

A NOTE ON THE APPROXIMATION OF THE UNIFORM EMPIRICAL PROCESS

BY PETER MAJOR

Mathematical Institute of the Hungarian Academy of Sciences

Mason and van Zwet gave an approximation of the uniform empirical process by a Brownian bridge. Their result is a refinement of a result of Komlós, Major and Tusnády in the case when the process is considered in a small interval. In this note we show that in such cases a much better Poissonian approximation is possible which seems to be better applicable in certain cases. We also prove a multidimensional version of this result, where a sequence of uniform empirical processes is simultaneously approximated by partial sums of independent Poisson processes.

1. Introduction. Let $\varepsilon_1, \varepsilon_2, \dots$, be a sequence of independent and on the interval $[0, 1]$ uniformly distributed random variables and define the empirical distribution function $F_n(t)$, $n = 1, 2, \dots$, as $F_n(t) = (1/n)\#\{\varepsilon_j, j \leq n, \varepsilon_j \leq t\}$, $0 \leq t \leq 1$. In this paper we investigate the (uniform) empirical process $\sqrt{n}(F_n(t) - t)$. Mason and van Zwet (1987) proved the following.

THEOREM A. *For all $n \geq 1$ a Brownian bridge $B_n(t)$, $0 \leq t \leq 1$, and a sequence of independent and on the interval $[0, 1]$ uniformly distributed random variables $\varepsilon_1, \varepsilon_2, \dots$ can be constructed in such a way that the empirical distribution function $F_n(t)$ defined with the help of the above ε 's satisfies the relation*

$$(1.1) \quad \mathbf{P} \left(\sup_{0 \leq s \leq d/n} |n(F_n(s) - s) - \sqrt{n} B_n(s)| > C \log d + x \right) < K e^{-\lambda x}$$

with some universal positive constants C, K and λ for all $0 \leq x < \infty, 1 \leq d \leq n$.

This is a refinement of a former result of Komlós, Major and Tusnády (1975), where the same estimate is proved in the special case $d = n$. Theorem A is a real improvement only for $d < n^\varepsilon$, where $\varepsilon > 0$ can be chosen arbitrarily small. For such numbers d the empirical distribution function can be much better approximated in the interval $[0, d/n]$ by a Poisson process. In Theorem 1 formulated below we present such an approximation.

Let $P_n(t)$, $0 \leq t \leq 1$, denote a Poisson process with parameter n , i.e., let $P_n(t)$ be a process with independent stationary increments, $P_n(0) = 0$, and let $P_n(v) - P_n(u)$, $0 \leq u < v \leq 1$, be Poisson distributed with parameter $n(u - v)$.

THEOREM 1. *For all $n \geq 1$ a Poisson process $P_n(t)$ with parameter n on the interval $[0, 1]$ and a sequence of independent and on the interval $[0, 1]$*

Received January 1988; revised December 1988.

AMS 1980 subject classifications. Primary 60F99, 60F17.

Key words and phrases. Empirical process, Poisson approximation, Brownian bridge approximation.

uniformly distributed random variables $\varepsilon_1, \dots, \varepsilon_n$ can be constructed simultaneously in such a way that the empirical distribution function defined with the help of these ε 's satisfies the relation

$$\mathbf{P}\left(\sup_{0 \leq s < n^{-2/3}} |n(F_n(s) - s) - (P_n(s) - ns)| > C\right) < K \exp(-\frac{1}{8}\sqrt{n} \log n)$$

with some universal constants $C > 0$ and $K > 0$.

Theorem 1 states that the empirical process can be approximated by a Poisson process in a small interval in such a way that their difference is bounded by a constant independent of the sample size n with probability very close to 1. We formulate a version of this result which states that this difference can be taken even as zero if we only want that the probability of the exceptional set where this relation does not hold tend to zero as $n \rightarrow \infty$.

THEOREM 1'. *Let $t_n \rightarrow 0$ be such that also $\sqrt{n}t_n \rightarrow 0$. For all $n \geq 1$ a Poisson process $P_n(t)$ with parameter n on the interval $[0, 1]$ and a sequence of independent and on the interval $[0, 1]$ uniformly distributed random variables $\varepsilon_1, \dots, \varepsilon_n$ can be constructed simultaneously in such a way that the empirical distribution function defined with the help of these ε 's satisfies the relation*

$$\mathbf{P}\left(\sup_{0 \leq s < t_n} |n(F_n(s) - s) - (P_n(s) - ns)| = 0\right) \rightarrow 1.$$

As the author learned from an Associate Editor, this result is contained in a work of Horváth (1990) recently submitted to the *Annals* in a more general form. Hence the proof of Theorem 1' is omitted.

