The Annals of Probability
2001, Vol. 29, No. 4, 1625–1636

SPEED OF CONVERGENCE OF CLASSICAL EMPIRICAL PROCESSES IN p-VARIATION NORM

BY YEN-CHIN HUANG AND R. M. DUDLEY[1]

Inspex, Inc. and Massachusetts Institute of Technology

Let F be any distribution function on $\mathbb{R}$, and F_n be the nth empirical distribution function based on variables i.i.d. (F). It is shown that for $2 < p < \infty$ and a constant $C(p) < \infty$, not depending on F, on some probability space there exist F_n and Brownian bridges B_n such that for the Wiener-Young p-variation norm $\|\cdot\|_{[p]}$, $E\|n^{1/2}(F_n - F) - B_n \circ F\|_{[p]} \le C(p)n^{(2-p)/(2p)}$, where $(B_n \circ F)(x) = B_n(F(x))$. The expectation can be replaced by an Orlicz norm of exponential order. Conversely, if F is continuous, then for any stochastic process $V(t, \omega)$ continuous in t for almost all ω, such as $B_n \circ F$, summation over n distinct jumps shows that $\|n^{1/2}(F_n - F) - V\|_{[p]} \ge n^{(2-p)/(2p)}$, so the upper bound in expectation is best possible up to the constant $C(p)$. In the proof, B_n is linked to F_n by the Komlós, Major and Tusnády construction, as for the supremum norm $(p = \infty)$.

1. Introduction. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with uniform law $U[0, 1]$ having distribution function U. Let $U_n(t)$ be the empirical distribution function based on $X_1, X_2, \ldots, X_n$ and $\alpha_n(t)$ the corresponding empirical process, that is, $\alpha_n(t) = \sqrt{n}(U_n(t) - t)$, $t \in [0, 1]$. Donsker [5] proved in 1952, except for some measurability problems, that the empirical process $\alpha_n(t)$ converges in law to a Brownian bridge $B(t)$ with respect to the sup norm. A sharp bound for the speed of this convergence was indicated by Komlós, Major and Tusnády [13]. They stated in 1975 that on some probability space there exist X_i i.i.d. $U[0, 1]$ and Brownian bridges B_n such that

$$P\left(\sup_{0\le t\le 1} |\sqrt{n}(\alpha_n(t) - B_n(t))| > c\log n + x\right) < Ke^{-\lambda x} \tag{1.1}$$

for all n and x, where c, K, and λ are positive absolute constants. Komlós, Major and Tusnády [13] specified a joint distribution for α_n and B_n but beyond that published very little proof of (1.1). Csörgő and Révész ([4], Section 4.4), gave a partial proof in which a crucial lemma attributed to Tusnády was not proved. Bretagnolle and Massart [2] gave a proof, complete in principle, in which several steps were sketched. For versions of the Bretagnolle-Massart proof see also Csörgő and Horváth [3], pages 116-139, and [8]. Mason and van Zwet [18] give an alternative proof, also applying to subintervals, in which some steps were sketched. Mason [17] gives more details.

Received November 1999; revised January 2001
[1]Supported in part by the NSF.
AMS 2000 *subject classifications*. Primary 62G30, 60F17.
Key words and phrases. Brownian bridge, Orlicz norms.

On the other hand, in practice, the sup norm is often not strong enough. For example, several statistical functionals of interest are not Fréchet differentiable in the sup norm but are Fréchet differentiable in some p-variation norms (see Dudley [6], [7], Dudley and Norvaiša [9]) defined as follows. For a real-valued function f on an interval J and $0 < p < \infty$, let its *p-variation* on J be $v_p(f, J) := \sup\{\sum_{i=1}^m |f(t_i) - f(t_{i-1})|^p : t_0 \in J, t_0 < t_1 < \cdots < t_m \in J, \ \ m = 1, 2, \ldots\}$. Let f be such that $v_p(f) < \infty$. For $1 \leq p < \infty$, the p-variation seminorm is defined by $\|f\|_{(p)} := v_p(f)^{1/p}$. Let $\|f\|_\infty := \sup_{x \in J} |f(x)|$. Then the p-variation norm is defined by $\|f\|_{[p]} := \|f\|_{(p)} + \|f\|_\infty$.

For a function f which is 0 somewhere in J, as α_n and B_n are at the endpoints of $[0, 1]$, we have for $1 \leq p < \infty$,

$$\|f\|_\infty \leq \|f\|_{(p)} \text{ and so } \|f\|_{[p]} \leq 2\|f\|_{(p)}. \tag{1.2}$$

Dudley [6] showed that the convergence in law of $\alpha_n(t)$ to $B(t)$ still holds with respect to the p-variation norm for $p \in (2, \infty)$. This article focuses on finding the speed of convergence. We will use some Orlicz norms. A *Young-Orlicz modulus* is a convex, increasing function from $[0, \infty)$ onto itself. Let g be a Young-Orlicz modulus and $(X, \mathscr{S}, \mu)$ a measure space. Let $\mathscr{L}_g(X, \mathscr{S}, \mu)$ denote the set of all measurable real-valued functions on X such that

$$\|f\|_g := \inf\left\{t > 0 : \int g(|f(x)|/t) d\mu(x) \leq 1\right\} < \infty.$$

