COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
11. LIMIT THEOREMS OF PROBABILITY THEORY, KESZTHELY (HUNGARY), 1974.

WEAK CONVERGENCE AND EMBEDDING

J. KOMLÓS – P. MAJOR – G. TUSNÁDY

1. INTRODUCTION

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables (i.i.d.r.v.) with distribution function $F \in \mathcal{F}_r$, where

$$\mathcal{F}_r = \{F\colon F \text{ is a distribution function}, \quad \int_{-\infty}^{\infty} x\,dF = 0, \quad \int_{-\infty}^{\infty} x^2\,dF = 1, \quad \int_{-\infty}^{\infty} |x|^r\,dF < \infty\}. \tag{1.1}$$

Let the function s_n be defined on $[0, 1]$ by

$$s_n(t) = \frac{1}{\sqrt{n}}\left[X_1 + \ldots + X_{k-1} + n\left(t - \frac{k-1}{n}\right)X_k\right] \quad \text{for} \quad \frac{k-1}{n} \leqslant t \leqslant \frac{k}{n}; \qquad k = 1, 2, \ldots, n; \tag{1.2}$$

and let the probability measure S_n^F be generated by $s_n(t)$ on the space $C(0, 1)$ of continuous functions on $[0, 1]$. It is well known that the sequence S_n^F converges to the Wiener measure W on $C(0, 1)$ for any

$F \in \mathscr{F}_2$. Similarly, if $Y_1, Y_2, \ldots$ are i.i.d.r.v.-s uniformly distributed on $[0, 1]$, the function z_n is defined by

$$z_n(t) = \sqrt{n}\left(\frac{1}{n} \sum_{i:\, Y_i < t} 1 - t\right) \qquad \text{for} \qquad 0 \leqslant t \leqslant 1\,, \tag{1.3}$$

and Z_n is the distribution of z_n on the space $D(0, 1)$ of real functions on $[0, 1]$ without discontinuities of second kind, then Z_n tends weakly to the distribution B of the Brownian bridge.

The weak convergence can be metrized by the Prohorov distance, which is defined in a metric space M by

$$\begin{aligned} \rho(\mathsf{P}_0, \mathsf{P}_1) = \inf\{\epsilon > 0\colon\ & \mathsf{P}_0(E) \leqslant \epsilon + \mathsf{P}_1\{y\colon \exists x \in E, d(x, y) < \epsilon\} \\ & \text{for all closed } E\}\,, \end{aligned} \tag{1.4}$$

where d is the metric of the space, especially on $C(0, 1)$ it is

$$d(x, y) = \sup_{0 \leqslant t \leqslant 1} |x(t) - y(t)|\,, \tag{1.5}$$

and the metric on $D(0, 1)$ is

$$\begin{aligned} d_0(x, y) = \inf\{\epsilon > 0\colon\ & \exists \text{ a homeomorphism } \varphi \\ & \text{of } [0, 1]\colon |\varphi(t) - t| \leqslant \epsilon, \text{ and } \sup_{0 \leqslant t \leqslant 1} |x(\varphi(t)) - y(t)| \leqslant \epsilon\}\,. \end{aligned} \tag{1.6}$$

Hence $\rho(\mathsf{S}_n^F, \mathsf{W})$ tends to 0 for any $F \in \mathscr{F}_2$, and also $\lim_{n \to \infty} \rho(\mathsf{Z}_n, \mathsf{B}) = 0$.

The investigation of the rate of convergence of $\rho(\mathsf{S}_n^F, \mathsf{W})$ or $\rho(\mathsf{Z}_n, \mathsf{B})$ is usually based on the following remark. Let M be any metric space with a metric d, and let $\mathsf{P}_0, \mathsf{P}_1$ be probability measures on $(M, \mathscr{B})$, where $\mathscr{B}$ is the σ-algebra of Borel-measurable sets of M. Let $p_0(\omega)$, $p_1(\omega)$ be Borel-measurable mappings from a probability space $(\Omega, \mathscr{A}, \mathsf{P})$ to $(M, \mathscr{B})$ such that p_i generates the measure P_i on M $(i = 0, 1)$. Then

$$\rho(\mathsf{P}_0, \mathsf{P}_1) \leqslant \inf_{\epsilon > 0} (\epsilon + \mathsf{P}(d(p_0, p_1) \geqslant \epsilon))\,. \tag{1.7}$$

We shall call such a pair p_0, p_1 a *coupling* of the measures $\mathrm{P}_0, \mathrm{P}_1$. Especially, if $(\Omega, \mathscr{A}, \mathrm{P})$ equals $(M, \mathscr{B}, \mathrm{P}_0)$, and φ_0 is the identity, then φ_1 will be called an embedding of P_1 into P_0.