In Theorems 1 and 1' we have approximated the empirical process by a Poisson process in a small neighbourhood of zero. This approximation can be combined by a Brownian bridge approximation of the empirical process in the remaining domain with the help of Theorem 3 in Komlós, Major and Tusnády (1975). We formulate this result in the case of Theorem 1' in the following.

PROPOSITION. *In addition to the processes $P_n(t)$ and $F_n(t)$ in Theorem 1' a process $B_n(s)$, $t_n \leq s \leq 1$, can be constructed which is the restriction of a Brownian bridge to the interval $[t_n, 1]$ in such a way that*

$$(1.2) \quad \mathbf{P}\left(\sup_{t_n \leq s \leq 1} |\sqrt{n}B_n(s) - n(F_n(s) - s)| > C \log n + x\right) < Ke^{-\lambda x},$$

where $C > 0$, $K > 0$ and $\lambda > 0$ are appropriate constants and the processes $B_n(s)$, $t_n \leq s \leq 1$, and $P_n(t)$, $0 \leq t \leq t_n$, are conditionally independent with respect to the events $nF_n(t_n) = k$ for all $k = 0, 1, \dots$.

Theorems 1 and 1' can be useful if we want to investigate the limit distribution of a sequence of random variables which is obtained when a sequence of functions \mathcal{F}_n are applied to the empirical processes, i.e., when the functional

may also depend on n . Such a problem is investigated, e.g., in the paper of Csörgő, Csörgő, Horváth and Mason (1986). If the functional \mathcal{F}_n depends only on the values of the empirical process in the interval $[0, t_n]$, then by Theorem 1' the uniform empirical process can be replaced by a standardized Poisson process and the limit distribution of our sequences will be the same after this replacement. A Gaussian approximation does not always work, because the Gaussian approximation of the empirical process is not as good as the Poissonian one. If the functional \mathcal{F}_n depends on the values of the empirical process on the whole interval $[0, 1]$, but this dependence is much stronger on the interval $[0, t_n]$, then Theorem 1' can be applied together with the proposition. Theorem 1 can be useful if we want to get better information about the error committed by the Poissonian approximation.

Another problem where the Gaussian approximation is not always applicable and a Poissonian approximation may be useful is the investigation of the law of iterated logarithms for the empirical process in small intervals. This problem is studied in Kiefer's (1972) paper. If one tries to study this problem in the usual way, then one has to apply the Borel–Cantelli lemma several times, and the main technical problem is to check whether certain sums of probabilities are convergent or divergent. If one substitutes these probabilities by those suggested by the Gaussian approximation one gets in certain cases wrong results, because the error committed by this substitution is too large. On the other hand, Theorem 1 would allow us to make a Poissonian approximation also in such cases. Nevertheless, it is more appropriate to have a result which automatically guarantees that the empirical process can be replaced by a Poisson process in such problems. This is done by Theorem 2 formulated below, or more precisely by its part (b).

THEOREM 2. (a) *For all $n = 1, 2, \dots$, and $\frac{1}{2} \geq \alpha \geq 0$ a sequence $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ of independent and on the interval $[0, 1]$ uniformly distributed random variables can be constructed together with a sequence $P_1(t), P_2(t), \dots, P_n(t)$, of independent Poisson processes with parameter 1 on the interval $[0, 1]$ in such a way that*

$$(1.3) \quad \mathbf{P} \left(\sup_{k \leq n} \sup_{0 \leq t < n^{-(1/2+\alpha)}} \left| k(F_k(t) - t) - \sum_{j=1}^k (P_j(t) - t) \right| > m \right) \leq C(m)n^{-2(m+1)\alpha}$$

for all $m = 0, 1, 2, \dots$, where the constants $C(m)$ depend only on m .

