A NOTE ON THE APPROXIMATION OF THE UNIFORM EMPIRICAL PROCESS

Péter Major

Mathematical Institute of the Hungarian Academy of Sciences

Summary: D. Mason and W. van Zwet (1987) gave an approximation of the uniform empirical process by a Brownian bridge which is a refinement of a result of Komlós–Major–Tusnády (1975). Their result is an improvement only if 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, on the interval $[0,1]$ uniformly distributed random variables, and define the empirical distribution function $F_n(t)$, $n=1,2,\dots,$ as $F_n(t)=\frac{1}{n}\#\{\varepsilon_j,\ j\le n,\ \varepsilon_j\le t\}$, $0\le t\le 1$. In this paper we investigate the (uniform) empirical process $\sqrt{n}(F_n(t)-t)$. Mason and Zwet (1987) proved the following

Theorem A. *For all $n\ge 1$ a Brownian bridge $B_n(t)$, $0\le t\le 1$, 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*

$$\mathbf{P}\left(\sup_{0\le s\le \frac{d}{n}} |n(F_n(s)-s)-\sqrt{n}B_n(s)|>C\log d+x\right)<Ke^{-\lambda x} \tag{1.1}$$

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

This is a refinement of a former result of Komlós, Major and Tusnády (KMT) (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 $P_n(t)$, $0\le t\le 1$, 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 $P_n(v)-P_n(u)$, $0\le u<v\le 1$, be Poisson distributed with parameter $n(u-v)$.

Typeset by $\mathcal{AMS}$-TEX

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

$$\mathbf{P}\left(\sup_{0\le s<n^{-2/3}} |n(F_n(s)-s)-(P_n(s)-ns)|>C\right) < K\exp\left(-\tfrac18\sqrt{n}\log n\right)$$

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

We shall also prove the following variant of Theorem 1.

Theorem 1′. *Let $t_n \to 0$ be such that also $\sqrt{n}t_n \to 0$. For all $n \ge 1$ a Poisson process $P_n(t)$ with parameter n on the interval $[0,1]$ and a sequence of independent, 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*

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

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 the KMT (1975) paper. We formulate this result in the case of Theorem 1′ in the following

Proposition. *Beside the processes $P_n(t)$ and $F_n(t)$ in Theorem 1′ a process $B_n(s)$, $t_n \le s \le 1$, can be constructed which is the restriction of a Brownian bridge to the interval $[t_n,1]$ in such a way that*

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

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

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 functionals $\mathcal{F}_n$ 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 M. Csörgő *et al.* (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 is not always allowed, because the Gaussian approximation of the empirical process is not so 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 a 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 logarithm 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 result, 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. *Part a) For all* $n = 1, 2, \dots,$ *and* $1/2 \ge \alpha \ge 0$ *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*

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

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

Part 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*

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

Part b) of Theorem 2 implies in particular that the restrictions of the empirical processes to the intervals $[0, t_n]$, $t_n \to 0$, satisfy the same laws of iterated logarithm as the averages $\frac{1}{n}\sum_{i=1}^{n} P_i(t)$ of independent Poisson processes restricted to the same intervals $[0, t_n]$, if $n^{1/2+\delta}t_n \to 0$ 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 $t_n = \frac{L(n)}{n}$, and $L(n)$ is bounded, or slowly tends to infinity.

2. The proof of Theorem 1, 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

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

$$\eta = k \qquad \text{if} \ \ G(k) \le \gamma < G(k+1),\ k = 0, 1, \dots. \tag{2.1$'$}$$

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*

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

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

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

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

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

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

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

(We define $g(k) = 0$ for $k < 0$ in (2.3$''$).) Relations 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 $0 \le k \le np - C$. Finally, for $|k - np| \le C$ 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 $\frac{g(k)}{f(k)}$ and $\frac{f(k+1)}{f(k)}$. Since $\lambda = np$,

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

and

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

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

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

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

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

Since

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

the relations $\frac{f(k+j)}{f(k)} > (3/2)^j$ and $\frac{f(k-j)}{f(k)} < (4/5)^j$ 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 $\frac{f(k-j)}{f(k)} < (1-\frac{1}{2}n^{-1/3})^j < 1-\frac{j}{4}n^{-1/3}$ and $\frac{f(k+j)}{f(k)} > (1+n^{-1/3})^j > 1+jn^{-1/3}$ if $k > np+C$ and $j < C$ with a sufficiently large $C > 0$ independent of n. This relation together with (2.5$'$) imply (2.3$'$) for $np+C < k < \frac{3}{2}np$. A similar estimation of $\frac{f(k\pm j)}{f(k)}$ 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 $\frac{f(k)}{1-F(k)}$ and $\frac{g(k)}{1-G(k)}$ are bounded away from zero (and naturally also from infinity) for $k > \sqrt{n}-C$, and then the estimates given on $\frac{f(k\pm j)}{f(k)}$ and $\frac{g(k)}{f(k)}$ imply the required inequality. Lemma 1 is proved.

