A GLIMPSE OF THE KMT (1975) APPROXIMATION OF EMPIRICAL PROCESSES BY BROWNIAN BRIDGES VIA QUANTILES

by Miklós Csörgő [1]

Carleton University

running head: Empiricals Approximated Via Quantiles

We deduce a partial version of the KMT (1975) inequality for coupling the uniform empirical process with a sequence of Brownian bridges via the construction used by Csörgő and Révész (CsR) (1978) for their similar coupling of the uniform quantile process with another sequence of Brownian bridges. These constructions are pivoted on the KMT (1975, 1976) inequalities for approximating partial sums by a Wiener process (Brownian motion).

1. Introduction and results.

Let $U_1, U_2, \cdots,$ be independent uniform $(0,1)$ random variables (r.v.'s). For each integer $n \geq 1$, define

$$
\begin{aligned}
(1.1)\qquad G_n(t) &:= n^{-1} \sum_{i=1}^{n} \mathbb{1}\{U_i \leq t\}, \quad 0 \leq t \leq 1, \\
&= \begin{cases} 0, & \text{if } \ 0 \leq t < U_{1,n}, \\ k/n, & \text{if } \ U_{k,n} \leq t < U_{k+1,n},\ 1 \leq k \leq n-1, \\ 1, & \text{if } \ U_{n,n} \leq t \leq 1, \end{cases}
\end{aligned}
$$

the *uniform empirical distribution function* based on $U_1, \ldots, U_n$ via indicator function $\mathbb{1}\{\cdot\}$ or, equivalently, on their corresponding uniform order statistics $U_{1,n} \leq U_{2,n} \leq \cdots \leq U_{n,n}$, and let

$$
(1.2)\qquad \alpha_n(t) = \sqrt{n}\,\{G_n(t) - t\}, \quad 0 \leq t \leq 1,
$$

be the corresponding *uniform empirical process.*

Next, define the *uniform empirical quantile function* as

$$
\begin{aligned}
(1.3)\qquad G_n^{-1}(t) &:= \inf\{s : G_n(s) \geq t\}, \quad 0 \leq t \leq 1,\ G_n^{-1}(0) := G_n^{-1}(0+), \\
&= \begin{cases} U_{1,n}, & \text{if } \ t = 0, \\ U_{k,n}, & \text{if } \ (k-1)/n < t \leq k/n,\ 1 \leq k \leq n, \end{cases}
\end{aligned}
$$

i.e., G_n^{-1} is defined to be the left continuous inverse of the right continuously defined uniform empirical distribution function G_n, and let

$$
(1.4)\qquad u_n(t) = \sqrt{n}\,\{G_n^{-1}(t) - t\}, \quad 0 \leq t \leq 1,
$$

be the corresponding *uniform quantile process.*

---

[1] Research supported by an NSERC Canada Discovery Grant at Carleton University, Ottawa.

AMS 2000 subject classifications. Primary 60F17; secondary 60F15, 60G50, 62G30.

Key words and phrases. Empirical and quantile processes, Brownian bridge approximations, Hungarian construction.

In his second landmark paper on invariance principles, Donsker (1952) proved that the empirical process $\alpha_n(t)$ converges in law to a Brownian bridge $B(t)$ with respect to the sup norm, except for some measurability problems that were sorted out later on (cf., e.g., Billingsley (1968, Section 18)). Komlós, Major and Tusnády (KMT) (1975) established a sharp bound for the speed of this weak convergence that reads as follows: *On a suitable probability space for the independent uniform* $(0, 1)$ *r.v.'s* $U_1, U_2, \ldots$, *the uniform empirical process* $\{\alpha_n(t),\, 0 \le t \le 1\}$ *can be approximated by a sequence of Brownian bridges* $\{B_n(t),\, 0 \le t \le 1\}$ *such that*

$$P\Big\{\sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| > n^{-1/2}(a\log n + x)\Big\} \le be^{-cx} \tag{1.5}$$

for all integers $n \ge 1$ *and* $x > 0$*, where* a, b *and* c *are positive absolute constants.*

It takes quite a bit of effort to arrive at (1.5) even after the KMT (1975) construction that establishes a joint distribution for $\alpha_n(\cdot)$ and $B_n(\cdot)$. For an insightful review of the construction of the latter joint distribution, we refer to Huang and Dudley (2001, Section 4) who study the speed of the convergence in law of $\alpha_n(t)$ to $B(t)$ with respect to the p–variation norm for $p \in (2, \infty)$ that was first established by Dudley (1992).

Concerning the proof of the KMT inequality (1.5) that is inequality (1.1) in Huang and Dudley (2001), they write: "Komlós, Major and Tusnády [42] specified a joint distribution for α_n and B_n, but beyond that published very little proof of (1.1). Csörgő and Révész ([23], Section 4.4), gave a proof in which a crucial lemma attributed to Tusnády was not proved. Bretagnolle and Massart [6] gave a proof, complete in principle, in which several steps were sketched. For versions of the Bretagnolle–Massart proof see also Csörgő and Horváth [16], pages 116–139, and [34]. Mason and van Zwet [47] give an alternative proof, also applying to subintervals, in which some steps were sketched. Mason [45] gives more details."

We note in passing that in [23] we gave a partial proof only for having

$$\sum_{n=1}^{\infty} P\Big\{\sup_{0\le t\le 1} \sqrt{n}\,|\alpha_n(t) - B_n(t)| > a\log n\Big\} < \infty$$

with some absolute positive constant a. In their proof, Bretagnolle and Massart (1989) concluded (1.5) with $a = 12$, $b = 2$ and $c = 1/6$. The more detailed version of the Bretagnolle–Massart proof in Csörgő and Horváth (1993, pages 116–139) also reproduces the same respective values for the constants a, b, c. For details on, and discussions of, the role and significance of Tusnády's lemma in the proof of the KMT inequality of (1.5), we refer to Csörgő and Révész (1981), Massart (2002), and Carter and Pollard (2004). In his 48 page manuscript, Major (2000) details the original KMT (1975) proof of (1.5). For a bootstrapped version of (1.5) with the same optimal speed for weak convergence we refer to Csörgő, Horváth and Kokoszka (2000).