(b) *For any $\delta > 0$ an infinite sequence $\varepsilon_1, \varepsilon_2, \dots$ of independent random variables with uniform distribution on the interval $[0, 1]$ and an infinite sequence of independent Poisson processes $P_1(t), P_2(t), \dots$ with parameter 1 on the interval $[0, 1]$ can be constructed in such a way that*

$$(1.4) \quad \mathbf{P} \left(\sup_n \sup_{0 < t < n^{-1/2-\delta}} \left| n(F_n(t) - t) - \sum_{j=1}^n (P_j(t) - t) \right| < \infty \right) = 1.$$

Part (b) of Theorem 2 implies in particular that the restrictions of the empirical processes to the intervals $[0, t_n]$, $t_n \rightarrow 0$, satisfy the same law of iterated logarithms as the averages $(1/n)\sum_{i=1}^n P_i(t)$ of independent Poisson processes restricted to the same intervals $[0, t_n]$, if $n^{1/2+\delta}t_n \rightarrow 0$ with some $\delta > 0$. Hence the problems investigated by Kiefer can be reformulated to equivalent problems about the sums of independent Poisson processes. The condition $t_n < n^{-1/2-\delta}$ could be slightly weakened, but it is not essential, since it is satisfied in the most interesting and important case, namely when $t_n = [L(n)]/n$ and $L(n)$ is bounded or slowly tends to infinity.

2. The proof of Theorem 1 and the proposition. Let $F(x) = F_\lambda(x)$ denote the (right continuous) Poisson distribution function with parameter λ and $G(x) = G_{n,p}(x)$ the (right continuous) binomial distribution function with parameters n and p . The following Lemma 1 plays an important role in the proof of Theorem 1. In Lemma 1 two random variables are constructed with distributions F and G which are close to each other if the parameters λ , n and p are appropriately chosen. We apply the quantile transformation, i.e., we make the following construction. Let γ be a uniformly distributed random variable on $[0, 1]$ and put

$$(2.1) \quad \xi = k \quad \text{if } F(k) \leq \gamma < F(k + 1), \quad k = 0, 1, \dots,$$

$$(2.1') \quad \eta = k \quad \text{if } G(k) \leq \gamma < G(k + 1), \quad k = 0, 1, \dots$$

Clearly ξ has a distribution F and η a distribution G . We prove the following.

LEMMA 1. *Choose $\lambda = np$ and $p = n^{-2/3}$ for the parameters of the distribution functions F and G . There is a constant $C > 0$ and a threshold n_0 in such a way that for $n > n_0$ the random variables ξ and η defined by (2.1) and (2.1') satisfy the relation*

$$|\xi - \eta| < C \quad \text{if } \xi \leq \sqrt{n}.$$

PROOF. It is enough to show that under the conditions of Lemma 1 there is some $C > 0$ such that

$$(2.2) \quad G(x - C) \leq F(x) \leq G(x + C) \quad \text{if } x \leq \sqrt{n}.$$

Let $f(k)$ and $g(k)$, $k = 0, 1, 2, \dots$, denote the density functions of F and G . We shall prove (2.2) with the help of the following relations. There are some integers $C > 0$ and $n > n_0$ such that if $n > n_0$, then

$$(2.3) \quad 1 - G(\sqrt{n} - C) > 1 - F(\sqrt{n}) > 1 - G(\sqrt{n} + C),$$

$$(2.3') \quad g(k - C) > f(k) > g(k + C) \quad \text{if } np + C < k < \sqrt{n},$$

$$(2.3'') \quad g(k - C) < f(k) < g(k + C) \quad \text{if } 0 \leq k < np - C.$$

[We define $g(k) = 0$ for $k < 0$ in (2.3'').] Relations (2.3)–(2.3'') imply (2.2) (with constant $2C$ instead of C). Indeed, by summing up (2.3) and (2.3') for $j = k, k + 1, \dots, \sqrt{n} - 1$, we get that $1 - G(k - C) > 1 - F(k) > 1 - G(k + C)$ for $np + C < k < \sqrt{n}$ and relation (2.3'') similarly implies that $G(k - C) < F(k) < G(k + C)$ for $0 \leq k \leq np - C$. Finally, for $|k - np| \leq C$ these relations imply that $G(k - 2C) < F(k - C) < F(k) < F(k + C) < G(k + 2C)$.

To prove (2.3) and (2.3') we estimate the ratios $[g(k)]/[f(k)]$ and $[f(k + 1)]/[f(k)]$. Since $\lambda = np$,

$$\begin{aligned} \frac{g(k)}{f(k)} &= \frac{n(n-1) \cdots (n-k+1)}{n^k(1-p)^k} (1-p)^n e^{np} \\ &= \exp \left\{ n(p + \log(1-p)) - k \log(1-p) + \sum_{j=0}^{k-1} \log \left(1 - \frac{j}{n} \right) \right\} \end{aligned}$$

and

$$(2.4) \quad \frac{g(k)}{f(k)} = \exp \left\{ O \left(np^2 + kp + \frac{k^2}{n} \right) \right\} \quad \text{for } k \leq \sqrt{n}.$$

