A LAW OF ITERATED LOGARITHM FOR ONE SAMPLE RANK-ORDER STATISTICS AND AN APPLICATION

by

Pranab Kumar Sen and Malay Ghosh

Department of Biostatistics
University of North Carolina, Chapel Hill

Institute of Statistics Mimeo Series No. 687

June, 1970
ABSTRACT

A LAW OF ITERATED LOGARITHM FOR ONE SAMPLE RANK-ORDER STATISTICS
AND AN APPLICATION by Pranab Kumar Sen and Malay Ghosh University
of North Carolina at Chapel Hill

A law of iterated logarithm has been established for one sample
rank-order statistics, and a test procedure for the classical one
sample location problem has been proposed which is of power 1 and
arbitrarily small type I error.
A LAW OF ITERATED LOGARITHM FOR ONE SAMPLE RANK-ORDER STATISTICS AND AN APPLICATION*

BY PRANAB KUMAR SEN AND MALAY GHOSH

University of North Carolina at Chapel Hill

1. Introduction and summary. The law of iterated logarithm for sample sums of iidrv (independent and identically distributed random variables) was first proved by Khintchine [7] for Bernoulli variables, and, later on, more general results in this direction were due to Kolmogorov [8], Petrovski [9], Erdös [3], Feller [4] and others. An excellent exposition of these results is available in Feller [4], and Strassen [11], while the latter extends these results to martingales. More recently, these results have been generalized to sample quantiles by Bahadur [1] and to U-statistics by Ghosh and Sen [5]. In the present paper, we find a similar law for one-sample rank-order statistics, and the details are provided in section 2. In section 3, for the classical one-sample location problem, a test procedure has been proposed along the lines of Darling and Robbins [2] with zero type II error, and arbitrarily small type I error. To achieve this, application is made of results of section 2, and also of some strong convergence results in connection with one-sample rank-order statistics as given in Sen [10].

2. A law of iterated logarithm for one-sample rank-order statistics.

*Work supported by National Institutes of Health, Grant GM-12868.
Let \( \{X_1, X_2, \ldots \} \) be a sequence of iidrv defined on the measure spaces \((\Omega, A, P_\theta)\), having a df(distribution function) \( F_\theta(x) = F(x-\theta) \), where \( \theta \in \Theta \) is an unknown parameter, and \( F \in \mathcal{I}_0 \), the class of all df's continuous with respect to Lebesgue measure, and symmetric about zero, i.e. \( F(x) + F(-x) = 1 \), for all real \( x \). Define for each positive integer \( n \),

\[
(2.1) \quad X_n = (X_1, \ldots, X_n), \quad 1_n = (1, \ldots, 1), \quad R_{ni} = \frac{1}{2} + \frac{1}{n} \sum_{j=1}^{n} c(|X_i| - |X_j|), \quad i=1,2,\ldots,n, \quad \text{where } c(u) = 1, 1/2 \text{ or 0 according as } u >, =, \text{ or } < 0.
\]

Let \( J_n(i/(n+1)) = EJ(U_{ni}) \), \( i=1,2,\ldots,n \), and,

\[
(2.2) \quad T_n = T_n(X_n) = n^{-1} \sum_{i=1}^{n} \text{sgn } X_i EJ(U_{ni}), \quad n \geq 1,
\]

where \( \text{sgn } x = 2c(x)-1 \) (\( x \) real), \( U_{n1} < \ldots < U_{nn} \) are \( n \) ordered rv's (random variables) from a rectangular \((0,1)\) df, and \( J(u) \), \( 0 < u < 1 \), is a score function satisfying \( J(u) = \psi^{-1}(1-\frac{1}{2}u) \), \( \psi \) being a symmetric df defined on \(( -\infty, \infty) \), i.e.,

\[
(2.3) \quad \psi(x) + \psi(-x) = 1, \quad \text{for all real } x.
\]

Define \( \mu = \int_0^1 J(u)du \), and, \( A^2 = \int_0^1 J^2(u)du \). Whenever, \( \psi(x) \) is non-degenerate, \( A^2 > 0 \). We assume that \( 0 < A^2 < \infty \). Note that if \( \psi(x) \) is uniform over \((-1,1)\) or is the standard normal df, the corresponding \( T_n \) is termed the signed-rank or normal-scores statistic. The following theorem is proved.

