The Annals of Probability
1982, Vol. 10, No. 3, 863-868

A LOG LOG LAW FOR MAXIMAL UNIFORM SPACINGS[1]

BY LUC DEVROYE

McGill University

Let $X_1, X_2, \cdots$ be a sequence of independent uniformly distributed random variables on [0, 1] and K_n be the kth largest spacing induced by $X_1, \cdots, X_n$. We show that $P(K_n \le (\log n - \log_3 n - \log 2)/n \text{ i.o.}) = 1$ where $\log_j$ is the j times iterated logarithm. This settles a question left open in Devroye (1981). Thus, we have

$$\liminf(nK_n - \log n + \log_3 n) = -\log 2 \quad \text{almost surely,}$$

and

$$\limsup(nK_n - \log n)/2\log_2 n = 1/k \quad \text{almost surely.}$$

1. Introduction. Consider a sequence $X_1, X_2, \cdots$ of independent identically distributed random variables with a uniform distribution on [0, 1], and let $S_1(n), \cdots, S_{n+1}(n)$ be the spacings induced by $X_1, \cdots, X_n$ on [0, 1]. Let K_n be the kth largest spacing among $S_i(n)$, $1 \le i \le n+1$. Devroye (1981) has shown that

$$\limsup(nK_n - \log n)/(2\log_2 n) = 1/k \quad \text{a.s.,} \tag{1.1}$$

and that

$$\liminf(nK_n - \log n + \log_3 n) = c \quad \text{a.s.} \tag{1.2}$$

where $-\log 2 \le c \le 0$. The strong upper bound (1.1) is now completely known for the case $k = 1$. In fact, we have for $p \ge 4$,

$$P\left(nK_n \ge \log n + \frac{2}{k}\log_2 n + \log_3 n + \cdots + \log_{p-1} n + (1+\delta)\log_p n \text{ i.o.}\right)$$
$$= \begin{cases} 0 & \text{when } \delta > 0 \text{ (Devroye, 1981)} \\ 1 & \text{when } \delta < 0 \text{ and } k = 1 \text{ (Deheuvels, 1982).} \end{cases}$$

The purpose of this paper is to show that the constant c in (1.2) is $-\log 2$.

THEOREM. *Let M_n be the maximal spacing among $S_i(n)$, $1 \le i \le n+1$. Then*

$$P(M_n \le (\log n - \log_3 n - \log 2)/n \text{ i.o.}) = 1.$$

COROLLARY *Since $K_n \le M_n$, we may combine this result with Theorem* 4.2 *of Devroye* (1981):

$$P(K_n \le (\log n - \log_3 n - \log 2 - \delta)/n \text{ i.o.}) = \begin{cases} 1 & \textit{when } \delta = 0 \\ 0 & \textit{when } \delta > 0. \end{cases}$$

2. Some Lemmas.

LEMMA 2.1. [Tail of the gamma distribution] (Devroye, 1981, Lemma 3.1). *If X is gamma (n) distributed, then for all $\varepsilon > 0$,*

$$P(X < n(1-\varepsilon)) \le \exp(-n\varepsilon^2/2).$$

Received June 1981.
[1] This research was sponsored in part by National Research Council of Canada Grant No. A3456.
AMS 1980 *subject classification.* Primary 60F15.
Key words and phrases. Law of the iterated logarithm, uniform spacings, strong laws, almost sure convergence, order statistics.

LEMMA 2.2. [Tail of the binomial distribution] (Dudley, 1978, page 907). *If X is a binomial (n, p) random variable where $n \geq 1$, $p \in (0, 1)$, then*

$$P(X \geq k) \leq \left(\frac{np}{k}\right)^k e^{k-np}, \qquad k \geq np, \quad k \text{ integer.}$$

PROOF. See Dudley (1978). The proof is based upon one of Okamoto's inequalities (Okamoto, 1958).

LEMMA 2.3. [Tail of the binomial distribution]. *If X is a binomial (n, p) random variable where $n \geq 1$, $p \in (0, 1)$, then*

$$P(X \geq np + \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2np} + \frac{\varepsilon^3}{2n^2p^2}\right), \quad \varepsilon > 0, \quad np \geq e.$$

PROOF. We use Lemma 2.2 and note that $(np/k)^k e^{k-np}$ is decreasing in k for $k > e$. Thus, by the inequality $\log(1 + u) > u - u^2/2$, $u > 0$,

$$P(X \geq np + \varepsilon) \leq \left(\frac{np}{np + \varepsilon}\right)^{np+\varepsilon} e^{\varepsilon} \leq \exp\left(-(np + \varepsilon)\left(\frac{\varepsilon}{np} - \frac{\varepsilon^2}{2n^2p^2}\right) + \varepsilon\right)$$
$$= \exp\left(-\frac{\varepsilon^2}{2np} + \frac{\varepsilon^3}{2n^2p^2}\right).$$

