BOL. SOC. BRAS. MAT. 8.2(1977), 109–115

A note on the law of the iterated logarithm for the empirical distribution function

Barry R. James

1. Introduction and principal result.

Let $X_1, X_2, \ldots$ be i.i.d. random variables, each uniformly distributed in the interval $[0, 1]$, and let F_n be the empirical d.f. at stage n, i.e. for $n \geq 1$ and $0 \leq t \leq 1$,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i \leq t]},$$

where $I_{[X_i \leq t]}$ is the indicator function of the event $[X_i \leq t]$. The purpose of this paper is to obtain uniform upper bounds for $t/F_n(t)$, where we consider only those values of t above $\min(X_1, \ldots, X_n) \equiv Z_{n1}$. In other words, we will study the behavior of the sequence

$$M_n \equiv \sup_{Z_{n1} \leq t \leq 1} \frac{t}{F_n(t)}.$$

As a first step in this direction, Kiefer proved (1972, Theorem 2) that if $Z_{n1} \leq Z_{n2} \leq \ldots \leq Z_{nn}$ are the order statistics of $X_1, \ldots, X_n$ and k is a fixed positive integer, then

$$\limsup_{n\to\infty} \frac{n Z_{nk}}{\log \log n} = 1 \quad a.s.$$

Using the inequalities

$$\frac{n Z_{n1}}{\log \log n} \leq \sup_{[Z_{n1},\, Z_{nk}]} \frac{t}{F_n(t) \log \log n} \leq$$

$$\leq \frac{Z_{nk}}{F_n(Z_{n1}) \log \log n} = \frac{n Z_{nk}}{\log \log n},$$

together with Kiefer's result, we see that

$$\limsup_{n\to\infty} \sup_{[Z_{n1},\, Z_{nk}]} \frac{t}{F_n(t) \log \log n} = 1 \quad a.s. \qquad k \geq 1.$$

This result was extended in [3] to the whole interval under consideration; lemma 4 of [3] gives a law of the iterated logarithm (LIL) for M_n:

$$\limsup_{n\to\infty} \frac{M_n}{\log\log n} = 1 \quad a.s.$$

We note here that with regard to this LIL, the behavior of M_n is determined by the behavior of the minimum, in the sense that the extreme behavior of M_n is attained along the sequence $(Z_{n1})_{n\geq 1}$.

Frankel (1976) published the following result extending Kiefer's LIL to an upper/lower class theorem: if $(c_n)_{n\geq 1}$ is a sequence of positive real numbers such that

(a) the sequence is eventually (ultimately) increasing and converges to $+\infty$, and

(b) $(c_n/n)_{n\geq 1}$ is eventually decreasing to zero, then $P(Z_{nk} \geq c_n/n$ infinitely often) $= 0$ or 1 according as

$$\sum_{n=1}^{\infty} e^{-c} \frac{c_n^k}{n} < \infty \quad \text{or} \quad = \infty.$$

Remark. Frankel's expression (5) contains a very unfortunate typographical error (the exponent k has been omitted), and the statement above is the corrected version.

In trying to obtain an analogous upper/lower class result for M_n, we note that the behavior of the kth order statistic now depends on k, which changes the situation slightly. It turns out that M_n no longer follows the same behavior pattern as minimum, but rather that its extreme behavior is determined by the (values immediately below the) second order statistic. Our mains result is that the sequence $(c_n)_{n\geq 1}$ belongs to the upper class (i.e. $M_n \leq c_n$ eventually a.s.) if and only if $\Sigma e^{-c_n} c_n^2/n < \infty$:

Theorem. *If $(c_n)_{n\geq 1}$ satisfies the monotonicity conditions (a) and (b) given above, then $P(M_n \geq c_n$ infinitely often$) = 0$ or 1 according as*

$$\sum_{n=1}^{\infty} e^{-c_n} \cdot \frac{c_n^2}{n} < \infty \quad or \quad = \infty$$

2. Proof of the Theorem.

(Note: the proof follows the general lines of the proof of lemma 4 in [3]. Since that result has not appeared in print, all details will be given). By Frankel, $nZ_{n2} \geq c_n$ i.o. with probability 0 or 1 according as $\Sigma e^{-c_n} c_n^2/n < \infty$ or $= \infty$. So if the series diverges, then

$$M_n \geq \frac{Z_{n2}}{F_n(Z_{n2}-)} = nZ_{n2} \geq c_n \text{ i.o. a.s.},$$

which proves the lower class part.