**THEOREM 2.1.** If \( J \in L_{2+\delta} \) for some \( \delta > 0 \), then under \( H_0: F \in \mathcal{I}_0 \),
\[ (2.4) \lim \sup_{n \to \infty} \sqrt{n} T_n / \left[ A(2 \log \log n)^{1/2} \right] = 1 \ a.s. ; \]

\[ (2.5) \lim \inf_{n \to \infty} \sqrt{n} T_n / \left[ A(2 \log \log n)^{1/2} \right] = -1 \ a.s. \]

**Proof.** Let \( B_n \) denote the \( \sigma \)-field generated by \( S_n = (\text{sgn} \ X_1, \ldots, \text{sgn} \ X_n) \) and \( R_n = (R_{n1}, \ldots, R_{nn}) \), \( n \geq 1 \); clearly \( B_n \uparrow \text{in} \ n \). Note that for \( F \in \mathcal{F}_0 \), \( \theta = 0 \), the two vectors of signs and ranks of absolute \( X \)'s are stochastically independent (see [6], p. 40). Write \( \tilde{T}_n = nT_n \).

Then, using the notation \( E_0 \) for \( E_{\mathcal{F}_0} \) (to be continued later on),

\[ (2.6) \quad E_0(\tilde{T}_n) = 0, \quad E_0(\tilde{T}_n^2) = nA_n^2, \]

where \( A_n^2 = n^{-1} \sum_{i=1}^n E[J(U_{ni})]^2 \leq A^2 < \infty \). Also,

\[ E_0(\tilde{T}_{n+1}'|B_n) = \sum_{i=1}^n \text{sgn} \ X_i E_0[J(U_{n+1R_{n+1i}}')|B_n] \]

\[ + E_0[\text{sgn} \ X_{n+1}]E_0[J(U_{n+1R_{n+1}})|B_n] \]

\[ = \sum_{i=1}^n \text{sgn} \ X_i [(1 - \frac{R_{ni}}{n+1})EJ(U_{n+1R_{ni}}) + \frac{R_{ni}}{n+1} EJ(U_{n+1R_{ni}+1})] \]

\[ = \sum_{i=1}^n \text{sgn} \ X_i EJ(U_{nR_{ni}}) = \tilde{T}_n , \]

where one uses the relation

\[ (2.8) \quad J_{n-1}(i/n) = (i/n)J_n ((i+1)/(n+1)) + ((n-i)/n)J_n (i/(n+1)) \]

Thus, \( (\tilde{T}_n, B_n, n \geq 1) \) form a martingale sequence.

To prove the theorem, we need now verify only the conditions of the theorem 4.4 of Strassen ([11], p. 334) which gives us access
via the extension of the Kolmogorov-Petrovski-Erdős criterion to martingales ([11], corollary 4.5, p. 337) to the law of iterated logarithm as given in (2.4) and (2.5).

First define, \( Z_1 = \frac{\tilde{T}_1}{T_1}, Z_k = \frac{\tilde{T}_k}{T_{k-1}} \) \((k \geq 2)\).

Then,

(i) \( E_0(Z_1^2) = 0, E_0(Z_k|B_{k-1}) = 0 \) \((k \geq 2)\);  

(ii) \( 0 < E_0(Z_1^2) = \mu^2 < \infty \);  

(iii) \( E_0(Z_n^2|B_{n-1}) = E\{\sum_{i=1}^{n-1} \text{sgn} \ X_i \{J_n(\frac{R_{in}}{n+1}) - J_{n-1}(\frac{R_{in}}{n})\} + \text{sgn} \ X_n \ J_n(\frac{R_{nn}}{n+1})^2 \} \) \(E_0\{\sum_{i=1}^{n-1} \text{sgn} \ X_i \{J_n(\frac{R_{in}}{n+1}) - J_{n-1}(\frac{R_{in}}{n})\} - J_{n-1}(\frac{R_{n-1i}}{n})^2 \} \) \(E_0\{J_n^2(\frac{R_{nn}}{n+1})|B_{n-1}\} \geq A_n^2\),

(using the stochastic independence of sign and rank vectors under \( H_0 \) and using the fact that \( E_0(\text{sgn} \ X_n|B_{n-1}) = E_0(\text{sgn} \ X_n) = 0 \).  

Thus, \( V_n = \sum_{i=2}^{n} E_0(Z_i^2|B_{i-1}) + E_0(Z_1^2) \geq \frac{n^2}{2} A_i^2 + \mu^2 \). Since, \( \sum \frac{A_i^2}{nA^2+1} \) as \( n \to \infty \), we have, \( V_n \to \infty \) as \( n \to \infty \). Define now

(2.9) \( f(t) = \frac{t(\log \log t)^2}{(\log t)^4}, t \geq 3; f(t) = 1, 0 < t \leq 3. \)

Then,

(iv) \( f(t) \uparrow, f(t)/t \uparrow \) in \( t \).