Proof of Theorem 1. 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 $[0, n^{-2/3}]$, $\varepsilon_1'', \varepsilon_2'', \dots$ 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 $n(F_n(s) - F_n(n^{-2/3})) = \#\{\varepsilon_j'',\ \varepsilon_j'' < s,\ j \le n - \eta\}$ if $n^{-2/3} < s < n - \eta$, 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, $P_n(s)$, $0 < s \le 1$, is a Poisson process with parameter n. On the other hand $|nF_n(s) - P_n(s)| \le |\xi - \eta|$ for $0 < s < n^{-2/3}$. Hence by Lemma 1

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

Since

$$\mathbf{P}(\xi > \sqrt{n}) \le const.\mathbf{P}(\xi = \sqrt{n}) \\ = const.\exp\left\{-n^{1/3} + \tfrac{1}{3}\sqrt{n}\log n - \log(\sqrt{n}!)\right\} \le Kexp\left\{-\tfrac{1}{8}\sqrt{n}\log n\right\}$$

the above estimates imply Theorem 1.

Proof of Theorem $1'$. We may assume without violating the generality that $nt_n \to \infty$. The proof of Theorem $1'$ is very similar to that of Theorem 1. The main difference is that now we need for each n, instead of the construction of Lemma 1, a pair of random variables (ξ_n, η_n) with distributions $F_\lambda(x)$ and $G_{n,p}(x)$, $\lambda = np_n$ and $p = p_n$ such that

$$\mathbf{P}(\xi_n = \eta_n) \to 1 \qquad \text{as } n \to \infty\,. \tag{2.7}$$

Then we can apply the same costruction as in the proof of Theorem 1, only now we have to work with the above pair (ξ_n, η_n), and the intervals $[0, n^{-2/3}]$ and $[n^{-2/3}, 1]$ must be replaced by the intervals $[0, t_n]$ and $[t_n, 1]$. So it remains to prove relation (2.7).

We claim that

$$\sum_{k=0}^{\infty} |\mathbf{P}(\xi_n = k) - \mathbf{P}(\eta_n = k)| \to 0 \tag{2.8}$$

First we show that relation (2.7) can be satisfied with an appropriate construction because of (2.8). Indeed, define the pair (ξ_n, η_n) by the following formulas:

$$\mathbf{P}(\xi_n = \eta_n = k) = \min\{\mathbf{P}(\xi_n = k), \mathbf{P}(\eta_n = k)\}$$

and

$$\mathbf{P}(\xi_n = k, \eta_n = j) \\ = B_n^{-1}\big(\mathbf{P}(\xi_n = k) - \min\{\mathbf{P}(\xi_n = k), \mathbf{P}(\eta_n = k)\}\big)\big(\mathbf{P}(\eta_n = j) \\ - \min\{\mathbf{P}(\xi_n = j), \mathbf{P}(\eta_n = j)\}\big) \quad \text{if } j \ne k\,,$$

where $B_n = 1 - \sum_{k=0}^{\infty} \min\{\mathbf{P}(\xi_n = k), \mathbf{P}(\eta_n = k)\}$. (If $B_n = 0$ then we have $\frac{0}{0}$ in the last formula which we define as 0.) Then the random variables ξ_n and η_n have the prescribed distribution, since

$$\mathbf{P}(\xi_n = k, \eta_n \ne k) = \mathbf{P}(\xi_n = k) - \min\{\mathbf{P}(\xi_n = k), \mathbf{P}(\eta_n = k)\}\,,$$

and a similar relation holds also for η_n. It follows from (2.8) that $B_n \to 0$, hence (2.7) also holds.

