The Annals of Probability
1977, Vol. 5, No. 2, 303-308

A MARTINGALE INEQUALITY FOR THE EMPIRICAL PROCESS

BY JON A. WELLNER

University of Washington

A martingale inequality for the ρ_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 $n \geqq 1$ let $\xi_1, \cdots, \xi_n$ be i.i.d. uniform (0, 1) rv's and let Γ_n denote their empirical df. The uniform empirical process U_n is the process on [0, 1] defined by $U_n = n^{\frac{1}{2}}(\Gamma_n - I)$ 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 ρ_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 ρ denotes the usual supremum metric. The convergence of U_n with respect to certain of these ρ_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 ρ_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 Dvoretzky, 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 $U_n(t)/(1 - t)$, $0 \leqq t < 1$ 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 $\{X_j, j = 1, \cdots, m\}$ be arbitrary rv's and let $\{r_j, j = 1, \cdots, m\}$ be positive and nondecreasing real numbers; for $k = 1, \cdots, m$ set

$$S_k = \sum_{j=1}^k X_j\,, \qquad D_k = \sum_{j=1}^k (X_j/r_j)\,.$$

Received May 13, 1976.

AMS 1970 *subject classifications.* Primary 62E10, 60G45; Secondary 60F15.

Key words and phrases. Inequalities, empirical process, martingale, law of the iterated logarithm.

LEMMA 1. $\max_{1\le k\le m} |S_k|/r_k \leqq 2 \max_{1\le k\le m} |D_k|$.

PROOF. Let $\Delta r_j = r_j - r_{j-1}$, $\Delta D_j = D_j - D_{j-1}$, $j = 2, \cdots, m$, $\Delta r_1 = r_1$, $\Delta D_1 = D_1$. 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 $|S_k|/r_k \leqq \max_{1\le i\le k} |D_k - D_{i-1}|$ and this implies the statement of the lemma. □

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

To state the second lemma, let $\{T_k, \mathscr{E}_k, k = 1, \cdots, m\}$ be a positive submartingale.

LEMMA 2. *For all* $\lambda > 0$

$$P(\max_{1\le k\le m} T_k \geqq 2\lambda) \leqq \lambda^{-1} E(T_m 1_{[T_m \ge \lambda]}) .$$

PROOF. Let $M_m = \max_{1\le k\le m} T_k$. From Doob's martingale inequality,

$$\begin{aligned} 2\lambda P(M_m \geqq 2\lambda) &\leqq E(T_m 1_{[M_m \ge 2\lambda]}) \\ &= E(T_m 1_{[M_m \ge 2\lambda, T_m \ge \lambda]}) + E(T_m 1_{[M_m \ge 2\lambda, T_m < \lambda]}) \\ &\leqq E(T_m 1_{[T_m \ge \lambda]}) + \lambda P(M_m \geqq 2\lambda) . \end{aligned}$$ □

Let $\mathscr{Q}$ 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 $\int_0^1 q^{-2}\, dI < \infty$. The functions $q(t) = [t(1-t)]^{\frac{1}{2}-\delta}$ with $0 < \delta \leqq \frac{1}{2}$ are all in $\mathscr{Q}$; so are the functions $q(t) = [t(1-t)]^{\frac{1}{2}}[-\log[t(1-t)]]^{\frac{1}{2}+\delta}$ with $\delta > 0$.

THEOREM 1. *Let* $q \in \mathscr{Q}$ *and* $\theta \in (0, \frac{1}{2}]$. *Then for all* $\lambda > 0$

$$P\left(\sup_{0<t\le\theta} \frac{|U_n(t)|}{q(t)} \geqq 4\lambda\right) \leqq \lambda^{-1} E(|T_n| 1_{[|T_n| \ge \lambda]}) \tag{1}$$

where $T_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} Y_i$, *the sum of the* i.i.d. rv's

$$Y_i = \frac{1}{q_\theta(\xi_i)} - \int_0^{\xi_i} \frac{1}{(1-I)q_\theta}\, dI$$

$i = 1, \cdots, n$ *with* $1/q_\theta = q^{-1} 1_{(0,\theta]}$. *Furthermore, the* Y_i*'s have* $E(Y_i) = 0$ *and* $\operatorname{Var}(Y_i) = \int_0^\theta q^{-2}\, dI$.

