta_0), n)]_{-\sqrt[4]{n}}^{+\sqrt[4]{n}} = O(\log^2 n) \sqrt[4]{n} + O(n^\varepsilon)$$

for any $\varepsilon > 0$,

$$\int_{-\infty}^{+\sqrt[4]{n}} = O(\sqrt{(n)}\, \alpha_n(\sqrt[4]{n}) - K(F(\sqrt[4]{(n)}; \theta_0), n)) = O(n^\varepsilon)$$

for some $\frac{1}{2} > \varepsilon > 0$, and the same is true for the integral $\int_{\sqrt[4]{n}}^{\infty}$; this proves our Lemma.

Proof of Theorem 1. Using the K.P. of Theorem B we clearly have

$$n(F_n(x) - F(x; \hat{\theta}_n)) = n(F_n(x) - F(x; \theta_0)) + n(F(x; \theta_0) - F(x; \hat{\theta}_n)) =$$

$$= K(F(x; \theta_0), n) + \delta_n(x) - n(\hat{\theta}_n - \theta_0)\, h(x, \theta_0) - n \frac{(\hat{\theta}_n - \theta_0)^2}{2} \frac{\partial^2 F(x; \theta')}{\partial \theta^2}$$

where $\min(\theta_0, \hat{\theta}_n) \leqq \theta' \leqq \max(\theta_0, \hat{\theta}_n)$.

Since by Theorem C $n(\hat{\theta}_n - \theta_0)^2 = O(n^\gamma)$ for any $\gamma > 0$ with probability 1, we have

$$n(F_n(x) - F(x; \hat{\theta}_n)) = K(F(x; \theta_0), n) - n(\hat{\theta}_n - \theta_0)\, h(x, \theta_0) + O_x(n^\varepsilon) =$$

$$= K(F(x; \theta_0), n) - \frac{h(x; \theta_0)}{I(\theta_0)} \sum_{j=1}^{n} \frac{\partial}{\partial \theta} \log f(X_j; \theta_0) - \sqrt{(n)}\, \varrho_n + O_x(n^\varepsilon) =$$

$$= K(F(x; \theta_0), n) - \frac{h(x; \theta_0)}{I(\theta_0)} \int \frac{\partial}{\partial \theta} \log f(x; \theta_0)\, d_x \sqrt{(n)}\, \alpha_n(x) + O_x(n^\varepsilon) =$$

$$= K(F(x; \theta_0), n) - \frac{h(x; \theta_0)}{I(\theta_0)} \int \frac{\partial}{\partial \theta} \log f(x; \theta_0)\, d_x K(F(x; \theta_0), n) + O_x(n^\varepsilon) =$$

$$= G(x, n) + O_x(n^\varepsilon)$$

where $O_x(n^\varepsilon)$ is a stochastic process for which

$$\overline{\lim_{n \to \infty}} \frac{1}{n^\varepsilon} \sup_x O_x(n^\varepsilon) < \infty$$

with probability 1.

To evaluate the covariance function of $G(x, n)$ is quite an elementary matter.

REMARK. Put

$$W(n) = \frac{1}{I(\theta_0)} \int \frac{\partial}{\partial \theta} \log f(x; \theta_0)\, d_x K(F(x; \theta_0), n).$$

It is worth while to mention that $W(n)$ is a G.P. with covariance

$$E\, W(n_1)\, W(n_2) = \min(n_1, n_2)$$

i.e. $W(n)$ is a Wiener Process.

2. ON THE EMPIRICAL PROCESS WHEN THE PARAMETER IS ESTIMATED FROM A HALF-SAMPLE

Let $\bar{\theta}_n$ be the m.l.e. of θ_0 based on a randomly choosen half of the sample $X_1, X_2, \ldots, X_n$. Without loss of generality we can assume that it is the first half: $X_1, X_2, \ldots, X_{[n/2]}$.

THEOREM 2. *Suppose that conditions* C.1 – C.12 *are fulfilled. Then, if the underlying probability space is rich enough, one can construct a K.P.* $\overline{K}(x, y)$ *such that*

$$\sup_x \left| n(F_n(x) - F(x; \bar{\theta}_n)) - \overline{K}(F(x; \theta_0), n) \right| = O(n^{\varepsilon})$$

with probability 1, *for some* $0 < \varepsilon < \frac{1}{2}$.