Again, since

(2.10) \( J_n(i/(n+1)) \leq J_{n-1}(i/n) \leq J_n((i+1)/(n+1)), 1 \leq i \leq n, \)
Then, we have,

\[ |Z_n| \leq \sum_{i=1}^{n-1} |J_n \left( \frac{R_i}{n+1} \right) - J_{n-1} \left( \frac{R_{i-1}}{n} \right)| + J_n \left( \frac{R_n}{n+1} \right) \]

\[ \leq \sum_{i=1}^{n-1} \left\{ \max |J_n \left( \frac{i}{n+1} \right) - J_{n-1} \left( \frac{i}{n} \right)|, |J_n \left( \frac{i+1}{n+1} \right) - J_{n-1} \left( \frac{i}{n} \right)| \right\} + J_n \left( \frac{n}{n+1} \right) \]

\[ \leq \sum_{i=1}^{n-1} \left[ J_n \left( \frac{i+1}{n+1} \right) - J_n \left( \frac{i}{n+1} \right) \right] + J_n \left( \frac{n}{n+1} \right) = 2J_n \left( \frac{n}{n+1} \right). \]

But,

\[ J_n \left( \frac{n}{n+1} \right) = EJ(U_{nn}) \leq [EJ^{2+\delta}(U_{nn})]^{1/(2+\delta)} \]

\[ = [n \int_0^1 J^{2+\delta}(u) \, du]^{1/(2+\delta)} = o(n^{2+\delta}). \]

Further, since,

\[ |\sum_{i=1}^{n-1} \left\{ J_n \left( \frac{R_i}{n+1} \right) - J_{n-1} \left( \frac{R_{i-1}}{n} \right) \right\} \, \text{sgn} \, X_i| \]

\[ \leq \sum_{i=1}^{n-1} \max\left\{ |J_n \left( \frac{R_{i-1}}{n+1} \right) - J_{n-1} \left( \frac{R_{i-1}}{n} \right)|, |J_n \left( \frac{R_{i-1}+1}{n+1} \right) - J_n \left( \frac{R_{i-1}}{n+1} \right)| \right\} \]

\[ \leq \sum_{i=1}^{n-1} \left[ J_n \left( \frac{R_{i-1}+1}{n+1} \right) - J_n \left( \frac{R_{i-1}}{n+1} \right) \right] \]

\[ = \sum_{i=1}^{n-1} \left[ J_n \left( \frac{i+1}{n+1} \right) - J_n \left( \frac{i}{n+1} \right) \leq J_n \left( \frac{n}{n+1} \right) \right], \]

\[ E \left( Z_n^2 | B_{n-1} \right) \leq J_n^2 \left( \frac{n}{n+1} \right) + A_n^2 \leq J_n^2 \left( \frac{n}{n+1} \right) + A^2, \quad n \geq 2. \]
So,

\[ V_n = E(z_1^2) + \sum_{i=1}^{n} E(Z_i^2 | B_{i-1}) \]

\[ < A^2 + \sum_{i=1}^{n} J_1^2 (\frac{i}{i+1}) + (n-1)A^2 = \sum_{i=1}^{n} J_1^2 (\frac{i}{i+1}) + nA^2. \]

Thus,

\[ f(V_n) = \frac{\log \log V_n}{V_n}^2 / (\log V_n)^4 \]

\[ \geq \frac{1}{2} A_1^2 + \mu^2 \right\} \frac{\log \log \left( \sum_{i=1}^{n} A_1^2 + \mu^2 \right)}{(\log \left( \sum_{i=1}^{n} J_1^2 (\frac{i}{i+1}) + nA^2 \right))^4} \]

Hence,

\[ |Z_n| / [f(V_n)]^{1/2} \]

\[ \leq 2J_n \left( \frac{n}{n+1} \right) \frac{\log \left( \sum_{i=1}^{n} J_1^2 (\frac{i}{i+1}) + nA^2 \right)}{(\sum_{i=1}^{n} A_1^2 + \mu^2)^{1/2}} \frac{1}{(\log \log \left( \sum_{i=1}^{n} A_1^2 + \mu^2 \right))^4} \]

Now,

\[ \sum_{i=1}^{n} J_1^2 (\frac{i}{i+1}) \leq \sum_{i=1}^{n} 0 \left( \frac{2}{2+\delta} \right) = 0 \left( \frac{4+\delta}{2+\delta} \right), \]

\[ J_n \left( \frac{n}{n+1} \right) = 0 \left( \frac{4+\delta}{2+\delta} \right), \frac{\sum_{i=1}^{n} A_1^2 + \mu^2}{nA^2} \to 1 \text{ as } n \to \infty. \]