Suppose now that the series converges, i.e. that

$$\sum_{n=1}^{\infty} e^{-c_n} \frac{c_n^2}{n} < \infty.$$

Set $a_n = \mu c_n/n$ and $b_n = \max(\mu c_n, \log^3 n)/n$, where $\mu > 1$ (its exact value to be determined later). Setting $m = [\mu + 1]$, we shall consider the supremum in the definition of M_n separately over the four intervals determined by Z_n, Z_{nm}, a_n, b_n and 1. (Note: Frankel's result guarantees that $Z_{nm} < a_n$ eventually a.s., so that all four intervals will eventually be non-empty. Cf. (ii) below).

(i) If $k \geq 1$,

$$\sup_{[Z_{nk},\ Z_{n,k+1}]} \frac{t}{F_n(t)} \leq \frac{Z_{n,k+1}}{F_n(Z_{nk})} = \frac{nZ_{n,k+1}}{k} \equiv Y_{nk}.$$

Now $Y_{n1} = nZ_{n2} < c_n$ eventually a.s., by Frankel. And if $k > 1$ is fixed, then $\Sigma e^{-2c_n}(2c_n)^{k+1}/n < \infty$, so that eventually a.s. we have $nZ_{n,k+1} < 2c_n$ and $Y_{nk} < 2c_n/k \leq c_n$. It follows that whatever the value of μ may be,

(1) $$\sup_{[Z_{n1}\ Z_{nm}]} \frac{t}{F_n(t)} < c_n \quad \text{eventually} \quad \text{a.s.}$$

(ii) Since $\Sigma e^{-\mu c_n}(\mu c_n)^m/n < \infty$, we have $Z_{nm} < a_n$ eventually a.s. In other words, the interval $[Z_{nm}, a_n]$ is eventually non-empty, a.s., and

(2) $$\sup_{[Z_{nm}\ a_n]} \frac{t}{F_n(t)} \leq \frac{a_n}{F_n(Z_{nm})} = \frac{na_n}{m} = \frac{\mu c_n}{m} < c_n.$$

(iii) Define $m_n = [e^{n/\log n}]$, and for convenience adopt the notation $c_{m_n} = c(m_n)$. We first note that $m_{n+1}/m_n \to 1$ as $n \to \infty$. Now, for n so large that $m_{n+1}/m_n < e$ and all monotonicity conditions are fulfilled, and for i such that

$$log(\mu c(m_{n+1})) - 2 \leq i \leq log(m_n b(m_n)),$$

set

$$p_{ni} = P(\sup_{\substack{e^i/m_n \leq t \leq e^{i+1}/m_n \\ m_n \leq k \leq m_{n+1}}} \frac{t}{F_k(t)c_k} \geq 1).$$

Since $(a_n)_{n\leq 1}$ and $(b_n)_{n\leq 1}$ are eventually decreasing, and since $[a(m_{n+1}), b(m_n)] \subset \cup_i [e^i/m_n, e^{i+1}/m_n]$ we have

$$P\Big(\sup_{\substack{a_k\leq t\leq b_k \\ m_n\leq k\leq m_{n+1}}} \frac{t}{F_k(t)c_k} \geq 1\Big) \geq P\Big(\sup_{\substack{a(m_{n+1})\leq t\leq b(m_n) \\ m_n\leq k\leq m_{n+1}}} \frac{t}{F_k(t)\,c_k} \geq 1\Big) \leq$$

$$\leq \sum_i p_{ni}$$

eventually, provided the set of i's in the summation is non-empty. But for this, note that for large n

$$\frac{c(m_{n+1})}{m_{n+1}} \leq \frac{c(m_n)}{m_n} \Rightarrow \frac{c(m_{n+1})}{c(m_n)} \leq \frac{m_{n+1}}{m_n} < e,$$

so that $c(m_{n+1}) \leq ec(m_n)$ and therefore $\log(\mu c(m_{n+1})) - 1 \leq \log(m_n\, b(m_n))$, i.e. there exists at least one such i.