The first coupling of S_n^F and W was made by Skorohod, and it is the following. For any $F \in \mathscr{F}_2$ there is a Wiener process $w(x)$ and a sequence of i.i.d.r.v.-s τ_j such that the distribution of the function $s_n(t)$ is S_n^F, where

$$
\begin{aligned}
&s_n(t) = \frac{\vartheta_k - t}{\vartheta_k - \vartheta_{k-1}}\, w(\vartheta_{k-1}) + \frac{t - \vartheta_{k-1}}{\vartheta_k - \vartheta_{k-1}}\, w(\vartheta_k) \\
&\text{for } \vartheta_{k-1} \leqslant t \leqslant \vartheta_k, \qquad k = 1, 2, \ldots, n;
\end{aligned}
\tag{1.8}
$$

and $\vartheta_k = \tau_1 + \ldots + \tau_k$, $\tau_0 = 0$. Here $\mathrm{E}\tau_1 = 1$, and if $F \in \mathscr{F}_4$ then $\mathrm{E}\tau_1^2 < \infty$. Using this coupling of S_n^F and W, Dudley [8] proved that if $F \in \mathscr{F}_r$, then

$$
\rho(\mathrm{S}_n^F, \mathrm{W}) = \begin{cases} O\Big((\log n)^{\frac{1}{2}} n^{-\frac{r-2}{2(r+1)}}\Big) & \text{for } 2 < r \leqslant 4, \\ O(n^{-\frac{1}{2}} \log n) & \text{for } r \geqslant 5, \end{cases}
\tag{1.9}
$$

and

$$
\rho(\mathrm{Z}_n, \mathrm{B}) = O(n^{-\frac{1}{4}} \log n) .
\tag{1.10}
$$

Sawyer [14] showed that the Skorohod embedding method cannot yield a speed of convergence faster than $O(n^{-\frac{1}{4}})$. Borovkov [3] proved that $\rho(\mathrm{S}_n^F, \mathrm{W}) = o\Big(n^{\frac{r-2}{2(r+1)}}\Big)$ for $2 \leqslant r \leqslant 3$.

The goodness of a coupling can be characterized in different ways. Let $\mathscr{C}(\mathrm{P}_0, \mathrm{P}_1)$ be the set of all possible couplings of $\mathrm{P}_0, \mathrm{P}_1$ (where also the space $(\Omega, \mathscr{A}, \mathrm{P})$ may vary). Strassen and Dudley [7] proved that if M is separable, then

$$
\rho(\mathrm{P}_0, \mathrm{P}_1) = \inf_{(p_0, p_1) \in \mathscr{C}(\mathrm{P}_0, \mathrm{P}_1)} \; \inf_{\epsilon > 0} \, (\epsilon + \mathrm{P}(d(p_0, p_1) \geqslant \epsilon)) .
\tag{1.11}
$$

In the light of this theorem we can find couplings which yield good estimations for the Prohorov distance. An other possibility for measuring the goodness of coupling is to check whether $\mathsf{E}d(p_0, p_1)$ is near to the Wasserstein-distance of $\mathsf{P}_0, \mathsf{P}_1$ which is defined by

$$(1.12) \qquad \rho_0(\mathsf{P}_0, \mathsf{P}_1) = \inf_{(p_0, p_1) \in \mathscr{C}(\mathsf{P}_0, \mathsf{P}_1)} \mathsf{E}d(p_0, p_1).$$

One may also try to find the coupling, which minimizes the probability $\mathsf{P}(d(p_0, p_1) > \epsilon)$ for a given ϵ.

A pair of random functions $\tilde{s}(t), \tilde{w}(t)$ defined for $0 \leqslant t < \infty$ will be called an infinite coupling of $\mathsf{S}_n^F, \mathsf{W}$ if the pair

$$(1.13) \qquad s_n(t) = \frac{\tilde{s}(nt)}{\sqrt{n}}; \quad w_n(t) = \frac{\tilde{w}(nt)}{\sqrt{n}}, \qquad 0 \leqslant t \leqslant 1$$

is a coupling of $\mathsf{S}_n^F, \mathsf{W}$ for $n = 1, 2, \ldots$. Strassen extended the Skorohod embedding to an infinite coupling, and proved in [15] that if $F \in \mathscr{F}_4$, then for the corresponding infinite coupling

$$(1.14) \qquad \sup_{0 \leqslant t \leqslant n} |\tilde{s}(t) - \tilde{w}(t)| = O(n^{\frac{1}{4}} (\log n)^{\frac{1}{2}} (\log \log n)^{\frac{1}{4}})$$

holds true with probability 1. It was widely guessed that the Skorohod embedding, at least asymptotically, is best possible. This guess was disproved by Csörgő and Révész [4]. Following their and Bártfai's ideas we proved in [13] the following theorems.

Theorem A. *If* $F \in \mathscr{F}_2$, *and* $\int_{-\infty}^{\infty} e^{tx} F(dx) < \infty$ *for* $|t| < t_0$ $(t_0 > 0)$, *then there is an infinite coupling* $\tilde{s}, \tilde{w}$ *of* $\mathsf{S}_n^F, \mathsf{W}$ *such that for all* $x > 0$ *and every* n

$$(1.15) \qquad \mathsf{P}(\sup_{0 \leqslant t \leqslant n} |\tilde{s}(t) - \tilde{w}(t)| \geqslant x) \leqslant K n^C e^{-\lambda x},$$

where C, K, λ *are positive constants depending only on* F.