In parallel to their approximation of $\alpha_n(\cdot)$ by $B_n(\cdot)$ as in (1.5), KMT (1975, 1976) also succeeded in approximating partial sums $S_n := X_1 + \cdots + X_n$, $n \ge 1$, $S_0 = 0$, of independent identically distributed random variables (i.i.d.r.v.'s) X, X_1, $X_2, \ldots$, that in case of having a finite moment generating function reads as follows: *On a rich enough probability space i.i.d.r.v.'s* $X, X_1, X_2, \ldots$, *with* $EX = 0$, $EX^2 = 1$ *and* $Ee^{tX} < \infty$ *in a neighbourhood of* $t = 0$*, together with a standard Wiener process* $\{W(s),\, 0 \le s < \infty\}$ *also on it, can be so constructed that for all* $x > 0$ *and integers* $n \ge 1$, *we have*

$$P\Big\{\sup_{0\le t\le 1}|S_{[nt]}-W(nt)|>a_F\log n+x\Big\}\le b_Fe^{c_Fx}, \tag{1.6}$$

where the positive constants a_F, b_F, c_F depend only on the distribution function F of X.

Building on this theorem, Csörgő and Révész (CsR) (cf. [20], [21], as well as [23], [10], [16]) proved the following analogue of the KMT inequality (1.5): *On a suitable probability space for the i.i.d. uniform $(0,1)$ r.v.'s $U_1,U_2,\ldots$, the uniform quantile process $\{u_n(t);\,0\le t\le 1\}$ can be approximated by a sequence of Brownian bridges $\{B_{1,n}(t);\,0\le t\le 1\}$ such that for all integers $n\ge 1$ and $x>0$, we have*

$$P\Big\{\sup_{0\le t\le 1}|u_n(t)-B_{1,n}(t)|>n^{-1/2}(a_1\log n+x)\Big\}\le b_1e^{-c_1x}, \tag{1.7}$$

where a_1, b_1, c_1 are positive absolute constants.

The approximation of $u_n(\cdot)$ by $B_{1,n}(\cdot)$ as stated in (1.7) is a slightly improved version of that of CsR (1978, 1981) in that here their restriction of $0<x\le cn^{1/2}$, $c>0$, is dropped in favour of all $x>0$. This improvement was established in Csörgő and Horváth (1993, Theorem 3.2.1).

We note that, due to their different constructions, the respective sequences of approximating Brownian bridges $B_n(\cdot)$ and $B_{1,n}(\cdot)$ of (1.5) and (1.7) are necessarily different. Indeed, as we will now see (cf. (1.12) and (1.13)), they cannot even in principle be identical sequences of Brownian bridges, due to the asymptotic behaviour of the sum

$$R_n(t):=\alpha_n(t)+u_n(t),\quad 0\le t\le 1, \tag{1.8}$$

of the uniform empirical and quantile processes of (1.2) and (1.4) respectively. $R_n(\cdot)$ is known in the literature as the *Bahadur–Kiefer process* (cf. Bahadur 1966, Kiefer 1967, 1970a). One of the famous results of Kiefer (1970a) reads

$$\limsup_{n\to\infty} n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4}\sup_{0\le t\le 1}|R_n(t)|=2^{-1/4}\quad\text{a.s.} \tag{1.9}$$

On the other hand, as $n\to\infty$, (1.5) and (1.7) respectively imply

$$\sup_{0\le t\le 1}|\alpha_n(t)-B_n(t)|=O(n^{-1/2}\log n)\quad\text{a.s.} \tag{1.10}$$

and

$$\sup_{0\le t\le 1}|u_n(t)-B_{1,n}(t)|=O(n^{-1/2}\log n)\quad\text{a.s.} \tag{1.11}$$

Consequently, on account of (1.9) and (1.10) we arrive at

$$\limsup_{n\to\infty} n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4}\sup_{0\le t\le 1}|u_n(t)+B_n(t)|=2^{-1/4}\quad\text{a.s.} \tag{1.12}$$

while, when (1.9) is combined with (1.11), we conclude

$$\limsup_{n\to\infty} n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4}\sup_{0\le t\le 1}|\alpha_n(t)+B_{1,n}(t)|=2^{-1/4}\quad\text{a.s.,} \tag{1.13}$$

in lieu of the respective approximations in (1.11) and (1.10).

On account of (1.9), resulting in (1.12) via (1.10) and in (1.13) via (1.11), it is impossible to hope for a simultaneous coupling of α_n and u_n *à la* (1.5) and (1.7) by the same sequence of Brownian bridges. However, extending the approach that resulted in the conclusion of the CsR (1978) inequality as in (1.7) (cf. [20], [21], as well as [23, Theorem 4.5.2], [10, Theorem 3.1.2 and Lemma 3.1.2], [16, Theorem 3.2.1]), it is possible to establish a partial version of (1.5) (cf. (1.15)) via (1.7) on the same probability space with relative ease. The aim of this exposition is to establish Theorem 1.1 in this regard.

We note in passing that our proof of (1.15), just like that of (1.7) $\equiv$ (1.14), hinges on one of the landmark KMT (1975, 1976) inequalities for approximating partial sums by Brownian motion as in (1.6). Furthermore, as already noted above, (1.7) via (1.9) and (1.11) leads to (1.13). Roughly speaking, this route of concluding (1.13) rhymes well with saying that Donsker's 1951 invariance principle for partial sums already implies his 1952 justification and extension of Doob's 1949 heuristic approach to the Kolmogorov-Smirnov theorems (cf. Breiman (1968, pp. 285-287)).