So if we can show that $\sum_{n,i} p_{ni} < \infty$, Borel-Cantelli will yied

$$\sup_{[a_n,\,b_n]} \frac{t}{F_n(t)} < c_n \quad \text{eventually} \quad \text{a.s.} \tag{3}$$

Now

$$p_{ni} \leq P\Bigg(\sup_{\substack{e_i/m_n\leq t\leq e_i{}^{+1}/m_n \\ m_n\leq k\leq m_{n+1}}} \frac{e^{i+1}}{kF_k(t)} \geq \frac{m_n\,c(m_n)}{m_{n+1}}\Bigg)$$

$$= P\left(\frac{m_{n+1}\,e^{i+1}}{m_n\,c(m_n)} \geq m_n\,F_m\left(\frac{e^i}{m_n}\right)\right),$$

where we have used the fact that $kF_k(t) =$ number of $X_1, \ldots, X_k \leq t$. Since $m_{n+1}/m_n < e$, and since binomial $(n, p) - np$ has the same distribution as $n(1-p)$ − binomial $(n, 1-p)$,

$$p_{ni} \leq P\left(F_{m_n}\left(\frac{e^i}{m_n}\right) - \frac{e^i}{m_n} \leq \frac{e^{i+2}}{m_n\,c(m_n)} - \frac{e^i}{m_n}\right) =$$

$$= P\left(F_m\left(1-\frac{e^i}{m_n}\right) - \left(1-\frac{e^i}{m_n}\right) \geq \frac{e^i}{m_n} - \frac{e^{i+2}}{m_n\,c(m_n)}\right)$$

Since $e^2/c(m_n) \downarrow 0$, we have

$$p_{ni} \leq P\left(F_{m_n}\left(1-\frac{e^i}{m_n}\right) - \left(1-\frac{e^i}{m_n}\right) \geq \frac{e^{i-1}}{m_n}\right),$$

eventually.

Now apply Bennett's inequality ([2], formula (2.12), using for convenience $b = 1$), to get

$$p_{ni} \leq \exp\left(-e^{i-1}\,h\left(\frac{1}{e(1-e^i/m_n)}\right)\right),$$

where $h(\lambda) = (1 + 1/\lambda)\log(1 + \lambda) - 1 > 0$ for $\lambda > 0$. Now $e^i/m_n \leq b(m_n) \downarrow 0$, so that we may assume $0 < 1 - e^i/m_n < 1$. Since $e^{i-1} \geq \mu\, c(m_{n+1})\, e^{-3}$ for the i's being considered, and since h is monotone increasing (see [4], lemma 2.5), we have

$$p_{ni} \leq \exp\left(-\frac{\mu c(m_{n+1})\,h(1/e)}{e^3}\right) \equiv r_n$$

uniformly in i for large n.

It follows that

$$\sum_{n,i} p_{ni} \leq \sum_n r_n \log(m_n\, b(m_n))$$

$$\leq \sum_n r_n \log(\mu c(m_n)) + \sum_n r_n \log(\log^3 m_n),$$

where the sums are taken over n sufficiently large. Now convergence of $\Sigma\, e^{-c_n} c_n^2/n$ implies convergence of $\Sigma\, e^{-c_n} c_n/n$, and so by lemma 8 of Robbins and Siegmund (1972), $\Sigma \exp(-c(m_n)) < \infty$. This implies convergence of the first series above if $\mu h(1/e)e^{-3} > 1$ (recall that $c(m_n) \leq c(m_{n+1})$). And since $\liminf_{n\to\infty}(c_n/\log\log n) \geq 1$ (see the proof of lemma 8 in [6]), we have

$$c(m_n) \geq \frac{1}{2}\log\log m_n = \frac{1}{6}\log(\log^3 m_n) \quad \text{eventually,}$$

i.e. the second series is bounded by $\Sigma\, 6\, r_n\, c(m_n)$, which again converges if $\mu h(1/e)\, e^{-3} > 1$. In summary, if we choose μ satisfying $\mu > e^3/h(1/e)$, (3) holds.