Let $L_g := L_g(X, \mathscr{S}, \mu)$ be the collection of equivalence classes of functions in $\mathscr{L}_g(X, \mathscr{S}, \mu)$ for equality μ-almost everywhere. It is known that L_g with the norm $\|\cdot\|_g$ is a Banach space, and that for $f \in \mathscr{L}_g$ with $\|f\|_g > 0$,

$$\int g(|f(x)|/\|f\|_g) d\mu(x) \leq 1, \tag{1.3}$$

for example, Luxemburg and Zaanen [15]. For $y \geq 0$ and $1 \leq p < \infty$ let

$$g_p(y) := (ey/p)^p 1_{0 \leq y < p} + e^y 1_{y \geq p}.$$

The left and right limits and derivatives of g_p at $y = p$ all equal e^p, so g_p is a Young-Orlicz modulus. Our main result is:

THEOREM 1. *For $2 < p < \infty$ there is a constant $A(p) < \infty$ such that if F is any d.f.* (*distribution function*) *on $\mathbb{R}$, then on some probability space there exist $X_1, X_2, \ldots,$ i.i.d.* (F) *and Brownian bridges B_n such that for all n, if F_n is the empirical d.f. based on $X_1, \ldots, X_n$, then*

$$\|\|n^{1/2}(F_n - F) - B_n \circ F\|_{[p]}\|_{g_p} \leq A(p) n^{-(p-2)/(2p)}. \tag{1.4}$$

In particular, for some constant $C(p) < \infty$,

$$E\|n^{1/2}(F_n - F) - B_n \circ F\|_{[p]} \leq C(p) n^{-(p-2)/(2p)}. \tag{1.5}$$

If F is continuous, then for $V = B_n \circ F$ or any sample-continuous process V,

$$\|n^{1/2}(F_n - F) - V\|_{(p)} \geq n^{-(p-2)/(2p)} \qquad a.s. \tag{1.6}$$

It follows from (1.3) that for $\|f\|_g > 0$

$$\mu(\{x: |f(x)| \geq t\|f\|_g\}) \leq e^{-t} \qquad \text{for } g = g_p \text{ and } t \geq p. \tag{1.7}$$

From (1.4) and (1.7) we get the exponential bound [cf. (1.1)]

$$P\left(\|\sqrt{n}(F_n - F) - B_n \circ F\|_{[p]} > xn^{-(p-2)/(2p)}\right) \leq e^{-x/A(p)} \qquad \text{for } x \geq pA(p).$$

In Section 2 we will define another Orlicz norm suitable for applying to p-variations v_p. The rest of the paper is then devoted to proving Theorem 1.

The lower bound (1.6) is elementary: since F is continuous, the X_i are a.s. distinct, so the function $n^{1/2}(F_n - F) - V$ has n distinct jumps of height $n^{-1/2}$, and (1.6) follows.

For the upper bound (1.4), the linkage of empirical processes and Brownian bridges will be by means of the Komlós, Major and Tusnády construction ([13], [2], [18]).

For the uniform d.f. U and its empirical d.f.'s U_n as above, clearly $F \equiv U \circ F$ and we can write $F_n \equiv U_n \circ F$. For any d.f. F and function h on $[0,1]$ it follows from the definition of p-variation that $\|h \circ F\|_{[p]} \leq \|h\|_{[p]}$, while if F is continuous, and h is right-continuous at 0 and left-continuous at 1, then $\|h \circ F\|_{[p]} = \|h\|_{[p]}$. So we can and will assume in the rest of the proof of Theorem 1 that F is the $U[0,1]$ d.f. By (1.2), it will be enough to prove (1.4) for the seminorm $\|\cdot\|_{(p)}$.

Section 3 starts with a brief outline of our approach in proving (1.4) and a description of special yet straightforward piecewise linear approximations of α_n and B_n that are used throughout the paper. Section 4 reviews the KMT construction, while Section 5 provides the details of our proof of (1.4). Throughout we assume $p > 2$ unless otherwise specified.

This paper evolved from a Ph. D. dissertation [12], which proved (1.5) and (1.6).

2. Another Orlicz norm. For $0 < \gamma := 1/p < 1$ and any $x \geq 0$ let $\psi_\gamma(x) := g_p(x^\gamma)$. Then $\psi_\gamma(x) = \kappa_\gamma x$ for $0 \leq x \leq x_\gamma$ where $\kappa_\gamma := (e\gamma)^p$ and $x_\gamma := p^p$. For $x > x_\gamma$, $\psi_\gamma(x) = \exp(x^\gamma)$ and one can check that $\psi_\gamma''(x) > 0$, so ψ_γ is convex and a Young-Orlicz modulus (the value and first derivative, but not the second, are continuous at x_γ).

In the rest of the paper p and $\gamma := 1/p$ will be fixed and we will set $g := g_p$, $\psi := \psi_\gamma$.