Theorem B. *If* $F \in \mathscr{F}_r$, $r > 3$, *then there is a constant* C *and a sequence* ϵ_n *tending to* 0 *with* $n \to \infty$ *such that for every* n *and* x, $n^{\frac{1}{r}} < x < C n^{\frac{1}{2}} (\log n)^{\frac{1}{2}}$, *there is a coupling* s_n, w, *of* $\mathsf{S}_n^F, \mathsf{W}$ *such that*

$$(1.16) \qquad \mathsf{P}(\sup_{0 \leq t \leq 1} n^{\frac{1}{2}} | s_n(t) - w(t) | \geq x) \leq n\epsilon_n x^{-r} .$$

Theorem C. *For every* n *there is a coupling* z_n, b *of* Z_n *and* B *such that for all* $x > 0$

$$(1.17) \qquad \mathsf{P}(\sup_{0 \leq t \leq 1} n^{\frac{1}{2}} | z_n(t) - b(t) | \geq x) \leq K n^C e^{-\lambda x} ,$$

where C, K, λ *are positive absolute constants.*

Our couplings are, in fact, embeddings of S_n^F (or Z_n) into W (or into B), and they are constructed by a new method called conditional quantile transformation on a diadic scheme. As we shall see, the investigation of couplings is closely related to the so-called stochastic geyser problem, so this is the topic of the next section.

2. THE STOCHASTIC GEYSER PROBLEM

Once upon a time there was a man who lived on a desolate island, and the only companion he had was a geyser. The geyser burst out periodically, and the man wanted to report the distribution of the random period of the consecutive bursts to his homeland. He had no watch and no possibility whatever to measure the hours. So he created a primitive calendar and put down carefully, day by day, the total number of bursts. Our man had a very long, let us say an infinitely long life, and only after he, and the geyser had died, found his fellow-countrymen the whole infinite sequence of his records. Can they figure out the distribution of the burst-time of the geyser?

The answer is yes, and it was given by Bártfai [1]. He investigated the following problem. Let $X = \{X_n;\ n = 1, 2, \ldots\}$ be i.i.d.r.v.-s with distribution function F, and $S_F = \{X_1 + \ldots + X_n;\ n = 1, 2, \ldots\}$ the sequence of partial sums. Let $e = \{e_n;\ n = 1, 2, \ldots\}$ be an arbitrary sequence of r.v.'s such that

$$(2.1) \qquad \mathsf{P}\left(\limsup_{n \to \infty} \frac{e_n}{a_n} \leq 1\right) = 1 ,$$

where $a = \{a_n;\ n = 1, 2, \ldots\}$ is a given monotone increasing sequence. We call the sequence a estimation-permitting in a set $\mathcal{F}$ of distributions, if the infinite sample

$$S_F(e) = \{X_1 + \ldots + X_n + e_n;\ n = 1, 2, \ldots\} \tag{2.2}$$

determines the distribution F with probability 1 for all $F \in \mathcal{F}$ and for all e having the property (2.1). This means that given a and $\mathcal{F}$ we can construct a system $\{A_F,\ F \in \mathcal{F}\}$ of disjoint Borel-measurable sets of infinite sequences of real numbers in such a way that for any $F \in \mathcal{F}$

$$\mathrm{P}(S_F(e) \in A_F) = 1 \tag{2.3}$$

for any choice of S_F and e such that e satisfies (2.1). Let $E(\mathcal{F})$ be the set of all monotone increasing sequences, estimation-permitting in $\mathcal{F}$. Combining Theorem A and the theorem of Bártfai we get

Theorem 1. *Let the set* $\mathcal{F}_\infty$ *be defined by*

$$\mathcal{F}_\infty = \{F\colon \text{ there is a } t_0 > 0 \text{ such that } \int_{-\infty}^{\infty} e^{tx} F(dx) < \infty \text{ for } |t| < t_0\}, \tag{2.4}$$

then $a \in E(\mathcal{F}_\infty)$ *if and only if*

$$\liminf_{n \to \infty} \frac{a_n}{\log n} = 0 . \tag{2.5}$$

This theorem will be proved later. First we define two other sets similar to $E(\mathcal{F})$. The first of them is the set $T(F_0, \mathcal{F})$ of monotone increasing test-permitting sequences in $\mathcal{F}$ with respect to F_0. A sequence a is test-permitting in $\mathcal{F}$ with respect to F_0 if the infinite sample $S_F(e)$ determines, with probability 1, whether $F = F_0$ or $F \in \mathcal{F}$ (of course $F_0 \notin \mathcal{F}$). This means that given $a, F_0, \mathcal{F}$ there is a measurable set A_0 such that for any choice of the error e satisfying (2.1) we have

$$\mathrm{P}(S_{F_0}(e) \in A_0) = 1 , \tag{2.6}$$

for any choice of S_{F_0} and

(2.7) $\quad P(S_F(e) \in A_0) = 0$

for any choice of $F \in \mathscr{F}$ and S_F.