THEOREM 1.1. *On a suitable probability space for the i.i.d. uniform* $(0,1)$ *r.v.'s* $U_1, U_2, \ldots,$ *there exists a sequence of Brownian bridges* $\{B_{1,n}(t),\ 0 \le t \le 1\}$ *such that for all integers* $n \ge 1$ *and* $x > 0$, *we have*

$$P\Big\{\sup_{0\le t\le 1} |u_n(t) - B_{1,n}(t)| > n^{-1/2}(a_1 \log n + x)\Big\} \le b_1 e^{-c_1 x}, \tag{1.14}$$

and, with $B_{2,n}(\cdot) := -B_{1,n}(\cdot)$,

$$P\Big\{\max_{0\le k\le n} \sup_{U_{k,n}\le t< U_{k,1,n}} |\alpha_n(t) - B_{2,n}(k/n)| > n^{-1/2}(a_2 \log n + x)\Big\} \le b_2 e^{-c_2 x}, \tag{1.15}$$

where a_i, b_i, c_i, $i = 1, 2$, *are suitable positive constants, and* $U_{0,n} := 0$, $U_{n+1,n} := 1$.

REMARK 1.1. We note that the CsR inequalities of (1.7) and (1.14) are identical, and emphasize that (1.15) concludes a partial view of the KMT inequality (1.5) via the sequence of Brownian bridges of of (1.7) $\equiv$ (1.14).

The approximations mentioned in this introduction have found wide ranging applications in probability theory and statistics, and have inspired many further works in these subjects. For a glimpse of this impact we may refer to Berkes and Philipp (1979), Csörgő and Révész (1981), Csörgő (1983), M. Csörgő, S. Csörgő and Horváth (1986), CsCsHM (1986), Shorack and Wellner (1986), Part II of the proceedings volume edited by Hahn, Mason and Weiner (1991), Csörgő and Horváth (1993, 1997), Mason (2001), along with the many references therein. Just recently, in appreciation of the KMT (1975) inequality of (1.5), Carter and Pollard (2004) write in their Introduction: "In one of the most important probability papers of the last forty years, Komlós, Major and Tusnády (1975) sketched a proof for a very tight coupling of the standardized empirical distribution function with a Brownian bridge, a result now often referred to as the KMT, or Hungarian, construction. Their coupling greatly simplifies the derivation of many classical statistical results – see Shorack and Wellner [(1986), Chapter 12 et seq.], for example." For a review of the Shorack-Wellner 1986 book along these lines we may refer to Csörgő (1987). Carter and Pollard (2004) in their Introduction continue with writing: "The construction has taken on added significance for statistics with its use by Nussbaum (1996) in establishing asymptotic equivalence of density estimation and white noise models."

2. Proof of Theorem 1.1. To begin with, let $U_1, \ldots, U_n$, $n = 1, 2, \ldots$, be independent uniform $(0,1)$ r.v.'s, and let E_1, $E_2, \ldots$ be independent exponential r.v.'s with distribution function $1 - e^{-x}$, $x \geq 0$. Put $Z_k = E_1 + \cdots + E_k$, $k = 1, 2, \ldots$, and $Z_0 = 0$. Let $U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n}$ be the order statistics of the random sample $U_1, \ldots, U_n$, $n \geq 1$. Then (cf., e.g., Proposition 13.15 in Breiman (1968), or Proposition 8.2.1 in Shorack and Wellner (1986))

$$\{U_{k,n}; 1 \leq k \leq n\} \stackrel{\mathcal{D}}{=} \left\{ \frac{Z_k}{Z_{n+1}}; \ 1 \leq k \leq n \right\} \ \textit{for each} \ \ n = 1, 2, \ldots, \tag{2.1}$$

i.e., the indicated two vectors of random variables have the same joint distribution for each $n = 1, 2, \ldots$.

In view of (1.1), (1.3) and (2.1), we define

$$\tilde{G}_n(t) := \begin{cases} 0, & \text{if } \ 0 \leq t < Z_1/Z_{n+1}, \\ k/n, & \text{if } \ Z_k/Z_{n+1} \leq t < Z_{k+1}/Z_{n+1}, \ 1 \leq k \leq n-1, \\ 1, & \text{if } \ Z_n/Z_{n+1} \leq t \leq 1, \end{cases} \tag{2.2}$$

and

$$\begin{aligned} \tilde{G}_n^{-1}(t) &:= \inf\{s : \tilde{G}_n(s) \geq t\}, \ 0 \leq t \leq 1, \ \tilde{G}_n^{-1}(0) := \tilde{G}_n^{-1}(0+) \\ &= \begin{cases} Z_1/Z_{n+1}, & \text{if } \ t = 0, \\ Z_k/Z_{n+1}, & \text{if } \ (k-1)/n < t \leq k/n, \ 1 \leq k \leq n. \end{cases} \end{aligned} \tag{2.3}$$

Consequently, on account of (2.1), we have (cf. (1.1) and (2.2), resp. (1.3) and (2.3))

$$\{\tilde{G}_n(t); \ 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{G_n(t); \ 0 \leq t \leq 1\} \ \textit{for each} \ \ n = 1, 2, \ldots, \tag{2.4}$$

$$\{\tilde{G}_n^{-1}(t); \ 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{G_n^{-1}(t); \ 0 \leq t \leq 1\} \ \textit{for each} \ \ n = 1, 2, \ldots, \tag{2.5}$$

and whence also

$$\{\tilde{\alpha}_n(t) := n^{1/2}(\tilde{G}_n(t) - t); 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{\alpha_n(t); \ 0 \leq t \leq 1\} \tag{2.6}$$

for each $n = 1, 2, \ldots$,

$$\{\tilde{u}_n(t) := n^{1/2}(\tilde{G}_n^{-1}(t) - t); \ 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{u_n(t); \ 0 \leq t \leq 1\} \tag{2.7}$$

for each $n = 1, 2, \ldots$, i.e., the indicated random elements of $D[0,1]$ as stochastic processes in $t \in [0,1]$ are equal in distribution for each $n = 1, 2, \ldots$.