Hence, \[ |Z_n| / [f(V_n)]^{1/2} = o(1). \] Thus, there exists an \( n_0 \) such that

\[ |Z_n| / [f(V_n)]^{1/2} < 1 \text{ for } n > n_0. \]
Hence,

\[
(\nu) \sum_{n \geq 1} [f(V_n)]^{-1} \int_{x > f(V_n)} x^2 \, dP\{Z_n < x | B_{n-1}\} = \sum_{n=1}^{n_0} [f(V_n)]^{-1} \int_{x > f(V_n)} x^2 \, dP\{Z_n < x | B_{n-1}\} < \infty.
\]

The proof of the theorem now follows directly from (i) - (\nu) and theorem 4.4 and corollary 4.5 of Strassen [11].

REMARK. If we define a process \( \tilde{T}(t) \), \( t \geq 0 \) by \( \tilde{T}(0) = 0 \), and

\[
(2.11) \quad \tilde{T}(t) = (t-n) \tilde{T}_{n+1} + [1-(t-n)] \tilde{T}_n \text{ for } n \leq t \leq n+1, \ n \geq 0,
\]

and consider a Brownian motion \( \xi(t) \) for which \( E\xi(t) = 0 \), \( E[\xi(s) \xi(t)] = A^2 s, 0 \leq s \leq t \leq \infty \), we have from (i) - (\nu) and theorem 4.4 of Strassen that

\[
(2.12) \quad \tilde{T}(t) = \xi(t) + o([t \log \log t]^{1/2}) \text{ a.s. as } t \to \infty.
\]

\textbf{Test of hypothesis with power 1}. We start with the same set up as in section 2, and assume \( F \in \mathcal{F}_0 \), \( J(u) \) strictly increasing in \( u \) \( (0 \leq u < 1) \). Consider,

\[
(3.1) \quad H_{01}: \ \theta = 0 \text{ against the alternatives } \theta > 0;
\]

\[
(3.2) \quad H_{02}: \ \theta = 0 \text{ against the alternatives } \theta \neq 0.
\]

Sen [10] has proved that if \( J \in L_1 \), then, \( \lim_{n \to \infty} T_n = \eta_0 \) a.s. (\( P_\theta \)), where, in our notations,
\[(3.3) \quad \eta_\theta = 2 \int_0^\infty J[F(x-\theta) - F(-x-\theta)]dF(x-\theta) - \mu.\]

When,

\[\theta > 0, \quad \eta_\theta \geq 2 \int_0^\infty J[F(x-\theta) - F(-x-\theta)]dF(x-\theta) - \mu\]

\[= 2 \int_0^\infty J[F(y) - F(-y-2\theta)]dF(y) - \mu > 2 \int_0^\infty J[F(y) - F(-y)]dF(y) - \mu\]

\[= 0, \text{ since } J(u) + u \text{ (strict) } F(x) + u \text{ (strict) } 0 \leq u < 1, x \text{ real}.\]

Similarly, for \(\theta < 0, \eta_\theta < 0.\)

Now, to test \(H_{01}\) define

\[(3.4) \quad N = \begin{cases} \text{first integer } n \geq n_0 \text{ such that } T_n \geq c_n/n, \\ \infty \text{ if no such } n \text{ occurs,} \end{cases}\]

where \(c_n\) is some sequence positive of constants such that \(c_n/n \to 0\) as \(n \to \infty.\) If \(H_{01}\) is false then \(T_n \to \eta_\theta(>0)\) a. s. as \(n \to \infty,\) and hence,

\[(3.5) \quad P_\theta(N=\infty) = \lim_{n \to \infty} P_\theta(N>n) \leq \lim_{n \to \infty} P(T_n < c_n/n) = 0.\]

Hence, if we agree to reject \(H_{01}\) as soon as we observe that \(N < \infty,\) while if \(N=\infty,\) we do not reject \(H_{01},\) then since \(P_\theta(N=\infty) = 1\) for \(\theta > 0,\) the test has power 1. Again, when \(H_{01}\) is true, by the law of iterated logarithm in section 2, \(\limsup_{n \to \infty} \sqrt{n} T_n / [A(2 \log \log n)^{1/2}] = 1.\)

Then,

\[(3.6) \quad P_\theta(N < \infty) = P_\theta(T_n \geq c_n/n \text{ for some } n \geq n_0) = P_\theta\left(\frac{\sqrt{n} T_n}{A(2 \log \log n)^{1/2}} \geq \frac{c_n}{\sqrt{2A(n \log \log n)^{1/2}}} \text{ for some } n \geq n_0\right)\]
which can be made arbitrarily small by taking $c_n = \sqrt{2}A(1+\varepsilon)$ 
$(n \log \log n)^{1/2}$, $\varepsilon > 0$, and $n_0$ sufficiently large.