To prove (2.8) observe first that similarly to (2.4) we have

$$\frac{\mathbf{P}(\xi_n = k)}{\mathbf{P}(\eta_n = k)} = \exp\left\{O\left(np_n^2 + kp_n + \frac{k^2}{n}\right)\right\} = \exp\{O(np_n^2)\} \qquad \text{if } k < 2np_n\,. \tag{2.9}$$

Hence

$$\sum_{k=0}^{2np_n} |\mathbf{P}(\xi_n = k) - \mathbf{P}(\eta_n = k)| \le \left\{\exp\{O(np_n^2)\} - 1\right\} \sum \mathbf{P}(\eta_n = k) = O(np_n^2)\ .$$

On the other hand

$$\sum_{k=2np_n}^{\infty} |\mathbf{P}(\xi_n = k) - \mathbf{P}(\eta_n = k)| \le \mathbf{P}(\xi_n \ge 2np_n) + \mathbf{P}(\eta_n \ge 2np_n) \to 0 \qquad \text{as } n \to \infty\ .$$

These relations imply (2.8). Theorem 1′ is proved.

Proof of the Proposition. The conditional distribution of the process $n[(F_n(s) - s) - (F_n(t_n) - t_n)]$, $t_n \le s \le 1$ 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 KMT (1975), a Brownian bridge $\bar{B}_n(u)$, $0 \le u \le 1$, on the conditional probability space $\mathbf{P}(\cdot\,|nF_n(t_n) = k)$ such that

$$\mathbf{P}\left(\sup_{t_n \le s \le 1}\left|n[(F_n(s) - F_n(t_n)) - (s - t_n)] - \sqrt{n-k}\,\bar{B}_n\left(\frac{s-t_n}{1-t_n}\right)\right| > C\log n + x \,\middle|\, nF_n(t_n) = k\right) \le Ke^{-\lambda x}\ . \tag{2.10}$$

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 KMT (1975) a standard Gaussian random variable ξ_n can be constructed with the help of the quantile transformation in such a way that

$$\mathbf{P}\left(\left|\sqrt{nt_n(1-t_n)}\xi_n - n(F_n(t_n) - t_n)\right| > C\right) \le Ke^{-\lambda x}\ . \tag{2.11}$$

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

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

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) that it has the covariance function $s(1-s')$ for $t_n \le s \le s' \le 1$. 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).

Remark first that (2.10) also implies that

$$\mathbf{P}\left(\sup_{t_n \le s \le 1}\left|n[(F_n(s) - F_n(t_n)) - (s - t_n)] - \sqrt{n(1-F_n(t_n))}\,\bar{B}_n\left(\frac{s-t_n}{1-t_n}\right)\right| > C\log n + x\right) \le Ke^{-\lambda x}\ . \tag{2.12}$$

Moreover, we claim that
(2.12′)

$$\mathbf{P}\left(\sup_{t_n\le s\le 1}\left|n[(F_n(s)-F_n(t_n))-(s-t_n)]-\sqrt{n(1-t_n)}\bar B_n\left(\frac{s-t_n}{1-t_n}\right)\right|>C\log n+x\right)\le Ke^{-\lambda x}\,,$$

(with possibly different constants C, K and λ.) To prove (2.12′) we have to show that a negligible error is committed if the coefficient of $\bar B(\cdot)$ in (2.12) $\sqrt{n(1-F_n(t_n))}$ is replaced by $\sqrt{n(1-t_n)}$. This follows from the following estimate:

$$\begin{aligned}&\mathbf{P}\left(\sup_{t_n\le s\le 1}\left|\left(\sqrt{n(1-t_n)}-\sqrt{n(1-F_n(t_n))}\right)\bar B_n\left(\frac{s-t_n}{1-t_n}\right)\right|>x\right)\\&\quad\le \mathbf{P}\left(\left|\sqrt{n(1-t_n)}-\sqrt{n(1-F_n(t_n))}\right|>\sqrt{x}\right)+\mathbf{P}\left(\sup_{0\le s\le 1}|\bar B_n(s)|>\sqrt{x}\right)\\&\quad\le \mathbf{P}\left(\sqrt{n}|F_n(t_n)-t_n|>\tfrac12\sqrt{x}\right)+K\exp\left(-\tfrac{x}{2}\right)\le Ke^{-\lambda x}\,.\end{aligned}$$

Now we can write for $t_n\le s\le 1$

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

Since the term $I_1(t_n)$ is bounded in (2.11) and the term $I_2(s,t_n)$ in (2.12′) the last identity implies (1.2). The Proposition is proved.