As a fundamental first step towards the proof of Theorem 1.1, we estimate the deviations

$$\begin{aligned} \tilde{A}_n &:= \max_{0 \leq k \leq n} \sup_{Z_k/Z_{n+1} \leq t < Z_{k+1}/Z_{n+1}} |\tilde{\alpha}_n(t) - \tilde{\alpha}_n(Z_k/Z_{n+1})| \\ &= \max_{0 \leq k \leq n} \sup_{Z_k/Z_{n+1} \leq t < Z_{k+1}/Z_{n+1}} |\tilde{\alpha}_n(t) + \tilde{u}_n(k/n)|, \quad n = 1, 2, \ldots. \end{aligned} \tag{2.8}$$

Lemma 2.1. *On the* KMT (1975, 1976) *probability space for having* (1.6)*, let* E*,* E_1*,* $E_2, \ldots$ *be i.i.d. exponential r.v.'s with mean* 1*, partial sums* $Z_0 = 0$*,* $Z_k = E_1 + \cdots + E_k$*,* $k = 1, 2, \ldots$*, together with a standard Wiener process* $\{W(s), 0 \leq s < \infty\}$ *so that for all* $x > 0$ *and integers* $n \geq 1$*, we have*

$$P\Big\{ \sup_{0 \leq t \leq 1} |(Z_{[nt]} - [nt]) - W(nt)| > C_1 \log n + x \Big\} \leq C_2 e^{-C_3 x}, \tag{2.9}$$

where the positive constants C_i, $i=1,2,3$, depend only on the distribution function $1-e^{-x}$ of E. Then, for all $x>0$ and integers $n\ge 1$, we have

$$P\Big\{\tilde{A}_n > n^{-1/2}(a_3\log n + x)\Big\} \le b_3 e^{-c_3 x}, \tag{2.10}$$

where a_3, b_3, c_3 are suitable positive constants.

Proof. First we note that

$$\tilde{A}_n \le 4n^{1/2}. \tag{2.11}$$

Hence, in view of (2.11), in order to obtain (2.10) for all $x>0$, it suffices to prove it for all $x\le (C_4+4)n$, where $C_4>1$ is a constant. We consider

$$\begin{aligned}
&\tilde{\alpha}_n(t)-\tilde{\alpha}_n\left(\frac{Z_k}{Z_{n+1}}\right)\\
&= n^{1/2}\left((\tilde{G}_n(t)-t)-\left(\frac{k}{n}-\frac{Z_k}{Z_{n+1}}\right)\right),\ 0\le t\le 1,\ k=0,1,\ldots,n,\\
&=\begin{cases}
n^{1/2}(0-t) & ,\ 0\le t<\frac{Z_1}{Z_{n+1}},\quad k=0\\
n^{1/2}\left(\left(\frac{k}{n}-t\right)-\left(\frac{k}{n}-\frac{Z_k}{Z_{n+1}}\right)\right) & ,\ \frac{Z_k}{Z_{n+1}}\le t<\frac{Z_{k+1}}{Z_{n+1}},\ 1\le k\le n-1,\\
n^{1/2}\left((1-t)-\left(1-\frac{Z_n}{Z_{n+1}}\right)\right) & ,\ \frac{Z_n}{Z_{n+1}}\le t\le 1,\quad k=n.
\end{cases}
\end{aligned}\tag{2.12}$$

Consequently, we conclude

$$\begin{aligned}
&\left|\tilde{\alpha}_n(t)-\tilde{\alpha}_n\left(\frac{Z_k}{Z_{n+1}}\right)\right|,\quad 0\le t\le 1,\quad k=0,1,\ldots,n,\\
&=\begin{cases}
n^{1/2}t & ,\ 0\le t<\frac{Z_1}{Z_{n+1}},\quad k=0\\
n^{1/2}\left(t-\frac{Z_k}{Z_{n+1}}\right) & ,\ \frac{Z_k}{Z_{n+1}}\le t<\frac{Z_{k+1}}{Z_{n+1}},\ 1\le k\le n-1,\\
n^{1/2}\left(t-\frac{Z_n}{Z_{n+1}}\right) & ,\ \frac{Z_n}{Z_{n+1}}\le t\le 1,\quad k=n,
\end{cases}
\end{aligned}\tag{2.13}$$

as well as

$$\begin{aligned}
&\max_{0\le k\le n}\ \sup_{Z_k/Z_{n+1}\le t<Z_{k+1}/Z_{n+1}}\left|\tilde{\alpha}_n(t)-\tilde{\alpha}_n\left(\frac{Z_k}{Z_{n+1}}\right)\right|\\
&\le n^{1/2}\left\{\frac{Z_1}{Z_{n+1}}\vee\left(\max_{1\le k\le n-1}\left(\frac{Z_{k+1}}{Z_{n+1}}-\frac{Z_k}{Z_{n+1}}\right)\right)\vee\left(1-\frac{Z_n}{Z_{n+1}}\right)\right\}\\
&= n^{1/2}\left\{\frac{E_1}{Z_{n+1}}\vee\left(\max_{1\le k\le n-1}\frac{E_{k+1}}{Z_{n+1}}\right)\vee\left(\frac{E_{n+1}}{Z_{n+1}}\right)\right\} =: \Delta_n,
\end{aligned}\tag{2.14}$$

where $\vee$ stands for maximum. Estimating now the latter r.v. for the sake of concluding (2.10), with some positive constant C_5 we consider

$$\begin{aligned}
&P\Big\{\Delta_n > n^{-1/2}(C_5\log n + x)\Big\}\\
&= P\left\{n^{1/2}\max_{0\le k\le n}\frac{E_{k+1}}{Z_{n+1}} > n^{-1/2}(C_5\log n+x)\right\}\\
&=: P_n.
\end{aligned}\tag{2.15}$$

Hence, in view of (2.11) and (2.14)–(2.15), in order to verify (2.10) it suffices to show that, with some positive constants C_6, C_7, we have

$$P_n \le C_6 e^{-C_7 x}, \tag{2.16}$$

for all integers $n \ge 1$ and $x \le (C_4+4)n$, with constant $C_4 > 1$.