It is easily seen that for a measurable real-valued function f on a measure space, $f \in \mathscr{L}_g$ if and only if $|f|^p \in \mathscr{L}_\psi$, with

$$\||f|^p\|_\psi = \|f\|_g^p. \tag{2.1}$$

In particular, for any stochastic process X such that $v_p(X)$ is measurable,

$$\|v_p(X)\|_\psi = \|\,\|X\|_{(p)}\|_g^p. \tag{2.2}$$

LEMMA 1. *Let $0 < \gamma = 1/p < 1$. Then for some $C = C_\gamma < \infty$, $g_p(y) \leq \exp(C_\gamma y) - 1$ for all $y > 0$.*

PROOF. For $y \geq p$ it will suffice to make $e^y \leq \exp(C_\gamma y)/2$ and $1 \leq \exp(C_\gamma y)/2$. Both hold if $C_\gamma \geq 1 + (\log 2)/p$. For $0 \leq y < p$ it suffices to make $C_\gamma y \geq \kappa_\gamma y^p$. Since $y \mapsto y^p$ is convex it will be enough to make $C_\gamma p \geq \kappa_\gamma p^p$. So we can set $C_\gamma := \max(\kappa_\gamma p^{p-1}, 1 + (\log 2)/p)$. $\square$

LEMMA 2. *Let $0 < \gamma = 1/p < 1$. Then there is a constant $K = K_\gamma$ depending only on γ such that whenever Y has a binomial $b(n, q)$ distribution with $nq \geq 1$, we have $\|Y^p\|_\psi \leq K(nq)^p$.*

PROOF. We apply (2.1) and the previous lemma. We have $Ee^{uY} = (qe^u + 1 - q)^n$, so for $s > 0$, $E\exp(C_\gamma Y/s) - 1 \leq 1$ if $q\exp(C_\gamma/s) + 1 - q \leq 2^{1/n}$. We have $e^x \leq 1 + 2x$ for $0 \leq x \leq 1$ and will take $s \geq C_\gamma$, so it will suffice if $1 + 2qC_\gamma/s \leq e^{\log 2/n}$, which will follow if $s \geq 2C_\gamma nq/(\log 2)$, or $s \geq 3C_\gamma nq$. Since $nq \geq 1$, $s \geq C_\gamma$ does hold. We thus have $\|Y\|_g \leq 3C_\gamma nq$ and the conclusion follows from (2.1), with $K = (3C_\gamma)^p$. $\square$

To deduce (1.5) from (1.4) we have:

LEMMA 3. *For any random variable Y, and $1 < p < \infty$, $E|Y| \leq (1 + p/e)\|Y\|_g$.*

PROOF. By homogeneity we can assume $\|Y\|_g = 1$. Thus by (1.3), $Eg(|Y|) \leq 1$. Then by Hölder's inequality

$$E|Y|1_{\{|Y| \leq p\}} \leq (E|Y|^p 1_{\{|Y| \leq p\}})^{1/p} \leq p/e,$$

while $E|Y|1_{\{|Y| > p\}} \leq 1$. The conclusion follows. $\square$

3. Piecewise linear interpolation and *p*-variation. To prove (1.4), we will define some piecewise linear interpolations, $[\alpha_n]_r$ and $[B_n]_r$, of α_n and B_n, then bound $\|\alpha_n - B_n\|_{(p)}$ above by

$$\|\alpha_n - [\alpha_n]_r\|_{(p)} + \|[\alpha_n]_r - [B_n]_r\|_{(p)} + \|[B_n]_r - B_n\|_{(p)}. \tag{3.1}$$

Specifically, for any $f: [0, 1] \mapsto \mathbb{R}$, $[f]_r$ will be the function equal to f at $k/2^{r+1}$ for $k = 0, 1, \ldots, 2^{r+1}$ and linear in between. If $f(0) = f(1) = 0$, as holds for $f = \alpha_n$ or B_n, then $[f]_r$ can be written as a sum as follows. For $j = 0, 1, \ldots,$ and $k = 0, 1, \ldots, 2^j - 1$, let $T_{j,k}$ be the "triangle function" such that $T_{j,k} = 0$ outside the interval $(k/2^j, (k+1)/2^j)$, $T_{j,k} = 1$ at the midpoint $(2k+1)/2^{j+1}$, and $T_{j,k}$ is linear in between. For each j and k, let

$$f_{j,k} := W_{j,k}(f) := f((2k+1)/2^{j+1}) - \tfrac{1}{2}\left[f(k/2^j) + f((k+1)/2^j)\right]. \tag{3.2}$$

Let $t(j) := 2^j - 1$ and $\phi_j(f) := \sum_{k=0}^{t(j)} f_{j,k} T_{j,k}$. Then

$$[f]_r = \sum_{j=0}^{r} \phi_j(f). \tag{3.3}$$

Clearly, if H is any monotone function and $1 \le p < \infty$,

$$v_p(H) = (\sup H - \inf H)^p \quad \text{and} \quad v_p(T_{j,k}) = 2 \tag{3.4}$$

for each j and k. To provide bounds for p-variations, the next fact will be useful.

LEMMA 4. *Let $1 \le p < \infty$ and let $f_1, f_2, \dots, f_K$ be real-valued functions, with supports included in $[a_1, b_1], [a_2, b_2], \dots, [a_K, b_K]$, respectively, such that* (i) $a_k < b_k$ *for* $k = 1, 2, \dots, K$; (ii) $b_{k-1} \le a_k$ *for each* $k = 2, 3, \dots, K$; *and* (iii) $f_k(a_k) = f_k(b_k) = 0$ *for every* $k = 1, 2, \dots, K$. *Let* $[a, b] = [a_1, b_K]$ *and* $f = \sum_{k=1}^{K} f_k$. *Then, we have*