Since $p = n^{-2/3}$ the last relation implies that

$$(2.5) \quad \alpha^{-1} < \frac{g(k)}{f(k)} < \alpha \quad \text{for } k < \sqrt{n}$$

with some $1 < \alpha < \infty$ and

$$(2.5') \quad \left| \frac{g(k)}{f(k)} - 1 \right| < \alpha n^{-1/3} \quad \text{if } |k - np| < \frac{np}{2}.$$

Since

$$(2.6) \quad \frac{f(k+1)}{f(k)} = \frac{\lambda}{k+1},$$

the relations $[f(k+j)]/[f(k)] > (\frac{3}{2})^j$ and $[f(k-j)]/[f(k)] < (\frac{4}{5})^j$ hold if $k > \frac{3}{2}np$ and $0 < j < np/4$. These inequalities together with (2.4) imply (2.3') for $\frac{3}{2}np < k < \sqrt{n}$ with a sufficiently large C which is independent of n . Relation (2.3'') for $k < np/2$ can be proved similarly. Since $\lambda = n^{1/3}$, relation (2.6) also implies that $[f(k-j)]/[f(k)] < (1 - \frac{1}{2}n^{-1/3})^j < 1 - (j/4)n^{-1/3}$ and $[f(k+j)]/[f(k)] > (1 + n^{-1/3})^j > 1 + jn^{-1/3}$ if $k > np + C$ and $j < C$ with a sufficiently large $C > 0$ independent of n . This relation together with (2.5') implies (2.3') for $np + C < k < \frac{3}{2}np$. A similar estimation of $[f(k \pm j)]/[f(k)]$ for $k < np - C$ yields (2.3'') in the remaining domain $np/2 < k < np - C$. To prove (2.3) it is enough to check that $[f(k)]/[1 - F(k)]$ and $[g(k)]/[1 - G(k)]$ are bounded away from zero (and naturally also from infinity) for $k > \sqrt{n} - C$ and then the estimates given on $[f(k \pm j)]/[f(k)]$ and $[g(k)]/[f(k)]$ imply the required inequality. Lemma 1 is proved. \square

PROOF OF THEOREM 1. The argument of the proof is very similar to that of Lemma 1 of Komlós, Major and Tusnády (1976). Let us construct a pair of random variables (ξ, η) with distributions F and G satisfying Lemma 1. Let $\varepsilon'_1, \varepsilon'_2, \dots$ be a sequence of independent uniformly distributed random variables on $[0, n^{-2/3}]$ and $\varepsilon''_1, \varepsilon''_2, \dots$ a sequence of independent uniformly distributed random variables on $[n^{-2/3}, 1]$ such that the pair (ξ, η) and the sequences $\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon''_1, \varepsilon''_2, \dots$ are independent of each other. Put $nF_n(s) = \#\{\varepsilon'_j, \varepsilon'_j < s, j < \eta\}$, $P_n(s) = \#\{\varepsilon'_j, \varepsilon'_j < s, j < \xi\}$ for $s < n^{-2/3}$ and define $n(F_n(s) - F_n(n^{-2/3})) = \#\{\varepsilon''_j, \varepsilon''_j < s, j \leq n - \eta\}$ if $n^{-2/3} < s < 1$

and $P_n(s) - P_n(n^{-2/3})$ as a Poisson process with parameter n on the interval $[n^{-2/3}, 1]$ independent of the above defined process $P_n(s)$, $0 < s < n^{-2/3}$. Then $F_n(s)$ can be considered as an empirical distribution function (corresponding to the sample obtained from the random permutations of $\varepsilon'_1, \dots, \varepsilon'_\eta, \varepsilon''_1, \dots, \varepsilon''_{n-\eta}$), since it has the prescribed conditional distribution under the condition $\eta = k$. Similarly, $P_n(s)$, $0 < s \leq 1$, is a Poisson process with parameter n . On the other hand, $|nF_n(s) - P_n(s)| \leq |\xi - \eta|$ for $0 < s < n^{-2/3}$. Hence by Lemma 1,

$$\begin{aligned} & \mathbf{P}\left(\sup_{0 < s < n^{-2/3}} |n(F_n(s) - s) - (P_n(s) - ns)| > C\right) \\ & < \mathbf{P}(|\xi - \eta| > C) < \mathbf{P}(\xi > \sqrt{n}). \end{aligned}$$