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 $s/(1-s)$, $s \leqq t$. Also let $r(t) = q(t)/(1-t)$. For $m = 2^h$, $h \geqq 1$ an integer, and $1 \leqq k \leqq m$, 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 $1 \leqq k \leqq [m\theta]$. Then, using Lemmas 1 and 2

$$\begin{aligned} P\left(\sup_{0<t\le\theta} \frac{|U_n(t)|}{q(t)} > 4\lambda\right) &= \lim_{h\to\infty} P\left(\max_{1\le k\le[m\theta]} \frac{|W_n(k/m)|}{r_k} > 4\lambda\right) \\ &= \lim_{h\to\infty} P\left(\max_{1\le k\le[m\theta]} \frac{|\sum_1^k X_j|}{r_k} > 4\lambda\right) \\ &\leqq \lim_{h\to\infty} P(\max_{1\le k\le[m\theta]} |\textstyle\sum_1^k (X_j/r_j)| > 2\lambda) \\ &\leqq \lim_{h\to\infty} \lambda^{-1} E(|\textstyle\sum_1^{[m\theta]} (X_j/r_j)| 1_{[|\sum_1^{[m\theta]}(X_j/r_j)| > \lambda]}) \end{aligned} \tag{2}$$

where the first inequality follows from Lemma 1 and the second inequality follows from Lemma 2 since, by Remark 1, $\{\sum_{j=1}^{k}(X_j/r_j), k = 1, \cdots, [m\theta]\}$ is a martingale. We now show that

$$T_n \equiv \lim_{h\to\infty} \textstyle\sum_{j=1}^{[m\theta]} (X_j/r_j) \tag{3}$$

exists for each $\omega \in \Omega$ and equals T_n of the statement of the theorem. Write $W_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} Q_i$ with $Q_i(t) = (1_{(0,t]}(\xi_i) - t)/(1-t)$. Using this together with the definition of X_j in (3) and interchanging the order of summation one obtains

$$T_n = n^{-\frac{1}{2}} \textstyle\sum_{i=1}^{n} \lim_{h\to\infty} \sum_{j=1}^{[m\theta]} \left\{Q_i\left(\frac{j}{m}\right) - Q_i\left(\frac{j-1}{m}\right)\right\}\Big/ r_j .$$

Since the Q_i are i.i.d. processes, it suffices to show that this last limit exists for $i = 1$ and equals Y_1 of the statement of the theorem. For $s < t$

$$Q_1(t) - Q_1(s) = \frac{1}{(1-t)} 1_{(s,t]}(\xi_1) - \frac{(t-s)}{(1-s)(1-t)} 1_{(s,1]}(\xi_1)$$

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

$$\begin{aligned}
&\textstyle\sum_{j=1}^{[m\theta]} \left\{Q_1\left(\frac{j}{m}\right) - Q_1\left(\frac{j-1}{m}\right)\right\}\Big/ r_j \\
&\quad = \textstyle\sum_{j=1}^{[m\theta]} \dfrac{1_{((j-1)/m,\, j/m]}(\xi_1)}{(1-(j/m))r_j} - \dfrac{1}{m} \textstyle\sum_{j=1}^{[m\theta]} \dfrac{1_{((j-1)/m,1]}(\xi_1)}{(1-(j-1)/m)(1-j/m)r_j} \\
&\quad \to \dfrac{1}{q_\theta(\xi_1)} - \textstyle\int_0^{\xi_1} \dfrac{1}{(1-I)q_\theta}\, dI \qquad h \to \infty \\
&\quad = Y_1 .
\end{aligned}$$

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 $\{\sum_{j=1}^{[m\theta]} (X_j/r_j), m \geqq 1\}$ is bounded in L_2 and hence uniformly integrable.