Lemma 1. *For any* $F_0 \notin \mathscr{F}_1$, $\mathscr{F} = \{F_0\} \cup \mathscr{F}_1 \subset \widetilde{\mathscr{F}}$, $\widetilde{\mathscr{F}}_1 \subset \mathscr{F}_1$

(2.8) $\quad E(\widetilde{\mathscr{F}}) \subset E(\mathscr{F}) \subset T(F_0, \mathscr{F}_1) \subset T(F_0, \widetilde{\mathscr{F}}_1)$

holds true.

Proof. The monotonity of the sets $E(\mathscr{F})$, $T(F_0, \mathscr{F}_1)$ is a trivial consequence of their definition. For proving the third statement choose $A_0 = A_{F_0}$.

A sequence a is couple-permitting with respect to F_0 and F_1 if there is an infinite coupling $\tilde{s}_0, \tilde{s}_1$ of $S_n^{F_0}$ and $S_n^{F_1}$ such that

(2.9) $$P\left(\limsup_{n \to \infty} \frac{|\tilde{s}_0(n) - \tilde{s}_1(n)|}{a_n} \leqslant 2\right) = 1 .$$

The set of all monotone increasing couple-permitting sequences with respect to F_0 and F_1 will be denoted by $C(F_0, F_1)$.

Lemma 2. *The sets* $C(F_0, F_1)$, $T(F_0, F_1)$ *are disjoint for any* F_0, F_1.

Proof. Assume that there is a sequence a such that $a \in C(F_0, F_1) \cap T(F_0, F_1)$. Then there is a set such that (2.6) and (2.7) hold true with $\mathscr{F} = \{F_1\}$, and there is an infinite coupling $\tilde{s}_0, \tilde{s}_1$ such that (2.9) holds true. Put $S_{F_0} = \{\tilde{s}_0(n);\ n = 1, 2, \ldots\}$, $S_{F_1} = \{\tilde{s}_1(n);\ n = 1, 2, \ldots\}$,

$e = \left\{\frac{\tilde{s}_0(n) - \tilde{s}_0(n)}{2};\ n = 1, 2, \ldots\right\}$, then e satisfies (2.1), and

$$P(S_{F_0}(e) \in A_0) = P(S_{F_1}(-e) \in A_0) =$$
$$= P\left(\left\{\tfrac{1}{2}(\tilde{s}_0(n) + \tilde{s}_1(n));\ n = 1, 2, \ldots\right\} \in A_0\right),$$

hence (2.6) and (2.7) can not hold true simultaneously.

Theorem 2. *For any* $F_0, F_1 \in \mathscr{F}_2 \cap \mathscr{F}_\infty$ *(cf.* (1.1) *and* (2.4)*) there are positive constants* $t(F_0, F_1) \leqslant c(F_0, F_1)$ *such that if a monotone increasing sequence* a *satisfies*

$$\liminf_{n\to\infty} \frac{a_n}{\log n} < t(F_0, F_1), \tag{2.10}$$

then $a \in T(F_0, F_1)$, *and if*

$$\liminf_{n\to\infty} \frac{a_n}{\log n} \geqslant c(F_0, F_1), \tag{2.11}$$

then $a \in C(F_0, F_1)$.

Proof. If $F_0, F_1 \in \mathscr{F}_2 \cap \mathscr{F}_\infty$, then Theorem A implies that there are infinite couplings $\tilde{s}_i, \tilde{w}$ of $\mathsf{S}_n^{F_i}$ and W such that

$$\mathsf{P}\left(\limsup_{n\to\infty} \frac{|\tilde{s}_i(n) - \tilde{w}(n)|}{\log n} \leqslant C_i\right) = 1 \qquad (i = 1, 2) \tag{2.12}$$

where C_0, C_1 are positive constants. As we have remarked, our couplings are embeddings of $\mathsf{S}_n^{F_i}$ in W, hence in the couplings $(\tilde{s}_0, \tilde{w}), (\tilde{s}_1, \tilde{w})$ the process $\tilde{w}$ may be chosen the same. Hence $(\tilde{s}_0, \tilde{s}_1)$ is an infinite coupling of $\mathsf{S}_n^{F_0}$ and $\mathsf{S}_n^{F_1}$, and (2.9) holds true with $a_n = \frac{1}{2}(C_0 + C_1)\log n$. Thus the second statement of Theorem 2 holds true with $c(F_0, F_1) = \frac{1}{2}(C_0 + C_1)$. The proof of the first statement is based on the following theorem of Erdős and Rényi [10].