(iv) By lemma 3.3 of [4], there is an $a > 0$ such that

$$\left(\frac{n}{\log\log n}\right)^{1/2} |\, F_n(t) - t \,| < t^{1/2} \log\frac{1}{t}$$

for all $t \in (0, a]$, eventually a.s. Since $nb_n \geq \log^3 n$, we see that with probability one, for n large and $t \in [b_n, a]$,

$$\frac{F_n(t)}{t} > 1 - \left(\frac{\log\log n}{nt}\right)^{1/2} \log\frac{1}{t}$$

$$\geq 1 - \left(\frac{\log\log n}{nb_n}\right)^{1/2} \log\frac{1}{b_n} \geq 1 - \left(\frac{\log\log n}{\log^3 n}\right)^{1/2} \log n.$$

This list expression converges to one as $n \to \infty$, and since Glivenko-Cantelli says

$$\sup_{[a,1]} \frac{t}{F_n(t)} \to 1 \quad \text{a.s.},$$

we conclude that

$$\sup_{[b_n,1]} \frac{t}{F_n(t)} \to 1 \quad \text{a.s.} \tag{4}$$

Formulas (1) through (4) now imply the theorem.

3. Some corollaries.

An immediate consequence of the theorem is the LIL for M_n, since $c_n \equiv (1+\varepsilon)\log\log n$ belongs to the upper (lower) class if $\varepsilon > 0 (= 0)$:

Corollary 1. (*Lemma* 4 *of* [3]).

$$\limsup_{n\to\infty} \frac{M_n}{\log\log n} = 1 \quad \text{a.s.}$$

Another easy consequence is the following corollary, which gives a functional bound for $t/F_n(t)$ near the origin:

Corollary 2 $\displaystyle \limsup_{n\to\infty} \sup_{[Z_n\, e^{-e}]} \frac{t}{F_n(t)\log\log 1/t} = 1 \quad \text{a.s.}$

Proof. Let $d_n = \log^3 n/n$ and take the supremum separately over the intervals $[Z_{n1}, d_n]$ and $[d_n, e^{-e}]$. By the theorem and the fact that

$$\inf_{[0\ d_n]} \log\log\frac{1}{t} = \log\log\frac{n}{\log^3 n} \sim \log\log n,$$

we have that for $\gamma > 1$,

$$\limsup_{n\to\infty} \sup_{[Z_{n1},d_n]} \frac{t}{F_n(t)\log\log 1/t} \leq$$

(5)

$$\limsup_{n\to\infty} \sup_{[Z_{n1},d_n]} \frac{t\gamma}{F_n(t)\log\log n} \leq \gamma \quad \text{a.s.}$$

For the second interval, we see that formula (4) holds with b_n substituted by d_n (the only property of b_n which was used was $b_n \geq \log^3 n/n = d_n$), i.e.

$$\sup_{[d_n,1]} \frac{t}{F_n(t)} \to 1 \quad \text{a.s.}$$

Since $1 = \log\log e^e \leq \log\log 1/t$ for $t \in [d_n, e^{-e}]$, and since $e^{-e}/F_n(e^{-e}) \to 1$ a.s., we have

$$\limsup_{n\to\infty} \sup_{[d_n,e^{-e}]} \frac{t}{F_n(t)\log\log 1/t} = 1 \quad \text{a.s.} \tag{6}$$

Expressions (5) and (6) now imply the corollary.

References

[1] Frankel, J. (1976). *A note on downcrossings for extremal processes. Ann. Probability* 4, 151-152.

[2] Hoeffding, W. (1963). *Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc.* 58, 13-30.

[3] James, B. R. (1971). *A functional law of the iterated logarithm weighted empirical distributions.* Ph. D. dissertation, University of California, Berkeley.

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

[5] Kiefer, J. (1972). *Iterated logarithm analogues for the sample quantiles when $p_n \downarrow 0$. Proc. 6th Berkeley Symp. Math. Statist. Prob.* I, 227-244.

[6] Robbins, H. and Siegmund, D. O. (1972). *On the law of the iterated logarithm for maxima and minima. Proc. 6th Berkeley Symp. Math. Statist. Prob.* II, 51-70.

Instituto de Matemática Pura e Aplicada
Rio de Janeiro