A Martingale Inequality for the Empirical Process

Jon A. Wellner


Stable URL:
http://links.jstor.org/sici?sici=0091-1798%28197704%295%3A2%3C303%3AAMIFTE%3E2.0.CO%3B2-E

*The Annals of Probability* is currently published by Institute of Mathematical Statistics.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ims.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
A MARTINGALE INEQUALITY FOR THE
EMPIRICAL PROCESS

BY JON A. WELLNER

University of Washington

A martingale inequality for the \( \rho_q \) distance from the uniform empirical process to zero is proved, compared with other inequalities for the process, and used to establish a law of the iterated logarithm.

1. Introduction. For \( n \geq 1 \) let \( \xi_1, \ldots, \xi_n \) be i.i.d. uniform \((0, 1)\) rv's and let \( \Gamma_n \) denote their empirical df. The uniform empirical process \( U_n \) is the process on \([0, 1]\) defined by \( U_n = n(t(\Gamma_n) - t) \) where \( I \) denotes the identity function \( I(t) = t \). If \( q \) is a nonnegative function approaching zero at the endpoints of the interval \([0, 1]\) and \( x, y \) are functions on \([0, 1]\), the \( \rho_q \)-metric is defined by

\[
\rho_q(x, y) = \rho(x/q, y/q) = \sup_{0 < t < 1} |x(t) - y(t)|/q(t)
\]

where \( \rho \) denotes the usual supremum metric. The convergence of \( U_n \) with respect to certain of these \( \rho_q \)-metrics has become an important tool in the study of linear rank statistics [11], linear combinations of order statistics [12], and sample quantiles [15].

Our main object here is to prove a martingale type inequality for the \( \rho_q \) distance from \( U_n \) to zero and show how it may be combined with a Berry–Esseen theorem of Katz [7] to prove a law of the iterated logarithm for \( U_n \). Theorem 1 presents the new inequality; Corollaries 1 and 2 relate it to inequalities for \( U_n \) due to Pyke and Shorack [11], and Dvoretsky, Kiefer and Wolfowitz [3]. Finally, the power of the new inequality is illustrated in the proof of Theorem 2. This theorem is in the spirit of Chover’s proof [2] of Strassen’s law of the iterated logarithm [14] which requires \( 2 + \delta \) moments with \( \delta > 0 \) as opposed to Strassen’s proof which requires only second moments. While the approach taken in the proof of Theorem 2 yields a result which is weaker than a theorem of James [6], it has the virtue of simplicity. In [15] we use the inequality of Theorem 1 to establish a different type of strong limit theorem for \( U_n \).

2. The Inequality. Our proof of Theorem 1 will rely upon the fact that the process \( U_n(t)/(1 - t), 0 \leq t < 1 \) is a martingale (cf. [8]) in conjunction with the following lemmas. Lemma 1 is a special case of Lemma 1 of [13]; Lemma 2 is a consequence of Doob’s martingale inequality.

Let \( \{X_j, j = 1, \ldots, m\} \) be arbitrary rv’s and let \( \{r_j, j = 1, \ldots, m\} \) be positive and nondecreasing real numbers; for \( k = 1, \ldots, m \) set

\[
S_k = \sum_{j=1}^{k} X_j, \quad D_k = \sum_{j=1}^{k} (X_j/r_j).
\]
Lemma 1. \( \max_{1 \leq k \leq m} |S_k|/r_k \leq 2 \max_{1 \leq k \leq m} |D_k| \).

Proof. Let \( \Delta r_j = r_j - r_{j-1}, \Delta D_j = D_j - D_{j-1}, j = 2, \ldots, m, \Delta r_1 = r_1, \Delta D_1 = D_1. \) Then, by writing \( X_j = r_j \Delta D_j = \sum_{i=1}^{j} \Delta r_i \Delta D_j \) and interchanging the order of summation, one obtains \( S_k = \sum_{i=1}^{k} \Delta r_i (D_k - D_{i-1}) \). Hence \( |S_k|/r_k \leq \max_{1 \leq k \leq m} |D_k - D_{i-1}| \) and this implies the statement of the lemma.