Theorem D. *Let* $X_1, X_2, \ldots$ *be i.i.d.r.v.-s with distribution function* $F \in \mathscr{F}_\infty$. *Let* $R(t)$ *be the moment-generating function of* F, $A^+ = \inf\{x\colon F(x) = 1\}$, $A^- = \sup\{x\colon F(x) = 0\}$ *and* $\pi(x) = \sup_t (tx - \log R(t))$. *Then for any* $\mathsf{E}X_1 < x < A^+$

$$\mathsf{P}\left(\lim_{n\to\infty} \sup_{1\leqslant j\leqslant n-m} \frac{1}{m}\sum_{k=j}^{j+m} X_k = x\right) = 1, \tag{2.13}$$

and for any $A^- < x < \mathsf{E}X_1$

(2.14) $$P\left(\lim_{n\to\infty} \inf_{1\leq j\leq n-m} \frac{1}{m}\sum_{k=j}^{j+m} X_k = x\right) = 1 ,$$

where $m = \left[\frac{\log n}{\pi(x)}\right]$.

Let us denote the corresponding quantities for F_i by A_i^+, A_i^- and $\pi_i(x)$ for $i = 0, 1$ and let $B_i^+ = \lim_{x=A_i^+-0} \pi_i(x)$, $B_i^+ = \lim_{x=A_i^-+0} \pi_i(x)$, $B^+ = \min(B_0^+, B_1^-)$, $B^- = \min(B_0^-, B_1^-)$. The functions $\pi_i(x)$ are monotone increasing for $x \geq 0$, let us denote their inverse here by $u_i(x)$, and let $v_i(x)$ be the inverse of $\pi_i(x)$ for $x \leq 0$. Then we prove the first statement of Theorem 2 with the following constant

$$t(F_0, F_1) = \frac{1}{4}\max\left(\sup_{0<x<B^+} \frac{|u_0(x) - u_1(x)|}{x}, \sup_{0<x<B^-} \frac{|v_0(x) - v_1(x)|}{x}\right).$$

Indeed, if a monotone increasing sequence a statisfies (2.10) with this $t(F_0, F_1)$, then there are real numbers x_0, x_1 such that $x_0 x_1 > 0$, $\pi_0(x_0) = \pi_1(x_1) = A$, and there is a sequence n_k such that

$$\lim_{k\to\infty} \frac{a_{n_k}}{\log n_k} < \frac{|x_0 - x_1|}{4A} .$$

We may assume that $0 < x_0 < x_1$. Let A_0 be the set of all sequences $\{c_k\}$ of real numbers such that

$$\limsup_{k\to\infty} \sup_{1\leq j\leq n_k - m_k} \frac{1}{m_k}(c_{j+m_k} - c_j) < \frac{x_0 + x_1}{2} ,$$

where $m_k = \left[\frac{\log n_k}{A}\right]$, then (2.6) and (2.7) hold true with this A_0, hence $a \in T(F_0, F_1)$.

Proof of Theorem 1. It is easy to see that for any $\epsilon > 0$ there are distributions F_i such that (2.12) holds with a $C_i < \epsilon$ (e.g. let F_i be the convolution power of some functions in $\mathscr{F}_\infty$). Hence Theorem 2, Lemma 1 and Lemma 2 imply that (2.5) is necessary to $a \in E(\mathcal{F}_\infty)$. The sufficiency is a consequence of Theorem D and the fact that the function $\pi(x)$ determines the distribution function F.

It is easy to see that Theorem D remains valid if we substitute the set $\mathcal{F}_\infty$ by

$$\widetilde{\mathcal{F}}_\infty = \{F: \text{ there is a } t \neq 0 \text{ such that } \int_{-\infty}^{\infty} e^{tx} F(dx) < \infty\}\,.$$

Hence Theorem 1 and the first statement of Theorem 2 hold true also for this $\widetilde{\mathcal{F}}_\infty$.

It is a natural question, what is the situation if in the above theorems we substitute $\mathcal{F}_\infty$ by $\mathcal{F}_r$. One can expect that functions with few moments are far from each other, farther than functions in $\mathcal{F}_\infty$. Surprisingly enough, this is not so. Halász and Major [12] proved that there are distributions F_0, F_1 such that there is an infinite coupling $\tilde{s}_0$, $\tilde{s}_1$ of $S_n^{F_0}, S_n^{F_1}$ such that

$$\mathsf{P}(|\tilde{s}_0(n) - \tilde{s}_1(n)| \leqslant K,\ n = 1, 2, \ldots) = 1$$

with some positive constant K. Actually they proved that even this K may be arbitrarily small and still $F_0, F_1 \in \bigcap_{r=2}^{\infty} \mathcal{F}_r$. Hence $E\left(\bigcap_{r=2}^{\infty} \mathcal{F}_r\right)$ is empty. On the other hand, it is easy to see that for any F_0, F_1 there is an ϵ such that the sequence $\{a_n = \epsilon,\ n = 1, 2, \ldots\}$ is an element of $T(F_0, F_1)$. It is also trivial that

$$\mathsf{P}(\lim_{n\to\infty} e_n = 0) = 1$$

implies that $S_F(e)$ determines F with probability 1 within any family $\mathcal{F}$.