Since

$$\begin{aligned} \mathbf{P}(\xi > \sqrt{n}) & < \text{const.} \mathbf{P}(\xi = \sqrt{n}) \\ & = \text{const.} \exp\left\{-n^{1/3} + \frac{1}{3}\sqrt{n} \log n - \log(\sqrt{n}!)\right\} \\ & \leq K \exp\left\{-\frac{1}{8}\sqrt{n} \log n\right\}, \end{aligned}$$

the above estimates imply Theorem 1. \square

PROOF OF THE PROPOSITION. The conditional distribution of the process $n[(F_n(s) - s) - (F_n(t_n) - t_n)]$, $t_n \leq s \leq 1$, under the condition $nF_n(t_n) = k$ agrees with the distribution of an empirical distribution function multiplied by $n - k$ from a sample of $n - k$ elements with uniform distribution on the interval $[t_n, 1]$. Hence we can construct, with the help of Theorem 3 of Komlós, Major and Tusnády (1975), a Brownian bridge $\bar{B}_n(u)$, $0 \leq u \leq 1$, on the conditional probability space $\mathbf{P}(\cdot | nF_n(t_n) = k)$ such that

$$\begin{aligned} & \mathbf{P}\left(\sup_{t_n \leq s \leq 1} \left|n[(F_n(s) - F_n(t_n)) - (s - t_n)] - \sqrt{n - k} B_n\left(\frac{s - t_n}{1 - t_n}\right)\right| \right. \\ (2.7) \quad & \left. > C \log n + x | nF_n(t_n) = k \right) \leq Ke^{-\lambda x}. \end{aligned}$$

Since the distribution of $\bar{B}_n(\cdot)$ does not depend on k it is a Brownian bridge also with respect to the original unconditional probability.

By Lemma 2 of Komlós, Major and Tusnády (1975) a standard Gaussian random variable ξ_n can be constructed with the help of the quantile transformation in such a way that

$$(2.8) \quad \mathbf{P}\left(|\sqrt{nt_n(1 - t_n)} \xi_n - n(F_n(t_n) - t_n)| > C\right) \leq Ke^{-\lambda x}.$$

Since $F_n(t_n)$ is independent of $\bar{B}_n(\cdot)$ the random variable ξ_n , which is obtained as a transformation of the random variable $F_n(t_n) - t_n$ with some randomization, is also independent of it. Define

$$B_n(s) = \sqrt{1 - t_n} \bar{B}_n\left(\frac{s - t_n}{1 - t_n}\right) + \frac{1 - s}{1 - t_n} \sqrt{t_n(1 - t_n)} \xi_n, \quad t_n \leq s \leq 1.$$

We claim that the above defined process $B_n(s)$ satisfies the proposition. It is a Gaussian process and simple calculation shows [by using the independence of $\bar{B}_n(\cdot)$ and ξ_n] that it has the covariance function $s(1 - s')$ for $t_n \leq s \leq s' \leq 1$. The processes $B_n(\cdot)$ and $P_n(\cdot)$ are conditionally independent with respect to the condition $F_n(t_n) = k$. We have to prove relation (1.2).

We remark first that (2.7) also implies that

$$(2.9) \quad \mathbf{P} \left(\sup_{t_n \leq s \leq 1} \left| n[(F_n(s) - F_n(t_n)) - (s - t_n)] - \sqrt{n(1 - F_n(t_n))} \bar{B}_n \left(\frac{s - t_n}{1 - t_n} \right) \right| > C \log n + x \right) \leq Ke^{-\lambda x}.$$

Moreover, we claim that

$$(2.9') \quad \mathbf{P} \left(\sup_{t_n \leq s \leq 1} \left| n[(F_n(s) - F_n(t_n)) - (s - t_n)] - \sqrt{n(1 - t_n)} \bar{B}_n \left(\frac{s - t_n}{1 - t_n} \right) \right| > C \log n + x \right) \leq Ke^{-\lambda x}$$

(with possibly different constants C , K and λ). To prove (2.9') we have to show that a negligible error is committed if the coefficient of $B(\cdot)$ in (2.9), $\sqrt{n(1 - F_n(t_n))}$, is replaced by $\sqrt{n(1 - t_n)}$. This follows from the estimate

$$\begin{aligned} & \mathbf{P} \left(\sup_{t_n \leq s \leq 1} \left| (\sqrt{n(1 - t_n)} - \sqrt{n(1 - F_n(t_n))}) \bar{B}_n \left(\frac{s - t_n}{1 - t_n} \right) \right| > x \right) \\ & \leq \mathbf{P}(|\sqrt{n(1 - t_n)} - \sqrt{n(1 - F_n(t_n))}| > \sqrt{x}) + \mathbf{P} \left(\sup_{0 \leq s \leq 1} |\bar{B}_n(s)| > \sqrt{x} \right) \\ & \leq \mathbf{P} \left(\sqrt{n} |F_n(t_n) - t_n| > \frac{1}{2} \sqrt{x} \right) + K \exp \left(-\frac{x}{2} \right) \leq Ke^{-\lambda x}. \end{aligned}$$