We have

$$\begin{aligned} P_n &\le P\left\{\frac{n}{Z_n} \max_{0\le k\le n} E_{k+1} > C_5 \log n + x\right\} \\ &=: P\{A_n(x)\}, \end{aligned} \tag{2.17}$$

$$\begin{aligned} &P\{A_n(x)\} \le P\{A_n(x),\, Z_n \ge n/2\} + P\{Z_n < n/2\} \\ &\le P\Big\{2 \max_{0\le k\le n} E_{k+1} > C_5 \log n + x\Big\} + P\{Z_n - n < -n/2\} \\ &=: P_{1,n} + P_{2,n}, \end{aligned} \tag{2.18}$$

where, immediately,

$$\begin{aligned} P_{1,n} &\le (n+1)P\{E_1 > (C_5 \log n + x)/2\} \\ &= (n+1)e^{-(C_5/2)\log n}\, e^{-x/2} \le C_8 e^{-C_9 x}. \end{aligned} \tag{2.19}$$

As to $P_{2,n}$, on remembering that we are working on the KMT probability space for having (2.9), we arrive at

$$\begin{aligned} P_{2,n} &= P\{(Z_n - n) - W(n)) + W(n) < -n/2\} \\ &\le P\{|(Z_n - n) - W(n)| > n/4\} + P\{|W(1)| > n^{1/2}/4\} \\ &\le C_{10} e^{-C_{11} n} \\ &\le C_{10} e^{-C_{12} x} \end{aligned} \tag{2.20}$$

for all $x \le (C_4+4)n$, with $C_4 > 1$, and $C_{12} = C_{11}/(C_4+4)$.

Thus, via (2.19) and (2.20), we conclude (2.16), and hence (2.10) as well. $\square$

In view of $\tilde{A}_n$ of (2.8), we define the deviations

$$\begin{aligned} A_n &:= \max_{0\le k\le n} \sup_{U_{k,n}\le t< U_{k+1,n}} |\alpha_n(t) - \alpha_n(U_{k,n})| \\ &= \max_{0\le k\le n} \sup_{U_{k,n}\le t< U_{k+1,n}} |\alpha_n(t) + u_n(k/n)|\,, \quad n = 1,2,\ldots. \end{aligned} \tag{2.21}$$

Then, on account of (2.1)–(2.7), we have

$$\tilde{A}_n \stackrel{\mathcal{D}}{=} A_n, \quad \text{for each } n = 1,2,\ldots \tag{2.22}$$

and, consequently, the following corollary to Lemma 2.1 as well.

COROLLARY 2.1. *With the positive constants of* a_3, b_3, c_3 *of* Lemma 2.1 *we also have*

$$P\left\{A_n > n^{-1/2}(a_3 \log n + x)\right\} \le b_3\, e^{-c_3 x} \tag{2.23}$$

for all $x>0$ and integers $n\geq 1$.

PROOF OF THEOREM 1.1. We have

$$|\alpha_n(t)-B_{2,n}(k/n)| = |\alpha_n(t)-\alpha_n(U_{k,n})+\alpha_n(U_{k,n})-B_{2,n}(k/n)| \\ = |\alpha_n(t)-\alpha_n(U_{k,n})+B_{1,n}(k/n)-u_n(k/n)|. \tag{2.24}$$

Therefore,

$$\max_{0\leq k\leq n}\ \sup_{U_{k,n}\leq t<U_{k+1,n}} |\alpha_n(t)-B_{2,n}(k/n)| \leq A_n+\max_{0\leq k\leq n}|B_{1,n}(k/n)-u_n(k/n)|. \tag{2.25}$$

Hence, on combining (1.7) $\equiv$ (1.14) with (2.23), we arrive at (1.15). $\square$

REMARK 2.1. The feasibility of the identical inequalities (1.7) and (1.14) being true can be seen via first approximating $\tilde{u}_n$ by the Brownian bridges

$$\tilde{B}_{1,n}(t) := n^{-1/2}(W(nt)-tW(n)), \quad 0\leq t\leq 1, \tag{2.26}$$

with Wiener process $W(\cdot)$ as in (2.9) of Lemma 2.1, by writing (cf. [21], [23, Theorem 4.5.2] and [16, Theorem 3.2.1])

$$\tilde{u}_n\left(\frac{k}{n}\right)-\tilde{B}_{1,n}\left(\frac{k}{n}\right) = n^{-1/2}\Big\{\Big(Z_k-k-W(k)\Big)-\frac{k}{n}(Z_n-n-W(n)\Big) \\ -\frac{k}{n}E_{n+1}+\left(\frac{n}{Z_{n+1}}-1\right)\left(Z_k-k-\frac{k}{n}\left(Z_{n+1}-n\right)\right)\Big\},$$

and keeping in mind Lemma 1 of [22] for estimating appropriate increments of a Wiener process, and noting that we have

$$\max_{0\leq k\leq n}\ \sup_{(k-1)/n<t\leq k/n}\left|\tilde{u}_n(t)-\tilde{u}_n\Big(\frac{k}{n}\Big)\right| \\ = \max_{0\leq k\leq n}\ \sup_{(k-1)/n<t\leq k/n} n^{1/2}\left|\Big(\frac{Z_k}{Z_{n+1}}-t\Big)-\Big(\frac{Z_k}{Z_{n+1}}-\frac{k}{n}\Big)\right| \leq 1/n^{1/2}.$$

This approach results in a coupling of $\tilde{u}_n$ and $\tilde{B}_{1,n}$ that is preliminary to that of u_n and $B_{1,n}$ as in (1.7) $\equiv$ (1.14), and on its own it yields

$$\sup_{0\leq t\leq 1}|\tilde{u}_n(t)-\tilde{B}_{1,n}(t)| = O(n^{-1/2}\log n) \quad \text{a.s.} \tag{2.27}$$

as $n\to\infty$.