One can expect that the sets $E(\mathcal{F})$, $T(F_0, \mathcal{F})$, $C(F_0, F_1)$ determine mutually each other in the sense that

$$E(\mathcal{F}) = \bigcap_{F_0 \in \mathcal{F}} T(F_0, \mathcal{F} \setminus \{F_0\})\,,$$

$$T(F_0, \mathscr{F} \setminus \{F_0\}) = \bigcap_{F_1 \in \mathscr{F} \setminus \{F_0\}} T(F_0, F_1),$$

and $T(F_0, F_1) \cup C(F_0, F_1)$ is the set of all monotone increasing sequences. Let $C_0(F_0, F_1)$ be the set of all monotone increasing sequences a such that there is an infinite coupling $\tilde{s}_0, \tilde{s}_1$ of $\mathsf{S}_n^{F_0}$ and $\mathsf{S}_n^{F_1}$ such that

$$\mathsf{P}\left(\limsup_{n \to \infty} \frac{|\tilde{s}_0(n) - \tilde{s}_1(n)|}{a_n} \leq 2\right) > 0 .$$

The sets $C_0(F_0, F_1)$, $T(F_0, F_1)$ are also disjoint for any F_0, F_1, hence our conjecture means that $C_0(F_0, F_1) = C(F_0, F_1)$. Nevertheless, we can not prove any statement of this type.

In particular, we cannot answer the following question. Let $\mathscr{F}$ be the set of all distributions F such that for any G we can test whether the sample-distribution is F or G if the error is bounded. Can we estimate the sample-distribution in $\mathscr{F}$ if the error is bounded?

3. THE PROHOROV DISTANCE

Theorem 3. *If $F \in \mathscr{F}_2 \cap \mathscr{F}_\infty$ (cf. (1.1) and (2.4)), then there are positive constants C_0, C_1 such that for all $n > 1$*

$$C_0 \frac{\log n}{\sqrt{n}} \leq \rho(\mathsf{S}_n^F, \mathsf{W}) \leq C_1 \frac{\log n}{\sqrt{n}} . \tag{3.1}$$

Proof. Theorem A implies that if $F \in \mathscr{F}_2 \cap \mathscr{F}_\infty$, then

$$\rho(\mathsf{S}_n(F), \mathsf{W}) \leq \inf_{0 < \epsilon < 1} (\epsilon + Kn^C e^{-\lambda\epsilon\sqrt{n}}) \leq C_1 \frac{\log n}{\sqrt{n}},$$

which is the upper part of (3 1). Theorem D implis that there are positive constants Δ_0, C_0, n_0 such that if A_n is the subset of $C(0, 1)$ consisting of all functions f for which

$$\sup_{0 \leq t \leq 1 - \Delta_n} |f(t + \Delta_n) - f(t)| \geq C_0 \frac{\log n}{\sqrt{n}}, \tag{3.2}$$

where $\Delta_n = \Delta_0 \frac{\log n}{n}$, then the probability of A_n and $\bar{A}_n$ is greater than 3/4 with respect to S_n^F and W, respectively, for any $n > n_0$ (or conversely, the probability of A_n is large with respect to W and the probability of $\bar{A}_n$ is large with respect to S_n^F). This yields the lower part of (3.1).

Theorem 4. *If* $r > 2$, $F \in \mathscr{F}_r$ *(cf.* (1.1)*), then*

$$\rho(\mathsf{S}_n^F, \mathsf{W}) = o\left(n^{-\frac{r-2}{2(r+1)}}\right) . \tag{3.3}$$

On the other hand, for any sequence ω_n *tending to* ∞ *with* $n \to \infty$ *there is an* $F \in \mathscr{F}_r$ *such that*

$$\limsup_{n\to\infty} \omega_n n^{-\frac{r-2}{2(r+1)}} \rho(\mathsf{S}_n^F, \mathsf{W}) = \infty . \tag{3.4}$$

Proof. For $2 < r \leqslant 3$ (3.3) was proved by Borovkov [3]. If $r > 3$ and $F \in \mathscr{F}_r$, Theorem B implies that

$$\rho(\mathsf{S}_n^F, \mathsf{W}) \leqslant \inf_{0<\epsilon<1} \left(\epsilon + \frac{n\epsilon_n}{(\epsilon\sqrt{n})^r}\right) \leqslant K\epsilon_n^{\frac{1}{r+1}} n^{-\frac{r-2}{2(r+1)}}$$

with some positive constant K. So the first part is proved.

Given any sequences ω_n, x_n tending to ∞ with $n \to \infty$ there is an $F \in \mathscr{F}_r$ such that

$$\limsup_{n\to\infty} \omega_n x_n^r (1 - F(x_n)) = \infty .$$

Hence if $X_1, X_2, \ldots$ are i.i.d.r.v.-s with distribution F, then

$$\limsup_{n\to\infty} \omega_n n^{\frac{r-2}{2(r+1)}} \mathsf{P}\left(\max (X_1, \ldots, X_n) \geqslant n^{\frac{3}{2(r+1)}}\right) = \infty .$$

This yields (3.4).