Remark 1. If \( \{X_j, j = 1, \ldots, m\} \) is a martingale-difference sequence then \( \{D_k, k = 1, \ldots, m\} \) is a martingale transform and under the present conditions is itself a martingale (confer [1]).

To state the second lemma, let \( \{T_k, \mathcal{S}_k, k = 1, \ldots, m\} \) be a positive submartingale.

Lemma 2. For all \( \lambda > 0 \)

\[ P(\max_{1 \leq k \leq m} T_k \geq 2\lambda) \leq \frac{\lambda^{-1} E(T_m 1_{\{T_m \geq \lambda\}})}{2\lambda} \cdot \]

Proof. Let \( M_m = \max_{1 \leq k \leq m} T_k \). From Doob's martingale inequality,

\[ 2\lambda P(M_m \geq 2\lambda) \leq E(T_m 1_{\{M_m \geq 2\lambda\}}) \leq E(T_m 1_{\{M_m \geq 2\lambda, T_m > \lambda\}}) + E(T_m 1_{\{M_m > 2\lambda, T_m < \lambda\}}) \leq E(T_m 1_{\{T_m > \lambda\}}) + \lambda P(M_m \geq 2\lambda). \]

Let \( \mathcal{C} \) denote the set of positive continuous functions on \([0, 1]\) which are nondecreasing on \([0, \frac{1}{2}]\), symmetric about \( \frac{1}{2} \), and have \( \int_0^1 q^{-2} dq < \infty \). The functions \( q(t) = [(t(1 - t))]^{1-s} \) with \( 0 < \delta \leq \frac{1}{2} \) are all in \( \mathcal{C} \); so are the functions \( q(t) = [(t(1 - t))]^{1-s} \) with \( \delta > 0 \).

Theorem 1. Let \( q \in \mathcal{C} \) and \( \theta \in (0, \frac{1}{2}] \). Then for all \( \lambda > 0 \)

\[ P \left( \sup_{0 < t \leq \theta} \frac{|U_n(t)|}{q(t)} \geq 4\lambda \right) \leq \lambda^{-1} E(|T_n| 1_{\{|T_n| \geq \lambda\}}) \]

where \( T_n = n^{-1} \sum_{i=1}^{n} Y_i \), the sum of the i.i.d. rv's

\[ Y_i = \frac{1}{q_\theta(\xi_i)} - \frac{1}{q_\theta} \int_{0}^{1} \frac{1}{q_\theta} \frac{q^{-1}_{\theta}}{q} \, dq \]

\( i = 1, \ldots, n \) with \( 1/q_\theta = q^{-1}_{\theta} 1_{(0, \theta]} \). Furthermore, the \( Y_i \)'s have \( E(Y_i) = 0 \) and \( \text{Var}(Y_i) = \frac{q^{-2}}{q^{-1}_{\theta}} \, dq \).

Proof. Let \( W_n(t) = U_n(t)/(1 - t); W_n \) is a martingale in \( t \) for each fixed \( n \) (cf. [8] or [10], page 42) with covariance \( s/(1 - s), s \leq t \). Also let \( r(t) = q(t)/(1 - t) \).

For \( m = 2^k, h \geq 1 \) an integer, and \( 1 \leq k \leq m \), define \( X_k = W_n(k/m) - W_n((k - 1)/m) \) and \( r_k = r(k/m) \). Note that the \( r_k \)'s are nondecreasing for \( 1 \leq k \leq [m\theta] \).

Then, using Lemmas 1 and 2

\[ P \left( \sup_{0 < t \leq \theta} \frac{|U_n(t)|}{q(t)} > 4\lambda \right) = \lim_{h \to \infty} P \left( \max_{1 \leq k \leq [m\theta]} \frac{|W_n(k/m)|}{r_k} > 4\lambda \right) \]

\[ = \lim_{h \to \infty} P \left( \max_{1 \leq k \leq [m\theta]} \frac{|\sum_{i=1}^{k} X_i|}{r_k} > 4\lambda \right) \leq \lim_{h \to \infty} P(\max_{1 \leq k \leq [m\theta]} |\sum_{i=1}^{k} (X_i/r_i)| > 2\lambda) \leq \lim_{h \to \infty} \lambda^{-1} E(|\sum_{i=1}^{[m\theta]} (X_i/r_i)| 1_{\{|\sum_{i=1}^{[m\theta]} (X_i/r_i)| > \lambda\}}) \]
where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 2 since, by Remark 1, \( \sum_{j=1}^{k} (X_j / r_j) \), \( k = 1, \ldots, [m\theta] \) is a martingale. We now show that