$$v_p(f, [a, b]) \le 2^{p-1} \sum_{k=1}^{K} v_p(f_k, [a_k, b_k]). \tag{3.5}$$

PROOF. Let A be the collection of all the a_j and b_j. Take any p-variation sum $S_p(f) := S_p(f, \{x_i\}_0^s) := \sum_{i=1}^{s} |f(x_i) - f(x_{i-1})|^p$, where $x_0 \in [a, b]$ and $x_0 < x_1 < \cdots < x_s \in [a, b]$. If $(x_{i-1}, x_i) \cap A = \varnothing$, then either $f(x_{i-1}) = f(x_i) = 0$ or $(x_{i-1}, x_i) \subset [a_k, b_k]$ for some k. Therefore, we have

$$|f(x_i) - f(x_{i-1})|^p \le |f_k(x_i) - f_k(x_{i-1})|^p \le 2^{p-1} |f_k(x_i) - f_k(x_{i-1})|^p$$

for some k. Or, if $(x_{i-1}, x_i) \cap A =: A_i \ne \varnothing$, let $c_i := \min A_i$ and $d_i := \max A_i$. We have by Jensen's inequality

$$\begin{aligned} |f(x_i) - f(x_{i-1})|^p &\le 2^{p-1}(|f(x_i) - 0|^p + |0 - f(x_{i-1})|^p) \\ &= 2^{p-1}(|f(x_i) - f(d_i)|^p + |f(c_i) - f(x_{i-1})|^p). \end{aligned}$$

Either $c_i = b_k$ for some k, and then $a_k \le x_{i-1} < b_k$, or $|f(c_i) - f(x_{i-1})|^p = 0$. Similarly, either $d_i = a_k$ for some k, and then $a_k < x_i \le b_k$, or $|f(x_i) - f(d_i)|^p = 0$. The lemma then follows. □

For $K = 2$, disjoint supports are not needed for (3.5):

LEMMA 5. *Let f, g be any two real-valued functions on an interval $[a, b]$. Then*

$$v_p(f + g, [a, b]) \le 2^{p-1}(v_p(f, [a, b]) + v_p(g, [a, b])). \tag{3.6}$$

PROOF. For any x_{i-1}, x_i we have

$$|(f + g)(x_i) - (f + g)(x_{i-1})|^p \le 2^{p-1}(|f(x_i) - f(x_{i-1})|^p + |g(x_i) - g(x_{i-1})|^p)$$

by Jensen's inequality as in the previous proof, and the result follows. □

As mentioned above, our goal is to find a good bound for each term in (3.1). We will fix r at first, then an appropriate r will be chosen to have a good total bound. Finally, note that each term in (3.1) is measurable. In particular, $\|\alpha_n - B_n\|_{(p)}$ and $\|\alpha_n - [\alpha_n]_r\|_{(p)}$ are measurable since we can restrict the points t_i in p-variation sums to be rational.

4. The KMT construction–a review. If B is a Brownian bridge on $[0,1]$ and Z is a $N(0,1)$ variable independent of the process B, then $Y(t) := B(t) + Zt$ gives a Brownian motion on $[0,1]$. From (3.2), $W_{j,k}(B) = W_{j,k}(Y)$ for each j and k. Then it's easy to check by covariances that:

LEMMA 6. *Each $W_{j,k}(B)$ has a $N(0, 2^{-j-2})$ distribution, and the $W_{j,k}(B)$ are independent for all j and k.*

For $m = 0, 1, 2, \ldots$, let $H(t|m) := \max\{k \le m : \sum_{i=0}^{k-1} \binom{m}{i} 2^{-m} < t\}$, $0 < t < 1$, be the usual (left-continuous) inverse of the binomial distribution function $b(m, 1/2)$. Let Φ be the standard normal distribution function. Given n and a Brownian bridge B_n, and hence the $W_{j,k}(B_n)$, Komlós, Major and Tusnády [13] constructed random variables $U^*_{j,k}$ iteratively as follows. Let $U^*_{0,0} := n$. Next, let $U^*_{1,0} := H(\Phi(2W_{0,0}(B_n))|U^*_{0,0})$ and $U^*_{1,1} := U^*_{0,0} - U^*_{1,0}$. Then, given $U^*_{j-1,k}$, for $j \ge 2$, for each $k = 0, \ldots, 2^{j-1} - 1$, let

$$U^*_{j,2k} := H\left(\Phi\left(2^{(j+1)/2} W_{j-1,k}(B_n)\right) |U^*_{j-1,k}\right) \tag{4.1}$$

and $U^*_{j,2k+1} := U^*_{j-1,k} - U^*_{j,2k}$. Since each $2^{(j+1)/2} W_{j-1,k}(B_n)$ has law $N(0,1)$ by Lemma 6, Φ of it has law $U[0,1]$, so that $U^*_{j,2k}$ has law $b(U^*_{j-1,k}, 1/2)$, given $U^*_{j-1,k}$. It is easily seen that by interpreting $U^*_{j,k}$ as the number of points in the interval $I_{j,k} := (k/2^j, (k+1)/2^j]$ for each j and k and letting $j \to \infty$, the $U^*_{j,k}$ in fact define n points in the interval $(0,1)$, so that intervals $I_{j,k}$ containing the points have $U^*_{j,k} = 1$ for j large enough. Let these n points be the ordered sample $X^*_{(1)}, X^*_{(2)}, \ldots, X^*_{(n)}$ and define $X^*_1, X^*_2, \ldots, X^*_n$ by a random permutation π, independent of $X^*_{(\cdot)}$: $X^*_i := X^*_{(\pi(i))}$, $i = 1, 2, \ldots, n$. If we let $X_1, X_2, \ldots, X_n$ be a sample of independent uniform $(0,1)$ random variables and $U_{j,k}$ the number of X_i in $I_{j,k}$, then one can easily show that the $U_{j,k}$ and the $U^*_{j,k}$ have the same joint distribution. So $X^*_1, X^*_2, \ldots, X^*_n$ are indeed i.i.d. with uniform $(0,1)$ distribution. By virtue of this fact, we will drop the superscript '$*$' from now on and treat $U^*_{j,k}$ and $U_{j,k}$, and thus X^*_i and X_i, as the same.