Theorem 5. *There are positive constants* C_0, C_1 *such that for every* $n > 1$

$$C_0 \frac{\log n}{\sqrt{n}} \leqslant \rho(\mathsf{Z}_n, \mathsf{B}) \leqslant C_1 \frac{\log n}{\sqrt{n}} . \tag{3.5}$$

Proof. The metric d_0 of $D(0,1)$ is majorated by the metric $C(0,1)$, hence Theorem C implies that

$$\rho(Z_n, B) \leqslant \inf_{0<\epsilon<1} (\epsilon + P(d_0(z_n, b) \geqslant \epsilon) \leqslant$$

$$\leqslant \inf_{0<\epsilon<1} (\epsilon + P(d(z_n, b) \geqslant \epsilon) \leqslant C_1 \frac{\log n}{\sqrt{n}}$$

with some positive constant C_1 where (z_n, b) is the coupling given by Theorem C.

For proving the lower part of (3.5) we shall apply again Theorem D, now for Poisson variables. If $X_1, X_2, \ldots$ are independent Poisson variables with parameter 1, then the moment generating function of $(X_1 - 1)$ is $R(t) = \exp\{e^t - 1 - t\}$, hence

$$\pi(x) = 1 + (x+1)(\log(x+1) - 1)\,. \tag{3.6}$$

Let $y(t, \lambda)$ be a Poisson process on the plane with parameter 1, and let $Y_\lambda(\Delta)$ be defined by

$$Y_\lambda(\Delta) = \sup_{0 \leqslant t \leqslant 1-\Delta} (\tilde{y}(t+\Delta, \lambda) - \tilde{y}(t, \lambda))\,, \tag{3.7}$$

where $\tilde{y}(t, \lambda) = (y(1, \lambda))^{-1/2}(y(t, \lambda) - ty(1, \lambda))$. In the same way, as Theorem D was proved, one can prove that

$$P\Big(\lim_{\lambda \to \infty} \frac{\sqrt{\lambda}}{\log \lambda} Y_\lambda\Big(\frac{\log \lambda}{\lambda \pi(x)}\Big) = \frac{x}{\pi(x)}\Big) = 1\,. \tag{3.8}$$

It is well known that if $w(x)$ is a Wiener process and $W(\Delta)$ is defined by

$$W(\Delta) = \sup_{0<\delta\leqslant\Delta}\ \sup_{0\leqslant t\leqslant 1-\delta} (w(t+\delta) - w(t))\,, \tag{3.9}$$

then

$$P\Big(\lim_{\lambda \to \infty} \frac{\sqrt{\lambda}}{\log \lambda} W\Big(\frac{\log \lambda}{\lambda \pi(x)}\Big) = \sqrt{\frac{2}{\pi(x)}}\Big) = 1\,. \tag{3.10}$$

For the function π defined by (3.6)

$$\lim_{x\to\infty} \left(\frac{x}{\pi(x)} - \sqrt{\frac{2}{\pi(x)}}\right) = \frac{1}{3}$$

holds true, hence it is enough to prove that (3.8) holds true if we substitute $Y_\lambda(\Delta)$ in it with

$$Z_n(\Delta) = \sup_{0\leq t\leq 1-\delta} (z_n(t+\Delta) - z_n(t)),$$

where $z_n(t)$ is defined by (1.3), i.e. if we prove that

$$\text{(3.11)} \qquad \mathsf{P}\left(\lim_{n\to\infty} \frac{\sqrt{n}}{\log n} Z_n\left(\frac{\log n}{n\pi(x)}\right) = \frac{x}{\pi(x)}\right) = 1,$$

where $\pi(x)$ is the same function as in (3.8). Let $\nu(x)$ be a Poisson process on the real line, and let the processes $\nu(\lambda)$, $\{z_n(t),\ n = 1, 2, \ldots\}$ be independent. Then the processes

$$\frac{\sqrt{\nu(\lambda)}}{\log \nu(\lambda)} Z_{\nu(\lambda)}\left(\frac{\log \nu(\lambda)}{\nu(\lambda)\pi(x)}\right), \quad \frac{\sqrt{\lambda}}{\log \lambda} Y_\lambda\left(\frac{\log \lambda}{\lambda\pi(x)}\right)$$

are equivalent (they generate the same probability measure), hence

$$\text{(3.12)} \qquad \mathsf{P}\left(\lim_{\lambda\to\infty} \frac{\sqrt{\nu(\lambda)}}{\log \nu(\lambda)} Z_{\nu(\lambda)}\left(\frac{\log \nu(\lambda)}{\nu(\lambda)\pi(x)}\right) = \frac{x}{\pi(x)}\right) = 1.$$

This implies (3.11), and thus the proof is complete.

Remark. The statement of Theorem 5 remains valid if we use the metric d of the space $C(0,1)$ instead of d_0 in the definition of $\rho(\mathsf{Z}_n, \mathsf{B})$. In fact the two versions of $\rho(\mathsf{Z}_n, \mathsf{B})$ are asymptotically equal.