REMARK. The above result does not provide any explicit upper bound 
for $P_0(N<\infty)$. However, if instead we take $c_n = \sqrt{2}A(1+\varepsilon)(n \log n)^{1/2}$, 
we can achieve this as follows:

$$(3.7) \quad P_0(T_n \geq c_n/n) = P_0(\tilde{T}_n \geq c_n) \leq \inf_{t > 0} E[\exp(t(\tilde{T}_n - c_n))]$$

But,

$$E[\exp(t(\tilde{T}_n - c_n))]$$

$$= e^{-tc_n} E[\exp(t \sum_{i=1}^n \text{sgn } X_i E(J(U_{nR_{ni}})))]$$

$$= e^{-tc_n} E[E[\exp(t \sum_{i=1}^n \text{sgn } X_i E(J(U_{nR_{ni}})))]|R_n]$$

Using the fact that $\text{sgn } X_n$ and $R_n$ are stochastically independent 
under $H_{01}$ and using the elementary inequalities $(e^x + e^{-x})/2 \leq \exp(x^2/2)$ 
for $x$ real, and $\sum_{i=1}^n [EJ(U_{ni})]^2 = nA_n^2 \leq nA^2$, one gets,

$$E[\exp(t \sum_{i=1}^n \text{sgn } X_i EJ(U_{nR_{ni}}))]|R_n]$$

$$= \prod_{i=1}^n [1/2\{\exp(t EJ(U_{nR_{ni}}))\} + \{\exp(-t EJ(U_{nR_{ni}}))\}]]$$

$$(3.8) \quad \leq \prod_{i=1}^n \exp\{(t^2/2)(EJ(U_{nR_{ni}}))^2\} = \exp\{(t^2/2) \sum_{i=1}^n (EJ(U_{ni}))^2\}$$

$$\leq \exp(nA^2 t^2/2).$$
We may remark that unlike the case of sample sum, we do not need the assumption the \( J(u) \) is exponentially integrable; \( J \in L_2 \) suffices the purpose.

Thus,

\[
P_0(T_n \geq c_n/n) \leq \inf_{t>0} \exp(-t c_n + \frac{nA^2 t^2}{2})
= \exp(-c_n^2/2nA^2) = n^{-(1+\varepsilon)}.
\]

Then,

\[
P_0(N < \infty) \leq \sum_{n=n_0}^{\infty} P_0(T_n \geq c_n) \leq \sum_{n=n_0}^{\infty} n^{-(1+\varepsilon)} < \infty.
\]

One may remark that a similar problem was faced by Darling and Robbins [2] in connection with the derivation of Kolmogorov-Smirnov tests with power 1, where to obtain an explicit upper bound for \( P_0(T_n \geq c_n/n) \), \( c_n \) was taken to be \( \{(1+\varepsilon) n \log n\}^{1/2} \) instead of \( \{(1+\varepsilon) n \log \log n\}^{1/2} \).

Now, for the testing problem \( H_{02} \), define

\[
N = \begin{cases} 
\text{first integer } n > n_0 \text{ such that } |T_n| \geq c_n/n \\
\infty \text{ if no such } n \text{ occurs},
\end{cases}
\]

(3.9) \( N = \infty \) if no such \( n \) occurs.

\( c_n \) defined in the same way as earlier. If \( H_{02} \) is false, \( T_n \to \eta_0 \) a.s. as \( n \to \infty \), where \( \eta_0 > (\varepsilon) 0 \) as \( \varepsilon > (\varepsilon) 0 \). Hence \( |T_n| \to |\eta_0| \) a.s. as \( n \to \infty \).

Then,

\[
P_\theta(N = \infty) = \lim_{n \to \infty} P_\theta(N > n) \leq \lim_{n \to \infty} P_\theta(|T_n| < \frac{c_n}{n}) = 0.
\]

Noting that \( T_n \) is distributed symmetrically about 0 under \( H_{02} \) and using similar arguments as before, one reaches the conclusion that the test is of power 1 and arbitrarily small type I error.
REFERENCES