That $E(Y_1) = 0$ and $\text{Var}\,(Y_1) = \int_0^\theta q^{-2}\, dI$ is easily verified by straightforward computation. □

REMARK 2. The process $\{B_n(t) \equiv (1+t)U_n(t/(1+t)), 0 \leqq t < \infty\}$ 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

$$T_n = \textstyle\int_0^{\theta^*} f\, dB_n$$

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

$$E(T_n^2) = \textstyle\int_0^{\theta^*} f^2\, dI = \int_0^\theta q^{-2}\, dI .$$

REMARK 3. For $q \in \mathscr{Q}$, $\int_0^\theta q^{-2}\,dI \to 0$ as $\theta \to 0$ and hence $\operatorname{Var}(Y_1)$ can be made arbitrarily small by choosing θ small.

REMARK 4. If $\int_0^1 q^{-2-\delta}\,dI < \infty$ for some $\delta > 0$, then the C_r and Jensen inequalities may be used to show that $E|Y_1|^{2+\delta} \leqq C(\delta) \int_0^\theta q^{-2-\delta}\,dI < \infty$ 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])

$$P\left(\sup_{0<t\leqq\theta} \frac{|U_n(t)|}{q(t)} \geqq \lambda\right) \leqq \lambda^{-2} \int_0^\theta q^{-2}\,dI . \tag{4}$$

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

$$P(\sup_{0\leqq t\leqq 1} |U_n(t)| \geqq \lambda) \leqq Ce^{-2\lambda^2} \tag{5}$$

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 $q \in \mathscr{Q}$ and $\lambda > 0$*

$$P\left(\sup_{0<t\leqq\theta} \frac{|U_n(t)|}{q(t)} \geqq \lambda\right) \leqq 16\lambda^{-2} \int_0^\theta q^{-2}\,dI . \tag{6}$$

PROOF. This follows immediately from (1) and $E(T_n^2) = \int_0^\theta q^{-2}\,dI$. □

COROLLARY 2. *For all $\lambda > 0$*

$$P(\sup_{0\leqq t\leqq 1} |U_n(t)| \geqq \lambda) \leqq 8(2\pi)^{-\frac{1}{2}}\lambda^{-1}e^{-\lambda^2/50} . \tag{7}$$

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

$$Y_i = 1 - \int_0^{\xi_i} (1 - I)^{-1}\,dI = 1 + \log(1 - \xi_i) = -(\exp(1) - 1)$$

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:

$$E(|T_n| 1_{[|T_n|\geqq\lambda]}) = \frac{n^{n+\frac{1}{2}}e^{-n}}{n!} \{(1 - \lambda_n)^n e^{n\lambda_n} + (1 + \lambda_n)^n e^{-n\lambda_n}\}$$

where $\lambda_n \equiv \lambda n^{-\frac{1}{2}}$. Use of Stirling's formula and the elementary inequalities $\log(1 - x) \leqq -x - \frac{1}{2}x^2$ and $\log(1 + x) \leqq x - \frac{8}{25}x^2$, $0 \leqq x \leqq \frac{1}{4}$ (recall that $\sup_{0\leqq t\leqq 1} |U_n(t)| \leqq n^{\frac{1}{2}}$ and hence we need only consider $4\lambda \leqq n^{\frac{1}{2}}$ or $\lambda_n \leqq \frac{1}{4}$) to bound this last expression yields

$$P(\sup_{0\leqq t\leqq 1} |U_n(t)| \geqq 4\lambda) \leqq (2/\pi)^{\frac{1}{2}}\lambda^{-1}e^{-(8/25)\lambda^2}$$

which implies (7). □

The inequalities (6) and (7) are not as sharp as the inequalities (4) and (5) essentially because of the two factors of two which enter through Lemmas 1 and

2. However, (1) holds for all $q \in \mathscr{Q}$ 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 $b_n = (2 \log \log n)^{\frac{1}{2}}$ and let

$$\mathbb{B} = \{f \in C[0, 1] : f(0) = 0 = f(1),\ f = \int_0^\cdot f'\, dI,\ \int_0^1 (f')^2\, dI \leqq 1\}\,.$$

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

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

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 ρ_q for a class of functions q which is slightly smaller than $\mathscr{Q}$. 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 $\mathscr{Q}_\delta$ denote the subset of $\mathscr{Q}$ having $\int_0^1 q^{-2-\delta}\, dI < \infty$.