REMARK. Since

$$P\left\{\sup_x \frac{1}{\sqrt{n}} \overline{K}(F(x; \theta_0), n) < y\right\} = 1 - e^{-2y^2} \quad \text{if} \quad y > 0$$

and

$$P\left\{\sup_x \frac{1}{\sqrt{n}} \left|\overline{K}(F(x; \theta_0), n)\right| < y\right\} = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2y^2} \quad \text{if} \quad y > 0$$

we have

$$\lim_{n \to \infty} P\{\sup_x \sqrt{(n)}\, (F_n(x) - F(x; \bar{\theta}_n)) < y\} = 1 - e^{-2y^2} \quad \text{if} \quad y > 0$$

and

$$\lim_{n \to \infty} P\{\sup_x \sqrt{(n)}\, |F_n(x) - F(x; \bar{\theta}_n)| < y\} = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2y^2} \quad \text{if} \quad y > 0\,.$$

Proof of Theorem 2. Let $F_n^{(1)}(x)$ resp. $F_n^{(2)}(x)$ be the empirical d.f.'s based on the sample $X_1, X_2, \ldots, X_{[n/2]}$ resp. $x_{[n/2]+1}, \ldots, X_n$, further let $K_1(x, y)$ resp. $K_2(x, y)$ be K.P.'s for which

$$\sup_x \left| \frac{n}{2} (F_n^{(1)}(x) - F(x; \theta_0)) - K_1(F(x; \theta_0), n/2) \right| = O(\log^2 n)$$

resp.

$$\sup_x \left| \frac{n}{2} (F_n^{(2)}(x) - F(x; \theta_0)) - K_2(F(x; \theta_0); n/2) \right| = O(\log^2 n)$$

with probability 1. Without the loss of generality we can assume that $K_2(\cdot, \cdot)$ is independent from $K_1(\cdot, \cdot)$ and also from the sample $X_1, X_2, \ldots, X_{[n/2]}$ (cf. Theorem 4 of [1]). Then we have

$$n(F_n(x) - F(x; \bar{\theta}_n)) = n \frac{F_n^{(1)}(x) - F(x; \theta_0)}{2} + n(F(x; \theta_0) - F(x; \bar{\theta}_n)) +$$

$$+ n \frac{F_n^{(2)}(x) - F(x; \theta_0)}{2} = K_1(F(x; \theta_0), n/2) + n(F(x; \theta_0) - F(x; \bar{\theta}_n)) +$$

$$+ K_2(F(x; \theta_0), n/2) + O(\log^2 n)\,.$$

Using the idea of the proof of Theorem 1 we get

$$n(F(x;\theta_0)-F(x;\bar{\theta}_n)) =$$

$$= -2\frac{h(x;\theta_0)}{I(\theta_0)}\int \frac{\partial}{\partial\theta}\log f(x;\theta_0)\,\mathrm{d}_x K_1(F(x;\theta_0), n/2) + \mathrm{O}_x(n^{\varepsilon}),$$

hence

$$n(F_n(x)-F(x;\bar{\theta}_n)) = K_1(F(x;\theta_0), n/2) + K_2(F(x;\theta_0), n/2) -$$

$$-2\frac{h(x;\theta_0)}{I(\theta_0)}\int \frac{\partial}{\partial\theta}\log f(x;\theta_0)\,\mathrm{d}_x K_1(F(x;\theta_0), n/2) + \mathrm{O}_x(n^{\varepsilon}).$$

Clearly the process

$$\bar{K}(F(x;\theta_0), n) = K_1(F(x;\theta_0), n/2) + K_2(F(x;\theta_0), n/2) -$$

$$-2\frac{h(x;\theta_0)}{I(\theta_0)}\int \frac{\partial}{\partial\theta}\log f(x;\theta_0)\,\mathrm{d}_x K_1(F(x;\theta_0), n/2)$$

is a G.P. and by a simple calculation one gets

$$E\,\bar{K}(F(x_1;\theta_0), n_1)\,\bar{K}(F(x_2;\theta_0), n_2) =$$

$$= \min(n_1, n_2)\,[\min(F(x_1;\theta_0), F(x_2;\theta_0)) - F(x_1;\theta_0)\,F(x_2;\theta_0)]$$

which proves that $\bar{K}$ is a K.P.