In other words, the KMT construction defines a joint distribution of a Brownian bridge B_n and empirical process α_n as follows. Begin with B_n, which has continuous sample functions, and for it, extend (3.3) to

$$B_n = \sum_{j=0}^{\infty} \sum_{k=0}^{t(j)} W_{j,k}(B_n) T_{j,k}$$

(which by piecewise linear interpolation clearly converges uniformly on $[0,1]$). Let $F_n(0) := 0$ and $F_n(1) := 1$. Apply (4.1) repeatedly to get

$$F_n(1/2) := U^*_{1,0}/n,\ F_n(1/4) := U^*_{2,0}/n,\ F_n(3/4) := F_n(1/2) + U^*_{2,2}/n, \tag{4.2}$$

and so on. Then $F_n(k/2^j)$ for $j = 0, 1, \ldots,$ $k = 0, 1, \ldots, 2^j - 1$, have their correct joint distribution, and $F_n(t)$ for $0 \le t \le 1$ is then defined by monotonicity and right-continuity. Let $\alpha_n := n^{1/2}(F_n - F)$.

Theorem 8.1 of Major [16] implies that if Z has law $N(0,1)$, then $Y = H(\Phi(Z)|m)$ minimizes $E(|m^{-1/2}(2Y - m) - Z|^p)$ among all Y with law $b(m, 1/2)$. This motivates the choice (4.1). To prove Theorem 1, we need to define a distribution of $\{(\alpha_n, B_n)\}_{n=1}^{\infty}$, although the joint distribution for different n has no effect on (1.4), (1.5) or (1.6). One way is as follows. Let $X_1, X_2, \dots$ be i.i.d. $U[0,1]$ and define α_n from $X_1, \dots, X_n$ as usual. For each n let $((X_1, \dots, X_n), B_n)$ have the KMT joint distribution as just defined. Then let B_n for $n \geq 1$ be conditionally independent given $\{X_i\}_{i=1}^{\infty}$ (the Vorob'ev [22], Berkes and Philipp [1] method). Good conditional distributions exist because the spaces (of sequences and $C[0,1]$) are complete separable metric spaces.

5. Piecewise linear approximations. First we will give a bound for $\|v_p(B - [B]_r)\|_{\psi}$, where as always $\psi := \psi_{\gamma}$, for any Brownian bridge B, for example, $B = B_n$. For a function f from a closed interval $[c,d]$ into $\mathbf{R}$, let $f_{[c,d]}$ be the linear function on $[c,d]$ which equals f at c and at d. On $[c,d]$ we define $\Delta(f;[c,d])(\cdot) := f - f_{[c,d]}$. The following two lemmas are straightforward to check by covariances. In both lemmas, as in (3.2) and Lemma 6, a Brownian bridge can again be replaced by a Brownian motion Y, simplifying the covariances.

LEMMA 7. *Suppose $t_1 \in (s_1, s_2) \subset [0,1]$ and $t_2 \in (s_3, s_4) \subset [0,1]$. Then $\Delta(B;[s_1,s_2])(t_1)$ and $\Delta(B;[s_3,s_4])(t_2)$ are independent if $s_2 \leq s_3$ or $t_2 \notin (s_1, s_2)$ $\subset [s_3, s_4]$.*

LEMMA 8. *Let $M(t), 0 \leq t \leq 1$, be a Brownian bridge. For any fixed $0 \leq u < v \leq 1$, define*

$$B(t;u,v) := (v-u)^{-1/2}[M(tv + (1-t)u) - tM(v) - (1-t)M(u)]$$

for $0 \leq t \leq 1$. Then $B(t;u,v)$ is a Brownian bridge. Specifically, for $0 \leq t \leq 1$, let $B_{j,k}(t) := B(t) - [B]_{j-1}(t)$ if $t \in [k/2^j, (k+1)/2^j]$, and otherwise $B_{j,k}(t) := 0$. Let $B^0_{j,k}(t) := (2^j)^{1/2} B_{j,k}((k+t)/2^j)$ for $0 \leq t \leq 1$. Then $B^0_{j,k}$ for $j = 0, 1, \dots$ and $k = 0, 1, \dots, 2^j - 1$ are Brownian bridges.

Now we are in a position to bound the last term in (3.1).