3. Some further notes and historical remarks. The respective coupling inequalities of (1.5) and (1.7) provide an optimal bound for the speed of the convergence in law of $\alpha_n(\cdot)$ and $u_n(\cdot)$ to a Brownian bridge $B(\cdot)$ (cf. (1.10) and (1.11)).

Brillinger (1969) was first in studying the almost sure behaviour of $\tilde{\alpha}_n$ with rates. Namely, based on Strassen's 1965 approximation of partial sums by a Wiener process (cf. [54]), he established

$$\sup_{0\leq t\leq 1}|\tilde{\alpha}_n(t)-\tilde{B}_{2,n}(t)| = O\left(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\right) \quad \text{a.s.} \tag{3.1}$$

as $n \to \infty$, with the sequence of Brownian bridges

$$\tilde{B}_{2,n}(t) := n^{-1/2}(\tilde{W}(nt) - t\tilde{W}(n)), \quad 0 \le t \le 1, \tag{3.2}$$

where Brownian motion $\tilde{W}(\cdot)$ is as in Strassen (1965).

The respective conclusions of (2.27) and (3.1) also imply strong theorems, but only for $\tilde{u}_n(\cdot)$ and $\tilde{\alpha}_n(\cdot)$, and not for $\alpha_n(\cdot)$ and $u_n(\cdot)$, whose respective weak convergences can of course be also deduced from them. As to the hinted at strong theorems, via Strassen's 1964 functional law of the iterated logarithm (LIL) for a standard Wiener process $W(\cdot)$ (cf. [53]), it can be shown for example that

$$\limsup_{n\to\infty} \sup_{0\le t\le 1} \frac{|W(nt) - tW(n)|}{(n \log\log n)^{1/2}} = 2^{-1/2} \quad \text{a.s.} \tag{3.3}$$

Consequently, on recalling the definitions of $\tilde{B}_{1,n}(\cdot)$ and $\tilde{B}_{2,n}(\cdot)$ (cf. (2.26) and (3.2)), via (2.27) and (3.1) we arrive at

$$\limsup_{n\to\infty} \sup_{0\le t\le 1} |\tilde{u}_n(t)|/(\log\log n)^{1/2} = 2^{-1/2} \quad \text{a.s.} \tag{3.4}$$

and

$$\limsup_{n\to\infty} \sup_{0\le t\le 1} |\tilde{\alpha}_n(t)|/(\log\log n)^{1/2} = 2^{-1/2} \quad \text{a.s.} \tag{3.5}$$

It is well known that the respective LIL of (3.4) and (3.5) also hold true for $u_n(\cdot)$ and $\alpha_n(\cdot)$, but the latter strong laws do not of course follow from (2.27) and (3.1), for the equalities in distribution of (2.6) and (2.7) hold true only marginally in n.

Kiefer (1970b) was first to call attention to the desirability of viewing the empirical process $\alpha_n(t)$ as a two-time parameter stochastic process, a function of t and n, and that a strong approximation theorem for $\alpha_n(t)$ should be given in terms of an appropriate two-time parameter Gaussian process. He also succeeded in giving the first solution to this problem (cf. Kiefer (1972)) by approximating $\alpha_n(t)$ by a Gaussian process whose covariance function coincides with that of $\sqrt{n}\,\alpha_n(t)$, i.e., with $E\sqrt{m}\alpha_m(s)\sqrt{n}\,\alpha_n(t) = m \wedge n(s \wedge t - st)$, where $\wedge$ stands for minimum. Preceeding this work, Müller (1970) proved a corresponding two-time parameter weak convergence for $\alpha_n(t)$ via using Rényi's (1953) exponential representation of the empirical process, and he also gave the first estimate of the error for the convergence in distribution of certain functionals of the sequence of empirical processes. Inspired by Kiefer's 1972 landmark paper [39], CsR (1975a) established the first coupling inequalities for approximating the uniform empirical process that is based on independent random d–vectors that are uniformly distributed on $I^d := [0,1]^d$, $d \ge 1$, by d–time parameter Brownian bridges $\{B_n(t),\, t \in I^d,\, d \ge 1\}_{n=1}^{\infty}$, as well as by a separable mean zero Gaussian process $K(\cdot,\cdot)$ on $[0,1]^d \times [0,\infty)$, which is a d–time parameter Brownian bridge in its first argument and Brownian motion in its second argument. Such a process was called a Kiefer process in CsR (1975a), and this has endured since, quite deservedly so, on account of Kiefer's fundamental first step in this regard in 1972. For a quick review of many further significant steps along these lines we refer to Csörgő (2002, pages 41–45), and Horváth and Szyszkowicz, eds. (2004, pages 8–14). For example, considering the empirical distribution on the unit cube $I^2 = [0,1]^2$, the best available Brownian bridge type approximation is due to Tusnády (1977) with the rate $O(n^{-1/2}(\log n)^2)$ that coincides with the KMT (1975) strong Kiefer type invariance principle for $d = 1$ (cf. also Castelle and Laurent–Bonvalot (1998)), which cannot be improved beyond $(\log n)$. By combining the latter KMT (1975) Kiefer type uniform strong approximation of $\alpha_n(t)$ by $K(t,n)/n^{1/2}$ with Kiefer's result of 1970 as in (1.9), CsR (1975b) observed the following invariance principle for the uniform quantile process $u_n(t)$:

$$\limsup n^{1/4}(\log n)^{-1/2}(\log\log n)^{-1/4} \sup_{0\le t\le 1} |u_n(t) - n^{-1/2}K_0(t,n)| = 2^{-1/4} \quad \text{a.s.,} \tag{3.6}$$

where $K_0(\cdot,\cdot) = -K(\cdot,\cdot)$. In other words, the same Kiefer process that KMT (1975) constructed for approximating $\alpha_n(\cdot)$, approximates $u_n(\cdot)$ as well as in (3.6) via (1.9). The same can be said about the first such two-time parameter Gaussian process that was constructed by Kiefer (1972) for strongly approximating $\alpha_n(\cdot)$ at the rate $O(n^{-1/6}(\log n)^{2/3})$. Deheuvels (1998) showed that approximating $u_n(\cdot)$ by any other, not for $\alpha_n(\cdot)$ constructed Kiefer process at a better rate than that of (3.6) is also impossible. Thus the rate of convergence in observation (3.6) for approximating $u_n(\cdot)$ by a Kiefer process is optimal not only for $\alpha_n(\cdot)$ constructed Kiefer processes, but also for any other Kiefer process. This is to be contrasted with approximating $\alpha_n(\cdot)$ and $u_n(\cdot)$ respectively, both at the same optimal rate, by two different sequences of Brownian bridges as in (1.5) and (1.7).

In conclusion we note that if $X_1, X_2, \ldots$ are independent random variables with a right continuously defined distribution function F, and F_n is the empirical distribution function of the first $n \ge 1$ of these random variables, then via (1.1) and (1.2) we have

$$F_n(x) = G_n(F(x)), \quad \text{and} \quad \beta_n(x) := \sqrt{n}\,(F_n(x) - F(x)) = \alpha_n(F(x)), \tag{3.7}$$

$x \in I\!R$, for all n, on account of $X_i \stackrel{\mathcal{D}}{=} F^{-1}(U_i)$, $i = 1, 2, \ldots$, where F^{-1} is the left continuous inverse of F, and if F is continuous, then

$$F_n(F^{-1}(t)) = G_n(t), \quad \text{and} \quad \beta_n(F^{-1}(t)) = \alpha_n(t),\ 0 \le t \le 1, \tag{3.8}$$

for all n, on account of $F(X_i) \stackrel{\mathcal{D}}{=} U_i$, $i = 1, 2, \ldots$.

Hence, it follows by (3.8) that (1.5), (1.10), (1.15) continue to hold true for β_n if F is continuous. Moreover, when F is arbitrary, then, inserting the function F into the argument of the random functions occuring in (1.5), (1.10), (1.15), by (3.7) and having also

$$\sup_{x\in I\!R} |\alpha_n(F(x)) - B_n(F(x))| \le \sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| \tag{3.9}$$

with any one of the sequences of Brownian bridges involved in the just mentioned results, they remain true for β_n also when F is an arbitrary distribution function on the real line.

Define now F_n^{-1} *à la* G_n^{-1} of (1.3), and the natural looking quantile process γ_n *à la* u_n of (1.4) as

$$\gamma_n(t) := \sqrt{n}(F_n^{-1}(t) - F^{-1}(t)), \quad 0 \le t \le 1.$$

Unfortunately, there is no such immediate simple route like that of (3.8) for transforming γ_n into its own corresponding uniform quantile process form

$$u_n(t) = \sqrt{n}\,(G_n^{-1}(t) - t) := \sqrt{n}\,(F(F_n^{-1}(t)) - t),\ 0 \le t \le 1, \tag{3.10}$$

so that one could quickly adapt results like the ones studied in this paper for u_n to γ_n as well. Hence, on assuming that F has a Lebesgue density function f on $I\!R$, CsR (1978) defined the *general quantile process*

$$\rho_n(t) := \sqrt{n}\, f(F^{-1}(t))(F_n^{-1}(t) - F^{-1}(t)),\ 0 \le t \le 1, \tag{3.11}$$

and, based on a law of the iterated logarithm of Csáki (1977) for the standardized version of the uniform empirical process α_n, succeeded in studying ρ_n via its deviations from its own u_n as in (3.10), under some natural conditions on F. The CsR (1978) conditions and their implications have since been further studied and utilized, for example in CsR (1981), Csörgő (1983), Csörgő et al. (1985), Shorack and Wellner (1986), Csörgő and Horváth (1993), Csörgő and Szyszkowicz (1998), Drees and De Haan (1999), Csörgő and Shi (2001), Csörgő and Zitikis (2002), del Barrio et al. (2005), and in many other works that are referred to in this selected list.

Acknowledgements. Discussions with Endre Csáki, Pál Révész and Barbara Szyszkowicz of an earlier version of this exposition have led to this revision. I wish to thank them for their much appreciated time and interest in this work.

References

[1] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[2] Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 19–54.

[3] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] Brillinger, D. (1969). An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* **17** 189–200.

[5] Breiman, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

[6] Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256.

[7] Carter, A. and Pollard, D. (2004). Tusnády's inequality revisited. *Ann. Statist.* **32** 2731–1741.

[8] Castelle, N. and Laurent-Bonvalot, F. (1998). Strong approximations of bivariate uniform empirical processes. *Ann. Inst. Henri Poincaré* **34** 425–480.

[9] Csáki, E. (1977). The law of the iterated logarithm for normalized empirical distribution function. *Z. Wahrsch. Verw. Gebiete* **38** 147–167.

[10] Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* SIAM, Philadelphia.

[11] Csörgő, M. (1987). Review of *Empirical Processes with Applications to Statistics*, by Galen R. Shorack and Jon A. Wellner. Wiley, New York 1986, in: *Bull. Amer. Math. Soc.* **17** 189–200.