REMARK. Let $G(x,\theta)$ be the inverse of $F(x;\theta)$ i.e. $F(G(x,\theta),\theta) = x$. Several times the sample $Y_k = Y_k^{(n)} = F(X_k;\bar{\theta}_n)$ $(k = 1, 2, \ldots, n)$ is investigated instead of $\{X_k\}_{k=1}^{n}$. Let $\bar{F}_n(x)$ be the empirical d.f. based on the sample $Y_1, Y_2, \ldots, Y_n$. Clearly $F_n(G(x,\bar{\theta}_n)) = \bar{F}_n(x)$, hence, by Theorem 2, we have

$$\sup_x |n(\bar{F}_n(x) - x) - K(F(G(x,\bar{\theta}_n);\theta_0), n)| = O(n^{\varepsilon})$$

and clearly

$$\sup_x |K(F(G(x,\bar{\theta}_n);\theta_0), n) - K(x, n)| = \mathrm{O}(n^{\varepsilon})$$

i.e.

$$\sup_x |n(\bar{F}_n(x) - x) - K(x, n)| = \mathrm{O}(n^{\varepsilon})$$

with probability 1.

3. AN ESTIMATION OF $\beta_n(x)$ INDEPENDENT FROM θ_0

In this paragraph it will be also assumed that $dI(\theta)/d\theta$, $\theta \in \mathscr{I}$ is bounded.
The following lemma can be immediately obtained from the definition of $G(x, n, \theta)$.

LEMMA. *We have*

$$\sup_x \left|G(x, n; \hat{\theta}_n) - G(x, n; \theta_0)\right| = O(n^\varepsilon)$$

with probability 1 *for some* $0 < \varepsilon < \frac{1}{2}$.

This lemma and Theorem 1 imply

THEOREM 3. *We have*

$$\sup_x \left|\sqrt{(n)}\,(F_n(x) - F(x; \hat{\theta}_n)) - G(x, n; \hat{\theta}_n)\right| = O(n^\varepsilon)$$

with probability 1 *for some* $0 < \varepsilon < \frac{1}{2}$.

4. THE MULTIVARIATE CASE

It is an important task to generalize these results to the case when x and θ are varying in a more dimensional Euclidean space, say $x \in R^k$, $\theta \in R^m$.

It can be seen that the most important tools of the proofs are Theorems B and C. Hence, if we have multivariate generalizations of these theorems then one can prove multivariate generalizations of our Theorems 1, 2, 3.

A multivariate generalization of Theorem B is known (see [6]). Theorem C is originally formulated for the case when $x \in R^k$, $\theta \in R^1$. It means that Theorems 1, 2, 3 can be generalized for the case $x \in R^k$, $\theta \in R^1$ without any difficulty and they could be generalized for the case $x \in R^k$, $\theta \in R^m$ if we could generalize Theorem C for this case.

The weak version of Theorem C is well-known (see e.g. [7], p. 500) what shows that weak versions of our Theorems can be proved in the general case too.

REFERENCES

[1] J. Komlós, P. Major, G. Tusnády: An approximation of Partial Sums of Independent RV's, and the Sample DF I. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete *32*, (1975), 111—131.

[2] J. Durbin: Distribution theory for tests based on the sample distribution function. Regional Conference Series in Applied Math. 9 (1973) SIAM, Philadelphia.

[3] J. Durbin: Weak convergence of the sample distribution function when parameters are estimated. Ann. Stat. *1* (1973), 279—290.

[4] И. А. Ибрагимов, Р. З. Хасьминский: Ассимптотическое поведение некоторых статистических оценок II. Теория вепоятностей и ее прим. *18* (1973), 78—93.

[5] I. A. Ibragimov, R. Z. Has'minskii: On the approximation of statistical estimators by sums of independent random variables Soviet Math. Dokl. *14* (1973), 883—887.

[6] M. Csörgő, P. Révész: A New Method to Prove Strassen Type Laws of Invariance Principle, II. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete *31* (1975), 261—269.

[7] H. Cramer: Mathematical Methods of Statistics. Princeton Univ. Press, Princeton 1946.

Carleton University Ottawa

Department of Mathematics

Hungarian Academy of Sciences

Mathematical Institute