LEMMA 9. *For any $p \in (2, \infty)$, there exists a constant $C_B(p)$ such that for $r = 1, 2, \dots$, $\| \|B - [B]_{r-1}\|_{(p)}\|_g \leq C_B(p)(2^r)^{-(p-2)/(2p)}$.*

PROOF. First, $v_p(B,[0,1]) < \infty$ a.s. since B a.s. satisfies a Hölder condition of any order $< 1/2$, [19], (9.11), or from the more precise results of Lévy [14], page 172, or S. J. Taylor [21]. Let $C_p := \| \|B\|_{(p)}\|_g$. By the Landau-Shepp-Marcus-Fernique theorem (specifically Fernique [10]; note that $\|\cdot\|_{(p)}$ is measurable but not separable), C_p is finite. Then by (2.2), $\|v_p(B,[0,1])\|_{\psi} =$

C_p^p. Recall that $t(r) := 2^r - 1$. By Lemma 4, for $r = 1, 2, \ldots$, we have

$$
\begin{aligned}
v_p(B - [B]_{r-1}, [0,1]) &= v_p\left(\sum_{k=0}^{t(r)} B_{r,k}, [0,1]\right) \\
&\le 2^{p-1} \sum_{k=0}^{t(r)} v_p\left(B_{r,k}, \left[\frac{k}{2^r}, \frac{(k+1)}{2^r}\right]\right) \\
&= 2^{p-1} \sum_{k=0}^{t(r)} (2^r)^{-p/2} v_p(B^0_{r,k}, [0,1]) \\
&= 2^{p-1}(2^r)^{-(p-2)/2} \sum_{k=0}^{t(r)} \frac{1}{2^r} v_p(B^0_{r,k}, [0,1]).
\end{aligned}
\tag{5.1}
$$

Since $p > 2$ and $B^0_{j,k}(t)$, $k = 0, 1, \ldots, 2^j - 1$, are all Brownian bridges, we have

$$\|v_p(B - [B]_{r-1}, [0,1])\|_\psi \le 2^{p-1}(2^r)^{-(p-2)/2} C_p^p.$$

By (2.2), the conclusion follows. □

Next we approximate the empirical process.

LEMMA 10. *For any $p > 2$ and $r = 1, 2, \ldots$, with $n/2^r \ge 1$, and K_γ from Lemma* 2,

$$\|v_p(\alpha_n - [\alpha_n]_{r-1})\|_\psi \le 4^p K_\gamma n^{p/2} 2^{-r(p-1)}.$$

PROOF. Since $\alpha_n - [\alpha_n]_{r-1} = 0$ at each point $k/2^r$, we have by Lemma 4

$$v_p(\alpha_n - [\alpha_n]_{r-1}) \le 2^{p-1} \sum_{k=0}^{t(r)} v_p\left(\alpha_n - [\alpha_n]_{r-1}, \left[\frac{k}{2^r}, \frac{k+1}{2^r}\right]\right).$$

Now note that the map $f \mapsto [f]_{r-1}$ is linear and that $F \equiv [F]_{r-1}$. Thus $\alpha_n - [\alpha_n]_{r-1} \equiv \sqrt{n}(F_n - [F_n]_{r-1})$. Recall that $U_{r,k}$ is the number of points X_i in the interval $I_{r,k}$, which almost surely equals the number in the closure $[k/2^r, (k+1)/2^r]$. On that interval, by Lemma 5, $v_p(F_n - [F_n]_{r-1}) \le 2^{p-1}(v_p(F_n) + v_p([F_n]_{r-1}))$. Applying monotonicity (3.4) then gives

$$v_p(\alpha_n - [\alpha_n]_{r-1}) \le 4^p n^{-p/2} \sum_{k=0}^{t(r)} U^p_{r,k}.$$

Since $U_{r,k}$ has a binomial $b(n, 1/2^r)$ distribution for each k, we can take $\|\cdot\|_\psi$ of both sides and apply Lemma 2. The conclusion follows. □

Now we approximate the difference $B_n - \alpha_n$. We have for $r = 1, 2, \ldots$, by (3.2) and (3.3),

$$[B_n]_r - [\alpha_n]_r = \sum_{j=0}^{r} \sum_{k=0}^{t(j)} W_{j,k}(B_n - \alpha_n) T_{j,k}. \tag{5.2}$$

Recall the inverse $H(\cdot|N)$ of the the binomial $b(N, 1/2)$ distribution function. For a $N(0,1)$ variable Y let $\beta_N := H(\Phi(Y)|N) - N/2$. Then according to a lemma of Tusnády whose first published proof was given as far as we know by Bretagnolle and Massart [2], Lemma 4,

$$\left|\beta_N - \frac{\sqrt{N}}{2}Y\right| \le 1 + Y^2/8.$$

Letting $N := U^*_{j-1,k}$ and $Y := Y_{j,k} := 2^{(j+1)/2}W_{j-1,k}(B_n)$ we get from (4.1) and (3.2)

$$W_{j-1,k}(\alpha_n) = n^{-1/2}\left(U^*_{j,2k} - \frac{N}{2}\right) = \beta_N/\sqrt{n}.$$

Then since $EN = n/2^{j-1}$,

$$\sqrt{n}|W_{j-1,k}(\alpha_n - B_n)| \le 1 + \tfrac{1}{8}Y^2_{j,k} + \tfrac{1}{2}|Y_{j,k}| \cdot |\sqrt{N} - \sqrt{EN}|. \tag{5.3}$$

Via the inequality $(a+b+c)^p \le 3^{p-1}(a^p + b^p + c^p)$ we will treat the three terms on the right of (5.3) separately.