Now we can write for $t_n \leq s \leq 1$,

$$\begin{aligned} & \sqrt{n} B_n(s) - n(F_n(s) - s) \\ & = \frac{1 - s}{1 - t_n} \left\{ \sqrt{nt_n(1 - t_n)} \xi_n - n(F_n(t_n) - t_n) \right\} \\ & \quad + \left\{ \sqrt{n(1 - t_n)} \bar{B}_n \left(\frac{s - t_n}{1 - t_n} \right) - n[F_n(s) - F_n(t_n) - (s - t_n)] \right\} \\ & = \frac{1 - s}{1 - t_0} I_1(t_n) + I_2(s, t_n). \end{aligned}$$

Since the term $I_1(t_n)$ is bounded in (2.8) and the term $I_2(s, t_n)$ in (2.9'), the last identity implies (1.2). The proposition is proved. \square

3. The proof of Theorem 2. First we formulate a lemma which is proved similarly to Theorem 1.

LEMMA 2. *Given some positive integer n and $\beta > 0$, let ξ_1, \dots, ξ_n be independent Poisson distributed random variables with parameters $n^{-\beta}$ and η_1, \dots, η_n independent random variables such that $\mathbf{P}(\eta_j = 1) = 1 - \mathbf{P}(\eta_j = 0) = n^{-\beta}$, $j = 1, \dots, n$. A sequence $\varepsilon_1, \dots, \varepsilon_n$ of independent, uniformly distributed random variables on $[0, 1]$ and a sequence $P_1(t), \dots, P_n(t)$, $0 < t < 1$, of independent Poisson processes with parameter 1 can be constructed in such a way that*

$$(3.1) \quad \begin{aligned} & \sup_{1 \leq k \leq n} \sup_{0 \leq s \leq n^{-\beta}} \left| k(F_k(s) - s) - \sum_{j=1}^k (P_j(s) - s) \right| \\ & = \sup_{1 \leq k \leq n} \left| \sum_{j=1}^k (\xi_j - \eta_j) \right|, \end{aligned}$$

where the functions $F_k(\cdot)$ are the empirical distribution functions defined with the help of the above ε 's.

Lemma 2 enables us to reduce part (a) of Theorem 2 to a simpler problem. Namely, if we have a sequence $\varepsilon_1, \dots, \varepsilon_n$ of independent uniformly distributed random variables on $[0, 1]$ and a sequence $P_1(t), \dots, P_n(t)$, $0 \leq t \leq 1$, of independent Poisson processes with parameter 1 define $\xi_j = P_j(n^{-(1/2+\alpha)})$ and $\eta_j = jF_j(n^{-(1/2+\alpha)}) - (j-1)F_{j-1}(n^{-(1/2+\alpha)})$, $j = 1, \dots, n$, where F_j , $j = 1, \dots, n$, are the empirical distribution functions defined with the help of the sample $\varepsilon_1, \dots, \varepsilon_j$. Then Lemma 2 implies that a new sequence of independent uniformly distributed random variables on $[0, 1]$, $\bar{\varepsilon}_j$, $j = 1, \dots, n$, and a new sequence of independent Poisson processes with parameter 1, $\bar{P}_j(t)$, $j = 1, \dots, n$ can be constructed in such a way that

$$\begin{aligned} & \sup_{1 \leq k \leq n} \sup_{0 < s < n^{-(1/2+\alpha)}} \left| k(\bar{F}_k(s) - s) - \sum_{j=1}^k (\bar{P}_j(s) - s) \right| \\ & = \sup_{1 \leq k \leq n} \left| k(F_k(n^{-(1/2+\alpha)}) - n^{-(1/2+\alpha)}) \right. \\ & \quad \left. - \sum_{j=1}^k (P_j(n^{-(1/2+\alpha)}) - n^{-(1/2+\alpha)}) \right|. \end{aligned}$$

This relation enables us to drop $\sup_{0 < t < n^{-(1/2+\alpha)}}$ from (1.3) and to consider only the argument $t = n^{-(1/2+\alpha)}$ instead.