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 $\xi_1,\dots,\xi_n$ be independent Poisson distributed random variables with parameters $n^{-\beta}$ and $\eta_1,\dots,\eta_n$ 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*

$$\sup_{1\le k\le n}\sup_{0\le s\le n^{-\beta}}\left|k(F_k(s)-s)-\sum_{j=1}^k(P_j(s)-s)\right|=\sup_{1\le k\le n}\left|\sum_{j=1}^k(\xi_j-\eta_j)\right|\,, \tag{3.1}$$

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 $P_1(t),\dots,P_n(t)$, $0\le t\le 1$ 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

$$\sup_{1\le k\le n}\sup_{0<s<n^{-(1/2+\alpha)}}\left|k(\bar F_k(s)-s)-\sum_{j=1}^k(\bar P_j(s)-s)\right|$$
$$=\sup_{1\le k\le n}\left|k(F_k(n^{-(1/2+\alpha)})-n^{-(1/2+\alpha)})-\sum_{j=1}^k\left(P_j(n^{-(1/2+\alpha)})-n^{-(1/2+\alpha)}\right)\right|.$$

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-s and η-s, 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 $1\le k\le n$

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

Proof of Theorem 2. Part a.) By Lemma 2 it is enough to construct a sequence of independent Poisson distributed random variables $\xi_1,\dots,\xi_n$ 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

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

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 $\xi_j+\xi'_j\ge 1$, and zero otherwise. Then both sequences $\xi_1,\dots,\xi_n$ and $\eta_1,\dots,\eta_n$ 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 $\sum_{j=1}^k(\eta_j-\xi_j)\le\sum_{j=1}^k\xi'_j$, and $\sum_{j=1}^k(\xi_j-\eta_j)\le\sum_{j=1}^k\bar\xi_j$ with $\bar\xi_j=\xi_j-1$ if $\xi_j>0$, and $\bar\xi_j=0$ if $\xi_j=0$. Hence

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

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

$$\begin{aligned}\mathbf{P}\left(\sup_{1\le k\le n}\sum_{j=1}^k(\xi_j-\eta_j)>m\right)&\le\mathbf{P}\left(\sum_{j=1}^n\bar\xi_j\ge m+1\right)\\&\le\exp(-2\alpha(m+1)\log n)E\exp\left(2\alpha\log n\sum_{j=1}^n\bar\xi_j\right)\\&=(E\exp(2\alpha\bar\xi_1\log n))^n n^{-2(m+1)\alpha}.\end{aligned}\tag{3.4}$$

We claim that $E\exp(2\alpha\bar\xi_1\log n)\le 1+\frac{C}{n}$ 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

$$\begin{aligned}E\exp(2\alpha\bar\xi_1\log n)&\le 1+\sum_{j=1}^\infty\frac{\exp(-n^{-(1/2+\alpha)}+2j\alpha\log n)}{(j+1)!}n^{-(j+1)(\alpha+1/2)}\\&\le 1+\sum_{j=1}^\infty\frac{1}{(j+1)!}n^{(j-1)(\alpha-1/2)-1}\le 1+\frac{C}{n}\end{aligned}$$

for $0\le\alpha\le 1/2$.

Proof of Part b.) Let us fix some $1/2>\delta>0$, 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

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

(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

$\varepsilon_j = \varepsilon^{(n)}_{j+1-2^n}$ for $2^n \le j < 2^{n+1}$, $n = 0, 1, 2, \dots$. Then relation (3.5) implies that the events A_n

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

satisfy the relation

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

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

$$\sum_{j=2^n}^{k} P_j(t) - t = k(F_k(t) - t) - (2^n - 1)[F_{2^n-1}(t) - t] \qquad \text{for all } 2^n \le k < 2^{n+1} \text{ and } t < k^{(-1/2+\delta)}$$

with probability 1. The last relation implies Part b) of Theorem 2.

References:

Kiefer, J. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. *Proc. 6.th Berkeley Sympos. Math. Statist. Probab.* Univ. of California Press 1, 227–244

Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahr. verw. Gebiete* **34** 33–58

Csörgő, M., Csörgő, S., Horváth, L., Mason, D.M. (1986). Normal and stable convergence of integral functions of the empirical distribution functions. *Ann. Probab.* **14** 86–118

Mason, D.M., van Zwet, W. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884