THEOREM 2. *Let $q \in \mathscr{Q}_\delta$ for some $\delta > 0$. Then with probability one the sequence $\{U_n/b_n,\ n \geqq 1\}$ is relatively compact with respect to ρ_q with limit set $\mathbb{B}$.*

PROOF. Suppose $q \in \mathscr{Q}_\delta$. In view of Finkelstein's [4] proof of the relative compactness with respect to the supremum metric ρ and symmetry of the process about $t = \frac{1}{2}$, it suffices to show that with probability one

$$\lim_{\theta \to 0} \limsup_{n \to \infty} \sup_{0 < t \leqq \theta} \frac{|U_n(t)|}{q(t) b_n} = 0\,. \tag{8}$$

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

Let $\sigma_\theta^2 = \mathrm{Var}(Y_1) = \int_0^\theta q^{-2}\, dI$, $C_\theta = E|(Y/\sigma_\theta)|^{2+\delta}$, and denote the standard normal density by ϕ. Using the Berry–Esseen bound, Mill's ratio, and $(a + b)^{\frac{1}{2}} \leqq a^{\frac{1}{2}} + b^{\frac{1}{2}}$ one obtains, for $n \geqq 3$,

$$\begin{aligned} P\left(\sup_{0<t\leqq\theta} \frac{|U_n(t)|}{q(t) b_n} \geqq \varepsilon\right) &\leqq \left(\frac{4}{\varepsilon b_n}\right) \sigma_\theta \left\{\left(\frac{8\sigma_\theta}{\varepsilon b_n}\right) \phi\left(\frac{\varepsilon b_n}{4\sigma_\theta}\right) + C \cdot C_\theta n^{-\delta/2}\right\}^{\frac{1}{2}} \\ &\leqq c_1 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon}{4\sigma_\theta}\right)^2 \log \log n\right) + c_2 n^{-\delta/4} \end{aligned}$$

where c_1, c_2 are constants depending on ε and θ but not on n. By Remark 3, θ

may be chosen so small that $\frac{1}{2}(\varepsilon/4\sigma_\theta)^2 > 1$; with this choice of θ the above inequality implies, via Borel–Cantelli, that with probability one the lim sup in (7) is less than ε 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 ε 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

[1] BURKHOLDER, D. L. (1966). Martingale transforms. *Ann. Math. Statist.* **37** 1494–1504.

[2] CHOVER, J. (1966). On Strassen's version of the log log law. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **8** 83–90.

[3] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[4] FINKELSTEIN, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615.

[5] MCKEAN, H. P. (1969). *Stochastic Integrals.* Academic Press, New York.

[6] JAMES, B. R. (1975). A functional law of the iterated logarithm for weighted empirical distributions. *Ann. Probability* **3** 762–772.

[7] KATZ, M. L. (1963). Note on the Berry-Esseen theorem. *Ann. Math. Statist.* **34** 1107–1108.

[8] KIEFER, J. (1972). Skorohod embedding of multivariate rv's and the sample df. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **24** 1–35.

[9] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

[10] PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

[11] PYKE, R. and SHORACK, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

[12] SHORACK, G. R. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

[13] SHORACK, G. R. and SMYTHE, R. T. (1976). Inequalities for max $|S_k|/b_k$ where $k \in N^r$. *Proc. Amer. Math. Soc.* **54** 331–336.

[14] STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 211–226.

[15] WELLNER, J. A. (1976). Sequential quantile and spacings processes. Submited to *Ann. Statist.*

[16] WELLNER, J. A. (1976). A law of the iterated logarithm for order statistics. To appear in *Ann. Statist.*

DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NEW YORK 14627