\[
T_n \equiv \lim_{h \to 0} \sum_{j=1}^{[m\theta]} (X_j / r_j)
\]

exists for each \( \omega \in \Omega \) and equals \( T_n \) of the statement of the theorem. Write \( W_n = n^{-i} \sum_{i=1}^{n} Q_i \) with \( Q_i(t) = (1_{c_{i1}}(\xi_i) - t) / (1 - t) \). Using this together with the definition of \( X_j \) in (3) and interchanging the order of summation one obtains

\[
T_n = n^{-i} \sum_{i=1}^{n} \lim_{h \to 0} \sum_{j=1}^{[m\theta]} \left\{ Q_i \left( \frac{j}{m} \right) - Q_i \left( \frac{j-1}{m} \right) \right\} / r_j.
\]

Since the \( Q_i \) are i.i.d. processes, it suffices to show that this last limit exists for \( i = 1 \) and equals \( Y_i \) of the statement of the theorem. For \( s < t \)

\[
Q_i(t) - Q_i(s) = \frac{1}{(1 - t)} 1_{[s,1]}(\xi_i) - \frac{(t - s)}{(1 - s)(1 - t)} 1_{[s,1]}(\xi_i)
\]

and hence, taking \( t = j/m \), \( s = (j - 1)/m \) and using the monotone convergence theorem

\[
\sum_{j=1}^{[m\theta]} \left\{ Q_1 \left( \frac{j}{m} \right) - Q_1 \left( \frac{j-1}{m} \right) \right\} / r_j
\]

\[
= \sum_{j=1}^{[m\theta]} \frac{1_{[j-1/m,j/m]}(\xi_i)}{(1 - (j/m))r_j} - \frac{1}{m} \sum_{j=1}^{[m\theta]} 1_{[j-1/m,1]}(\xi_i)/(1 - (j-1)/m)(1 - j/m)r_j
\]

\[
\rightarrow \frac{1}{q_\theta(\xi_i)} \int_0^{\xi_i} \frac{1}{(1 - \theta)} q_\theta \, dl \quad h \to \infty
\]

\[
= Y_i.
\]

Now the first assertion of the theorem follows if the limit on \( h \) and integration with respect to \( P \) in the last line of (2) may be interchanged; this follows easily from standard theorems (e.g., [9], page 52) since the sequence \( \{ \sum_{j=1}^{[m\theta]} (X_j / r_j) \}, m \geq 1 \) is bounded in \( L_2 \) and hence uniformly integrable.

That \( E(Y_i) = 0 \) and \( \text{Var}(Y_i) = \int q^{-3} \, dl \) is easily verified by straightforward computation. \( \square \)

Remark 2. The process \( \{ B_n(t) = (1 + t)U_n(t/(1 + t)), 0 \leq t < \infty \} \) is also a martingale and has the same covariance as Brownian motion, \( E(B_n(s)B_n(t)) = s \wedge t \). Note that the random variable \( T_n \) may be written in terms of the process \( B_n \) as

\[
T_n = \int_0^\theta f \, dB_n
\]

where \( f(t) = [(1 + t)q(t/(1 + t))]^{-1}, \theta^* = \theta/(1 - \theta) \), and the integral is to be interpreted as an improper (since \( f \) is unbounded near zero) Riemann–Stieltjes integral. In analogy with stochastic integrals (of deterministic \( L_2 \) functions) with respect to Brownian motion ([5], page 21) it is not surprising that

\[
E(T_n^2) = \int_0^\theta f^2 \, dl = \int_0^\theta q^{-3} \, dl.
\]
Remark 3. For \( q \in C \), \( \int_0^\infty q^{-2} \, dl \to 0 \) as \( \theta \to 0 \) and hence \( \text{Var}(Y_i) \) can be made arbitrarily small by choosing \( \theta \) small.