For a standard normal variable Y such as $Y_{j,k}$, $\||Y|^{2p}\|_\psi$ is a finite constant η_p depending on p. For the last term in (5.3) we will use

$$|Y_{j,k}|^p \cdot |\sqrt{N} - \sqrt{EN}|^p \le |Y_{j,k}|^{2p} + |\sqrt{N} - \sqrt{EN}|^{2p}. \tag{5.4}$$

For $t > 0$ we have by the relation $\psi_\gamma(x) \equiv g_p(x^{1/p})$ and Lemma 1 that

$$E\psi_\gamma(|\sqrt{N} - \sqrt{EN}|^{2p}/t) \le E\exp(C_\gamma(\sqrt{N} - \sqrt{EN})^2/t^{1/p}) - 1. \tag{5.5}$$

We have the following:

LEMMA 11. *Let N be a binomial $b(n,q)$ random variable and $\phi(s) := \phi_{n,q}(s) := E\exp(s(\sqrt{N} - \sqrt{EN})^2)$. Then for $0 < s < 1/2$ and all n, q, $\phi(s) \le e^s/\sqrt{1-2s}$.*

REMARKS. The supremum over (n,q) of $\phi(s)$ is $+\infty$ for $s > 1$, as can be seen by taking the Poisson limit $n \to \infty$, $q \to 0$, $nq \to \lambda > 0$, where the $N = 0$ term converges to $e^{-\lambda+s\lambda}$, then letting $\lambda \to \infty$. We have a $3\frac{1}{2}$-page proof that for $q \le 1/2$, $\phi(1)$ is uniformly bounded, by a large constant. Thus, Lemma 11 is not efficient for s close to 1/2. A referee and an Associate Editor pointed out to us that smaller values of s can yield better bounds on $\||\sqrt{N} - \sqrt{EN}|^{2p}\|_\psi$ and suggested most of the following proof.

Let $Z := (\sqrt{N} - E\sqrt{N})^2$. Then $(\sqrt{N} - \sqrt{EN})^2 - Z \le EN - (E\sqrt{N})^2$ since $E\sqrt{N} \le \sqrt{EN}$. Then $EN - (E\sqrt{N})^2 \le EN - (EN)^3/E(N^2)$, in other words $(EN)^3 \le (E\sqrt{N})^2 E(N^2)$, by Hölder's inequality for $f = N^{1/3} \in \mathscr{L}^{3/2}$, $g = N^{2/3} \in \mathscr{L}^3$. Next,

$$EN - (EN)^3/E(N^2) \le \mathrm{Var}(N)/EN \le 1$$

as follows from $[E(N^2) - (EN)^2]^2 \geq 0$. Thus if N' and N are i.i.d.,

$$\phi(s) \leq e^s E \exp\left(s\left(\sqrt{N} - E\sqrt{N}\right)^2\right) \leq e^s E \exp\left(s\left(\sqrt{N} - \sqrt{N'}\right)^2\right)$$

by Jensen's inequality, noting that $E\sqrt{N'} = E\sqrt{N}$. Let $S := N + N'$ and $D := N - N'$. Then

$$\left(\sqrt{N} - \sqrt{N'}\right)^2 \equiv S\left(1 - \sqrt{1 - (D/S)^2}\right) \leq D^2/S$$

as is easily checked. Now write $D = N - N' = \sum_{i=1}^n \xi_i$ where $\xi_i = \eta_i - \eta_i'$ and η_i, η_i', $i = 1, ..., n$, are i.i.d. Bernoulli (q). Then $\eta_i \equiv \eta_i^2$, $\eta_i' \equiv {\eta_i'}^2$, so $S \geq \sum_{i=1}^n \xi_i^2$. Conditional on ξ_i^2 for $i = 1, \ldots, n$, D is equal in distribution to $\sum_{i=1}^n \varepsilon_i \xi_i$ where ε_i are Rademacher functions, equal to ± 1 with probability $1/2$ each, independent of each other and the ξ_j. We will show that

$$\phi(s) \ \leq \ e^s \sup\left\{E \exp\left(s\left[\sum_{i=1}^n \alpha_i \varepsilon_i\right]^2\right) : \sum_{i=1}^n \alpha_i^2 \leq 1\right\}. \tag{5.6}$$

To see this, condition on the ξ_i and if not all ξ_i are 0, let $\alpha_i := \xi_i / (\sum_{j=1}^n \xi_j^2)^{1/2}$. If all ξ_i are 0 let $\alpha_i = 0$ for all i.

We next need the following known bound, which is asymptotically sharp, letting $\alpha_i = 1/\sqrt{n}$, $n \to \infty$.

LEMMA 12. *For* $0 < s < 1/2$, *a* $N(0, 1)$ *random variable* Z, *any* α_i *with* $\sum_{i=1}^n \alpha_i^2 \leq 1$, *and independent Rademacher* ε_i,

$$E \exp\left(s\left(\sum_{i=1}^n \alpha_i \varepsilon_i\right)^2\right) \leq E \exp(sZ^2).$$

PROOF. Let $Y := \sum_{i=1}^n \alpha_i \varepsilon_i$. It suffices to show that $E(Y^{2k}) \leq E(Z^{2k})$ for $k = 0, 1, \ldots$. We have from $Ee^{tZ} = \exp(t^2/2)$ that $E(Z^{2k}) = (2k)!/(2^k k!)$. Thus the lemma follows from Whittle [23], Theorem 1. The lemma as stated is also a special case of Pinelis [20], Corollary 2.7; see also [11], Theorem 1.1.