PROOF OF LEMMA 2. Let $\varepsilon'_1, \varepsilon'_2, \dots$, be a sequence of independent uniformly distributed random variables on $[0, n^{-\beta}]$ which is independent of the random variables ξ and η , put $S_k = \sum_{j=1}^k \xi_j$ and $T_k = \sum_{j=1}^k \eta_j$, $k = 1, \dots, n$. If $\eta_k = 1$ define $\varepsilon_k = \varepsilon'_l$ with $l = t_k$. Define the Poisson process $P_k(t)$ in the interval

$[0, n^{-\beta}]$ in the following way: It has jumps in the points $\varepsilon'_j, \dots, \varepsilon'_{j+p}$ with $j = S_{k-1} + 1, j + p = S_k$. (If $S_{k-1} + 1 > S_k$, then it has no jumps in this interval.) Define the Poisson processes $P_k(t), 0 < t < 1$, in such a way that on the interval $[0, n^{-\beta}]$ they agree with the already defined Poisson processes and the processes $P_k(t) - P_k(n^{-\beta}), n^{-\beta} < t < 1, k = 1, 2, \dots$, are independent of the processes $P_k(t), 0 < t < n^{-\beta}$, and of each other. Otherwise they are arbitrarily defined. Finally, let $\varepsilon'_1, \varepsilon'_2, \dots$ be a sequence of independent uniformly distributed random variables on the interval $[n^{-\beta}, 1]$, which is independent both of the sequence $\varepsilon'_1, \varepsilon'_2, \dots$ and the η 's, and define $\varepsilon_k = \varepsilon'_k$ if $\eta_k = 0$. Then the sequences $\varepsilon_1, \dots, \varepsilon_n$ and $P_1(t), \dots, P_n(t)$ have the prescribed distributions. On the other hand, relation (3.1) holds, since for all $1 \leq k \leq n$,

$$\sup_{0 \leq s < n^{-\beta}} \left| k(F_k(s) - s) - \sum_{j=1}^k (P_j(s) - s) \right| = |T_k - S_k|. \quad \square$$

PROOF OF THEOREM 2. (a) By Lemma 2 it is enough to construct a sequence of independent Poisson distributed random variables ξ_1, \dots, ξ_n with parameter $n^{-(1/2+\alpha)}$ and a sequence of independent random variables $\eta_1, \dots, \eta_n, \mathbf{P}(\eta_j = 1) = 1 - \mathbf{P}(\eta_j = 0) = n^{-(1/2+\alpha)}$ in such a way that

$$(3.2) \quad \mathbf{P} \left(\sup_{1 \leq k \leq n} \left| \sum_{j=1}^k (\xi_j - \eta_j) \right| > m \right) \leq C(m)n^{-2(m+1)\alpha}$$

for all $m = 0, 1, \dots$. Let us consider two independent Poisson processes $P_1(t)$ and $P_2(t)$ with parameter 1 and define $\xi_j = P_1(jn^{-(1/2+\alpha)}) - P_1((j-1)n^{-(1/2+\alpha)}), j = 1, 2, \dots, n$. To define η_j let us first introduce the random variables $\xi'_j = P_2(js) - P_2((j-1)s), j = 1, \dots, n$, where $e^{-s} = (1 - n^{-(1/2+\alpha)})\exp(n^{-(1/2+\alpha)})$. Let $\eta_j = 1$ if $\xi_j + \xi'_j \geq 1$ and zero otherwise. Then both sequences ξ_1, \dots, ξ_n and η_1, \dots, η_n consist of independent random variables with the right distribution. [Observe that $\mathbf{P}(\eta_j = 0) = \mathbf{P}(\xi_j = 0, \xi'_j = 0) = \exp(-n^{-(1/2+\alpha)} - s) = 1 - n^{-(1/2+\alpha)}$.] It remains to show that the above defined ξ_j 's and η_j 's satisfy (3.2). Observe that $\sum_{j=1}^k (\eta_j - \xi_j) \leq \sum_{j=1}^k \xi'_j$ and $\sum_{j=1}^k (\xi_j - \eta_j) \leq \sum_{j=1}^k \bar{\xi}_j$ with $\bar{\xi}_j = \xi_j - 1$ if $\xi_j > 0$ and $\bar{\xi}_j = 0$ if $\xi_j = 0$. Hence

$$(3.3) \quad \mathbf{P} \left(\sup_{1 \leq k \leq n} \sum_{j=1}^k (\eta_j - \xi_j) > m \right) \leq \mathbf{P} \left(\sum_{j=1}^n \xi'_j \geq m + 1 \right) \leq C(m)n^{-2(m+1)\alpha}$$