Remark 4. If \( \int_0^\infty q^{-2-\delta} \, dl < \infty \) for some \( \delta > 0 \), then the \( C_r \) and Jensen inequalities may be used to show that \( E[Y_i|S_{i+1}] \leq C(\delta) \int_0^\infty q^{-2-\delta} \, dl < \infty \) with \( C(\delta) = 3 \cdot 2^{1+\delta} \).

By use of the Birnbaum–Marshall inequality it may be shown that (confer [10], page 41 and Lemma 2.2 of [11])

\[
P\left( \sup_{0 < t < \delta} \frac{|U_n(t)|}{q(t)} \geq \lambda \right) \leq \lambda^{-2} \int_0^\infty q^{-2} \, dl.
\]

When \( q \equiv 1, \theta = 1 \) Dvoretzky, Kiefer and Wolfowitz [3] proved that

\[
P(\sup_{0 < t < 1} |U_n(t)| \geq \lambda) \leq Ce^{-2\lambda^2}
\]

for some absolute constant \( C > 0 \). The following corollaries of Theorem 1 shows that (1) implies versions of the inequalities (4) and (5) which differ from them by constant factors.

Corollary 1. For \( q \in C \) and \( \lambda > 0 \)

\[
P\left( \sup_{0 < t < \delta} \frac{|U_n(t)|}{q(t)} \geq \lambda \right) \leq 16\lambda^{-2} \int_0^\infty q^{-2} \, dl.
\]

Proof. This follows immediately from (1) and \( E(T_n) = \int_0^\infty q^{-2} \, dl \).

Corollary 2. For all \( \lambda > 0 \)

\[
P(\sup_{0 < t < 1} |U_n(t)| \geq \lambda) \leq 8(2\pi)^{-1/2}e^{-\lambda^{-2}/2}.
\]

Proof. For \( q \equiv 1 \) the inequality (1) holds for any \( 0 < \theta < 1 \) since \( r(t) = (1 - n)^{-1} \) is increasing on \([0, 1)\). Letting \( \theta \to 1 \) the \( Y_i \) of Theorem 1 become

\[Y_i = 1 - \int_0^\infty (1 - t)^{-1} \, dt = 1 + \log(1 - \xi_i) = -(\exp(1) - 1)\]

where \( \exp(1) \) denotes an exponential rv with scale parameter one. Therefore \( T_n = -n^{-\frac{1}{2}}(G_n - n) \) where \( G_n \) denotes a gamma \( (n, 1) \) rv and the right side of (1) may be computed exactly:

\[E(T_n)_{1 \leq (T_n < 2)} = \frac{n^{n+1}e^{-n}}{n!} \left\{ (1 - \lambda_n)e^{\lambda_n n} + (1 + \lambda_n)e^{-\lambda_n n} \right\}
\]

where \( \lambda_n \equiv \lambda n^{-1} \). Use of Stirling’s formula and the elementary inequalities \( \log(1 - x) \leq -x - \frac{1}{2}x^2 \) and \( \log(1 + x) \leq x - \frac{5}{6}x^2 \), \( 0 \leq x \leq \frac{1}{4} \) (recall that \( \sup_{0 < t < 1} |U_n(t)| \leq n^4 \) and hence we need only consider \( 4\lambda \leq n^4 \) or \( \lambda_n \leq \frac{1}{4} \)) to bound this last expression yields

\[P(\sup_{0 < t < 1} |U_n(t)| \geq 4\lambda) \leq (2/\pi)^{1/2}e^{-(4\lambda)^2/(8n^2)}
\]

which implies (7).
2. However, (1) holds for all \( q \in \mathcal{Q} \) and is more powerful than (4). In the following we use (1) to establish a law of the iterated logarithm for \( U_n \).

3. A law of the iterated logarithm for \( U_n \). Let \( b_n = (2 \log \log n)^{1/2} \) and let

\[
\mathcal{B} = \{ f \in C[0, 1] : f(0) = 0 = f(1), f = \int_0^1 f' \, dI, \int_0^1 (f')^2 \, dI \leq 1 \}.
\]

Finkelstein [4] has shown that with probability one the sequence \( \{ U_n/b_n, n \geq 1 \} \) is relatively compact with respect to the supremum metric \( \rho \) and has limit set \( \mathcal{B} \). James [6] extended this conclusion to the metrics \( \rho_q \) for a class of functions \( q \) which is slightly larger than \( \mathcal{Q} \); he shows that finiteness of the integral

\[
\int_0^1 q^{-2} \log \log (I(1 - I))^{-1} \, dI
\]

is both necessary and sufficient for this convergence.