Lemma 12 is proved. □

By Lemma 12 and (5.6), since $E \exp(sZ^2) = 1/\sqrt{1 - 2s}$, Lemma 11 follows. □

Now returning to the proof of Theorem 1, by (5.5) we have $\||\sqrt{N} - \sqrt{EN}|^{2p}\|_\psi \leq t$ if $\phi(C_\gamma/t^{1/p}) \leq 2$. Lemma 11 gives $\phi(0.28) < 2$. Thus

$$\||\sqrt{N} - \sqrt{EN}|^{2p}\|_\psi \leq (3.6C_\gamma)^p, \tag{5.7}$$

a constant depending only on p. Summing up the terms from (5.3) with (5.4), there is some $A_p < \infty$ such that $\|(\sqrt{n}|W_{j,k}(\alpha_n - B_n)|)^p\|_\psi \leq A_p$ for each j

and k. Thus for each j, by Lemma 4 and (3.4),

$$\left\| v_p \left(\sum_{k=0}^{t(j)} W_{j,k}(\alpha_n - B_n) T_{jk} \right) \right\|_{\psi} \le 2^{j+p} A_p n^{-p/2}.$$

Hence by (2.2), for some constant $B_p < \infty$,

$$\left\| \left\| \sum_{k=0}^{t(j)} W_{j,k}(\alpha_n - B_n) T_{jk} \right\|_{(p)} \right\|_g \le 2^{j/p} B_p n^{-1/2}.$$

Now, for a given n, we use the KMT construction (4.1) and (4.2) to construct α_n from B_n. We choose $r = r(n) = \lfloor \log_2(n) \rfloor$, where $\lfloor x \rfloor :=$ the largest integer $\le x$. Then $n/2^r \ge 1$ and

$$\left\| \left\| \sum_{j=0}^{r(n)} \sum_{k=0}^{t(j)} W_{j,k}(\alpha_n - B_n) T_{jk} \right\|_{(p)} \right\|_g \le B_p n^{-1/2} \sum_{j=0}^{r(n)} 2^{j/p} \le B_p n^{-1/2} n^{1/p} / (1 - 2^{-1/p}). \tag{5.8}$$

By Lemma 9 we have

$$\| \| B_n - [B_n]_r \|_{(p)} \|_g \le C_B(p)(2^{r+1})^{(2-p)/(2p)} \le C_B(p) n^{(2-p)/(2p)} \tag{5.9}$$

for some constant $C_B(p)$. Next, by (5.2) and (5.8) we have

$$\| \| [B_n]_r - [\alpha_n]_r \|_{(p)} \|_g \le C_D(p) n^{(2-p)/(2p)} \tag{5.10}$$

for some constant $C_D(p)$. Third, by Lemma 10 and (2.2) we have

$$\| \| \alpha_n - [\alpha_n]_r \|_{(p)} \|_g \le C_\alpha(p) n^{(2-p)/(2p)} \tag{5.11}$$

for some constant $C_\alpha(p)$. Combining (5.9), (5.10), and (5.11) as in (3.1), now (1.4) and Theorem 1 are proved. □

Acknowledgments. We thank an anonymous referee and an Associate Editor for helpful comments, Joel Zinn for suggestions about exponential bounds and David Mason for sending us a copy of [17].

REFERENCES

[1] BERKES, I. and PHILIPP, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrsch. Verw. Gebiete* **41** 115–137.

[2] BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256.

[3] CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, Chichester.

[4] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

[5] DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

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

[7] DUDLEY, R. M. (1994). The order of the remainder in derivatives of composition and inverse operators for p-variation norms. *Ann. Statist.* **22** 1–20.

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

[9] DUDLEY, R. M. and NORVAIŠA, R. (1999). *Differentiability of Six Operators on Nonsmooth Functions and p-Variation. Lecture Notes in Math.* **1703**. Springer, New York.

[10] FERNIQUE, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Sér. A* **270** 1698–1699.

[11] FIGIEL, T., HITCZENKO, P., JOHNSON, W. B., SCHECHTMAN, G. and ZINN, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997–1027.

[12] HUANG, Y.-C. (1994). Empirical distribution function statistics, speed of convergence, and p-variation. Ph. D. dissertation, MIT.

[13] 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.

[14] LÉVY, P. (1937). *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris.

[15] LUXEMBURG, W. A. J. and ZAANEN, A. C. (1956). Conjugate spaces of Orlicz spaces. *Indag. Math.* **18** *Akad. Wetensch. Amsterdam Proc. Ser. A* **59**. 217–228.

[16] MAJOR, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8** 487–517.

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

[18] MASON, D. M. and VAN ZWET, W. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.

[19] PALEY, R. E. A. C., WIENER, N. and ZYGMUND, A. (1933). Notes on random functions. *Math. Z.* **37** 647–668. [Reprinted (1976) *Norbert Wiener: Collected Works* (P. Masani, ed.) **1** 526–557. MIT Press.

[20] PINELIS, I. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* **22** 357–368.

[21] TAYLOR, S. J. (1972). Exact asymptotic estimates of Brownian path variation. *Duke Math. J.* **39** 219–241.

[22] VOROB'EV, N. N. (1962). Consistent families of measures and their extensions. *Theory Probab. Appl.* **7** 147–163.

[23] WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.

186 BELLOWS HILL ROAD
CARLISLE, MASSACHUSETTS 01741
E-MAIL: ychuang@alum.mit.edu

DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139
E-MAIL: rmd@math.mit.edu