[12] Csörgő, M. (2002). Life is a random walk with a local time: a tribute to Pál Révész. In: *Limit Theorems in Probability and Statistics I.* (I. Berkes, E. Csáki, M.Csörgő, eds.) 21–99. *János Bolyai Mathematical Society*, Budapest.

[13] Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes*. Lecture Notes in Statistics **33**, Springer-Verlag, Heidelberg.

[14] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

[15] Csörgő, M., Csörgő, S., Horváth, L. and Révész, P. (1985). On weak and strong approximations of the quantile process. In: *Proc. Seventh Conf. Probability Theory*, 81–95. Editura Academiiei, Bucuresti and Nuscience Press, Utrecht.

[16] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, Chichester.

[17] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, Chichester.

[18] Csörgő, M., Horváth, L. and Kokoszka, P. (2000). Approximation for bootstrapped empirical processes. *Proc. Amer. Math. Soc.* **128** 2457–2464.

[19] Csörgő, M. and Révész, P. (1975a). A new method to prove Strassen type laws of invariance principle II. *Z. Wahrsch. Verw. Gebiete* **31** 261–269.

[20] Csörgő, M. and Révész, P. (1975b). Some notes on the empirical distribution function and the quantile process. In: *Coll. Math. Soc. J. Bolyai* **11** *Limit Theorems of Probability Theory* (P. Révész, ed.) 59–71. North Holland, Amsterdam.

[21] Csörgő, M. and Révész, P. (1978). Strong approximation of the quantile process. *Ann. Statist.* **6** 882–894.

[22] Csörgő, M. and Révész, P. (1979). How big are the increments of a Wiener process? *Ann. Probab.* **7** 731–737.

[23] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[24] Csörgő, M. and Shi, Z. (2001). An L^p–view of a general version of the Bahadur–Kiefer process. *Journal of Mathematical Sciences* **105** 2534–2540.

[25] Csörgő, M. and Szyszkowicz, B. (1998). Sequential quantile and Bahadur-Kiefer processes. In: *Handbook of statistics*, Vol. 16 (N. Balakrishnan and C.R. Rao, eds.) 631–688. Elsevier Science, North-Holland.

[26] Csörgő, M. and Zitikis, R. (2002). On the general Bahadur–Kiefer, quantile, and Vervaat processes: old and new. In: *Limit Theorems in Probability and Statistics I*, (I. Berkes, E. Csáki, M. Csörgő, eds.) 389–426. *János Bolyai Mathematics Society*, Budapest.

[27] Deheuvels, P. (1998). On the approximation of quantile processes by Kiefer processes. *J. Theoret. Probab.* **11** 997–1018.

[28] del Barrio, E., Giné, E. and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189.

[29] Donsker, N, (1951). An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, No. 6.

[30] Donsker, N, (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277-281.

[31] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

[32] Drees, H. and de Haan, L. (1999). Conditions for quantile process approximations. *Comm. Statist. – Stoch. Models* **15** 485–502.

[33] Dudley, R.M. (1992). Fréchet differentiability, p–variation and uniform Donsker classes. *Ann. Probab.* **20** 1968–1982.

[34] Dudley, R.M. (1999). An exposition of Bretagnolle and Massart's proof of the KMT theorem for the uniform empirical process. Preprint.

[35] Hahn, M., Mason, D.M. and Weiner, D. (1991), eds. *Sums, Trimmed Sums and Extremes.* Birkhäuser, Boston.

[36] Horváth, L. and Szyszkowicz, B., eds. (2004). Path Properties of Forty Years of Research in Probability and Statistics: In conversation with Miklós Csörgő. *Technical Report Series of the Laboratory for Research in Statistics and Probability*, No. 400. Carleton University–University of Ottawa.

[37] Huang, Y.-C. and Dudley, R.M. (2001). Speed of convergence of classical empirical processes in p–variation. *Ann. Probab.* **29** 1625–1636.

[38] Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342.

[39] Kiefer, J. (1970a). Deviations between the sample quantile process and the sample DF. In: *Nonparametric Techniques in Statistical Inference*, (M.L. Puri, ed.) 299–319. Cambridge University Press, Cambridge.

[40] Kiefer, J. (1970b). Old and new methods for studying order statistics and sample quantiles. In: *Nonparametric Techniques in Statistical Inference* (M.L. Puri, ed.) 349–357. Cambridge University Press, Cambridge.

[41] Kiefer, J. (1972). Skorohod embedding of multivariate RVs and the sample DF. *Z. Wahrsch. Verw. Gebiete* **24** 1–35.

[42] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111-131.

[43] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58.

[44] Major, P. (2000). The approximation of the normalized empirical distribution function by a Brownian bridge. Notes available from www.renyi.hu/˜major/.

[45] Mason, D.M. (1998). Notes on the KMT Brownian bridge approximation to the uniform empirical process. Preprint.

[46] Mason, D.M. (2001). An exponential inequality for a weighted approximation to the uniform empirical process with applications. In: *State of the Art in Probability and Statistics* (Leiden, 1999) 477–498, *IMS Lecture Notes Monogr. Ser.* **36**, IMS, Beachwood, OH.

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

[48] Massart, P. (2002). Tusnády's lemma, 24 years later. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 991–1007.

[49] Müller, D.W. (1970). On Glivenko-Cantelli convergence. *Z. Wahrsch. Verw. Gebiete* **16**, 195–220.

[50] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

[51] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4** 191–232.

[52] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[53] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

[54] Strassen, V. (1965). Almost sure behaviour of sums of independent random variables and martingales. In: *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** 315–344.

[55] Tusnády, G. (1977). A remark on the approximation of the sample DF in the multidimensional case. Periodica Math. Hungar. **8** 53–55.

School of Mathematics and Statistics, Carleton University
1125 Colonel By Drive, Ottawa, Canada K1S 5B6
E-mail: mcsorgo@math.carleton.ca