Here we use Theorem 1 in conjunction with the Berry–Esseen estimate of Katz [7] to establish the relative compactness of \( U_n/b_n \) with respect to \( \rho_q \) for a class of functions \( q \) which is slightly smaller than \( \mathcal{Q} \). The proof is in the spirit of Chover’s [2] proof of Strassen’s law of the iterated logarithm under the assumption of a finite \( 2 + \delta \) moment, \( \delta > 0 \), and is considerably simpler than the proofs of [6]. In [16] we use the convergence given by Theorem 2 or [6] to prove a law of the iterated logarithm for linear combinations of order statistics; in [15] Theorem 1 is used to prove a different type of strong limit theorem for \( U_n \).

For \( \delta > 0 \) let \( \mathcal{Q}_\delta \) denote the subset of \( \mathcal{Q} \) having \( \int_0^1 q^{-2 - \delta} \, dI < \infty \).

Theorem 2. Let \( q \in \mathcal{Q}_\delta \) for some \( \delta > 0 \). Then with probability one the sequence \( \{ U_n/b_n, n \geq 1 \} \) is relatively compact with respect to \( \rho_q \) with limit set \( \mathcal{B} \).

Proof. Suppose \( q \in \mathcal{Q}_\delta \). In view of Finkelstein’s [4] proof of the relative compactness with respect to the supremum metric \( \rho \) and symmetry of the process about \( t = \frac{1}{2} \), it suffices to show that with probability one

\[
\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \sup_{0 < t \leq \theta} \frac{|U_n(t)|}{q(t)b_n} = 0.
\]

Let \( \varepsilon > 0 \) and take \( \lambda = \varepsilon b_n/4 \) in (1). Application of the Cauchy–Schwarz inequality to (1) yields a bound involving \( \{T_n \geq \varepsilon b_n/4\} \). Since \( q \in \mathcal{Q}_\delta \), Remark 4 implies that a \( 2 + \delta \) version of the Berry–Esseen theorem [7] may be used to bound this probability.

Let \( \sigma_q^2 = \text{Var}(Y_1) = \int_0^1 q^{-2} \, dI \), \( C_\theta = E[(Y_1/\sigma_q)^{1+\delta/2}] \), and denote the standard normal density by \( \phi \). Using the Berry–Esseen bound, Mill’s ratio, and \( (a + b)^{1/2} \leq a^{1/2} + b^{1/2} \) one obtains, for \( n \geq 3 \),

\[
P\left( \sup_{0 < t \leq \theta} \frac{|U_n(t)|}{q(t)b_n} \geq \varepsilon \right) \leq \left( \frac{4}{\varepsilon b_n} \right) \sigma_q \left\{ \left( \frac{8\sigma_q}{\varepsilon b_n} \right) \phi \left( \frac{\varepsilon b_n}{4\sigma_q} \right) + C_\theta n^{-3/2} \right\}^{1/2}
\]

\[
\leq c_1 \exp \left( -\frac{1}{2} \left( \frac{\varepsilon}{4\sigma_q} \right)^2 \log \log n \right) + c_2 n^{-3/4}
\]

where \( c_1, c_2 \) are constants depending on \( \varepsilon \) and \( \theta \) but not on \( n \). By Remark 3, \( \theta \)
may be chosen so small that $\frac{1}{2} (\varepsilon / 4 \sigma_0^2)^3 > 1$; with this choice of $\theta$ the above inequality implies, via Borel-Cantelli, that with probability one the lim sup in (7) is less than $\varepsilon$ for a subsequence of the form $n_k = [\alpha^k]$ with $\alpha > 1$. This is easily extended to the full sequence in the usual way using (the Banach space version of) Skorohod’s inequality, and since $\varepsilon$ is arbitrary (8) holds.

Acknowledgment. This paper is based on part of the author’s 1975 Ph. D. dissertation prepared at the University of Washington under the direction of Galen R. Shorack.

REFERENCES