4. APPLICATIONS

(a) *The Wasserstein metric.* Inequalities (3.1) and (3.5) remain valid if we substitute the Prohorov metric ρ by the Wasserstein metric ρ_0 defined by (1.12). The coupling given by Theorem B depends on the level x, hence it is not applicable for the estimation of the Wasserstein distance. It seems us, however, that there is no difficulty in getting rid of the effect of the level x in the construction, and taking so, to estimate the Wasserstein distance of S_n^F and W if $F \in \mathscr{F}_r$.

(b) *Embedding of functionals.* Let us say that the random variables ξ_n have a limit distribution F of rate $\frac{1}{\sqrt{n}}$ if the limit

$$\lim_{n \to \infty} \sqrt{n}\,(\mathrm{P}(\xi_n < x) - F(x)) = F_1(x)$$

exists for any continuity point of F. All the known functionals on the empirical process $z_n(t)$ defined by (1.3) have a limit distribution of rate $\frac{1}{\sqrt{n}}$, nevertheless there is no general theorem ensuring a limit distribution of rate for a whole class of functionals. As it was stated in [13], Theorem C have a Corollary on functionals fulfilling a Lipschitzian condition, but this Corollary gives only an estimation of rate $\frac{\log n}{\sqrt{n}}$. So it is an open question, whether all the Lipschitzian functionals on $z_n(t)$ have a limit distribution of rate $\frac{1}{\sqrt{n}}$, or not.

(c) *The law of iterated logarithm.* Strassen applied the Skorohod embedding for proving the extension of the law of iterated logarithm. The law of iterated logarithm holds true for any distribution having a finite second moment, hence our improvement of the Skorohod embedding for distributions having higher moments does not yield any improvement of the law of iterated logarithm. This holds true even for the law of iterated logarithm on the empirical process $z_n(t)$ made by Finkelstein [11]. One may expect improvements concerning the tail-behavior of $z_n(t)$, i.e. the investigation of the supremum $\sup\limits_{\epsilon_n < t < 1 - \epsilon_n} \frac{z_n(t)}{\sqrt{t(1-t)}}$. In connection with this problem we refer to the paper of Csörgő and Révész [5]. Another possible application of the new embedding scheme would be the investigation of the law of iterated logarithm in the multidimensional case. This is hampered by the fact that our embedding is extended only to two dimensions yet.

(d) *The estimation of the density function.* Bickel and Rosenblatt [2] extended the heuristic approach of Doob to the Kolmogorov theorem for investigating the estimation of the density function. Our new embedding yields only a minor extension of their theorem, so we do not state it here explicitely. A real extension of their theorem would be the investigation of the multidimensional case, hence again the multidimensional embedding is needed.

(e) *Goodness-of-fit in the presence of nuisance parameters.* Durbin [9] extended the heuristic approach of Doob to the Kolmogorov theorem for investigating the case when some parameters of the distribution were estimated. Csörgő, Révész and the present authors applied Theorem C for this problem in [6].

REFERENCES

[1] P. Bártfai, Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation, *Studia Sci. Math. Hung.*, 1 (1966), 161-168.

[2] P.I. Bickel – M. Rosenblatt, On some global measures of deviations of density function estimates, *Ann. Statist.*, 1 (1973), 1071-1095.

[3] A.A. Borovkov, On the rate of convergence for the invariance principle, (in Russian), *Teorija Verojatn. i Primenen.*, 18 (1973), 217-235.

[4] M. Csörgő – P. Révész, A new method to prove laws of iterated logarithm of Strassen's type, I-II, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, (to appear).

[5] M. Csörgő – P. Révész, Some notes on the empirical distribution function and the quantile process, (this volume).

[6] M. Csörgő – J. Komlós – P. Major – P. Révész – G. Tusnády, On the empirical process when parameters are estimated, *Transactions of the seventh Prague Conference*, 1974, (to appear).

[7] R.M. Dudley, Distances of probability measures and random variables, *Ann. Math. Statist.*, 39 (1968), 1563-1572.

[8] R.M. Dudley, Speeds of metric probability convergence, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 22 (1972), 323-332.

[9] J. Durbin, Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.*, 1 (1973), 279-290.

[10] P. Erdős – A. Rényi, On a new law of large numbers, *Jour. Analyse Mathématique*, 23 (1970), 103-111.

[11] H. Finkelstein, The law of iterated logarithm for empirical distribution, *Ann. Math. Statist.*, 42 (1971), 607-615.

[12] G. Halász – P. Major, On the geyser problem with bounded errors, *Ann. of Prob.*, (to appear).

[13] J. Komlós – P. Major – G. Tusnády, An approximation of partial sums of independent RV'-s, and the sample DF., I-II., *Z. Wahrscheinlichkeitstheorie verw. Geb.*, (to appear).

[14] S. Sawyer, Rates of convergence for some functionals in probability, *Ann. Math. Statist.*, 43 (1972), 273-284.

[15] V. Strassen, Almost sure behaviour of sums of independent random variables and martingales, *Proc. 5th Berkeley Sympos. Math. Statist. Probab.*, II. (part 1), 315-343. Univ. of Calif. Press, Berkeley, 1967.