(here we use the fact that $\sum_{j=1}^n \xi'_j$ is a Poisson distributed random variable with parameter $ns \leq n^{-2\alpha}$) and

$$(3.4) \quad \begin{aligned} \mathbf{P} \left(\sup_{1 \leq k \leq n} \sum_{j=1}^k (\xi_j - \eta_j) > m \right) &\leq \mathbf{P} \left(\sum_{j=1}^n \bar{\xi}_j \geq m + 1 \right) \\ &\leq \exp(-2\alpha(m+1)\log n) E \exp \left(2\alpha \log n \sum_{j=1}^n \bar{\xi}_j \right) \\ &= (E \exp(2\alpha \bar{\xi}_1 \log n))^n n^{-2(m+1)\alpha}. \end{aligned}$$

We claim that $E \exp(2\alpha\bar{\xi}_1 \log n) \leq 1 + C/n$ with some $C > 0$ independent of n . This inequality together with (3.4) and (3.3) imply (3.2) and hence also part (a) of Theorem 2. We have

$$\begin{aligned} E \exp(2\alpha\bar{\xi}_1 \log n) &\leq 1 + \sum_{j=1}^{\infty} \frac{\exp(-n^{-(1/2+\alpha)} + 2j\alpha \log n)}{(j+1)!} n^{-(j+1)(\alpha+1/2)} \\ &\leq 1 + \sum_{j=1}^{\infty} \frac{1}{(j+1)!} n^{(j-1)(\alpha-1/2)-1} \leq 1 + \frac{C}{n} \end{aligned}$$

for $0 \leq \alpha \leq \frac{1}{2}$.

(b) Let us fix some $\frac{1}{2} > \delta > 0$ and apply part (a) with $\alpha = \delta$ and 2^n , $n = 0, 1, \dots$. Because of formula (1.3) in the special case $m = 0$ we can construct a sequence of independent Poisson processes $P_1^{(n)}(t), \dots, P_{2^n}^{(n)}(t)$ with parameter 1 and a sequence of independent uniformly distributed random variables on $[0, 1]$, $\varepsilon_1^{(n)}, \dots, \varepsilon_{2^n}^{(n)}$, in such a way that

$$(3.5) \quad \mathbf{P} \left(\sum_{j=1}^k P_j^{(n)}(t) = kF_k^{(n)}(t) \text{ for } 0 \leq t < 2^{-(1/2+\delta)n} \text{ and all } k = 1, \dots, 2^n \right) \leq C2^{-2n\delta}.$$

(The upper index in $F_k^{(n)}$ denotes that this empirical distribution function is constructed with the help of the sample $\varepsilon_1^{(n)}, \dots, \varepsilon_{2^n}^{(n)}$.) We may assume that the above defined sequences $\varepsilon_j^{(n)}$ and $P_j^{(n)}$ are independent for different n . Define $P_j(t) = P_{j+1-2^n}^{(n)}(t)$ and $\varepsilon_j = \varepsilon_{j+1-2^n}^{(n)}$ for $2^n \leq j < 2^{n+1}$, $n = 0, 1, 2, \dots$. Then relation (3.5) implies that the events A_n ,

$$A_n = \left\{ \sum_{j=2^n}^k \mathbf{P}_j(t) = kF_k(t) - (2^n - 1)F_{2^n-1}(t) \right. \\ \left. \text{for all } 0 \leq t \leq 2^{-n(1/2+\delta)} \text{ and } 2^n \leq k < 2^{n+1} \right\},$$

satisfy the relation

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty.$$

Hence by the Borel–Cantelli lemma there is a random threshold $n(w)$ such that for $n > n(w)$,

$$\sum_{j=2^n}^k P_j(t) - t = k(F_k(t) - t) - (2^n - 1)[F_{2^n-1}(t) - t]$$

$$\text{for all } 2^n \leq k < 2^{n+1} \text{ and } t < k^{(-1/2+\delta)}$$

with probability 1. The last relation implies part (b) of Theorem 2. \square

REFERENCES

- CsÖRGÖ, M., CsÖRGÖ, S., HORVÁTH, L. and MASON, D. M. (1986). Normal and stable convergence of integral functions of the empirical distribution functions. *Ann. Probab.* **14** 86–118.
- KIEFER, J. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 227–244. Univ. California Press.
- 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. Wahrsch. 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. Wahrsch. verw. Gebiete* **34** 33–58.
- MASON, D. M. and VAN ZWET, W. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.

MATHEMATICAL INSTITUTE OF THE
HUNGARIAN ACADEMY OF SCIENCES
P.O.B. 127
BUDAPEST H-1364
HUNGARY