LEMMA 2.4. [Inequality for the multinomial distribution]. *If $X_1, \cdots, X_n$ are* i.i.d. *random variables uniformly distributed on* $[0, 1]$ *and $N_1, \cdots, N_k$ are the number of X_i's in the intervals $(0, a)$, $(a, 2a)$, $\cdots$, $((k-1)a, ka)$ respectively where $ka \leq 1$, $k \geq 1$, $a \geq 0$, then*

$$(1 - (1 - a)^n)^k \geq P(\min_{1 \leq i \leq k} N_i \geq 1)$$
$$\geq (1 - \exp(-an(1 - \varepsilon)))^k - \exp(-n\varepsilon^2/2), \quad \text{all} \quad \varepsilon \in (0, 1).$$

PROOF. The upper bound follows from Mallows' inequality (Mallows, 1968)

$$P(\min_{1 \leq i \leq k} N_i \geq 1) \leq \textstyle\prod_{i=1}^k P(N_i \geq 1).$$

The lower bound can be obtained by considering the i.i.d. sequence $X_1, X_2, \cdots$ of uniform $[0, 1]$ random variables, and an independent Poisson $(n(1 - \varepsilon))$ random variable Z. Clearly, $X_1, \cdots, X_Z$ can be considered as the arrival times in a homogeneous Poisson point process on $[0, 1]$ with intensity $n(1 - \varepsilon)$. Also, if $N'_1, \cdots, N'_k$ are the cardinalities of the intervals $(0, a)$, $(a, 2a)$, $\cdots$, $((k-1)a, ka)$ obtained from $X_1, \cdots, X_Z$, then

$$P(\min_{1 \leq i \leq k} N'_i \geq 1) = (1 - \exp(-an(1 - \varepsilon)))^k \leq P(\min_{1 \leq i \leq k} N_i \geq 1) + P(Z > n).$$

If G is a gamma (n) random variable, then, by Lemma 2.1,

$$P(Z \geq n) \leq P(G < n(1 - \varepsilon)) \leq \exp(-n\varepsilon^2/2).$$

LEMMA 2.5. *Let $u > 0$ and let $k \geq 1$ be integer. If K_n is the k*th *largest spacing $S_i(n)$, $1 \leq i \leq n + 1$, then*

$$P(K_n > u) \leq e^{-\sqrt{n/2}} + P(Z \geq k)$$

where Z is a binomial (p, n) random variable and $p = e^{-un}e^{un^{3/4}}$.

PROOF. We use the fact that $\{S_i(n), 1 \leq i \leq n + 1\}$ is distributed as $\{E_i/T, 1 \leq i \leq n + 1\}$ where $E_1, \cdots, E_{n+1}$ are i.i.d. exponentially distributed random variables and $T = \sum_{i=1}^{n+1} E_i$. If $E_{(k)}$ is the kth largest of the E_i's, then

$$P(K_n > u) = P(E_{(k)} > u \textstyle\sum_{i=1}^{n+1} E_i) \le P(\sum_{i=1}^{n+1} E_i < n - n^{3/4}) + P(E_{(k)} > u(n - n^{3/4}))$$
$$\le \exp(-\sqrt{n}/2) + P(Z \ge k)$$

by Lemma 2.1.

LEMMA 2.6. [A strong law for the k_nth largest spacing].
Let

$$u_n = (\log n - (1 + c)\log_3 n - \log 2)/n, \qquad c \ge 2,$$
$$p_n = \exp(-nu_n + n^{3/4}u_n),$$
$$\delta_n = \sqrt{2np_n} \cdot \sqrt{2\log_2 n + (2 + c + \theta)\log_3 n}, \qquad \theta > 0,$$

and
$$k_n = \lceil np_n + \delta_n \rceil \qquad (\lceil\ \rceil \text{ is the ceiling function}).$$

*If K_n is the k_n*th *largest spacing among $S_i(n)$, $1 \le i \le n + 1$, then*

$$P(K_n > u_n \quad \text{f.o.}) = 1.$$

NOTE. We will need good asymptotic estimates of p_n, δ_n and k_n in what follows. A quick check shows that

$$p_n = \frac{2(\log_2 n)^{1+c}}{n}\left(1 + O\left(\frac{\log n}{n^{1/4}}\right)\right),$$
$$\delta_n = (\sqrt{8} + o(1))(\log_2 n)^{1+c/2},$$

and

$$k_n = 2(\log_2 n)^{1+c}\left(1 + O\left(\frac{\log n}{n^{1/4}}\right)\right) + O((\log_2 n)^{1+c/2})$$
$$= 2(\log_2 n)^{1+c}(1 + O((\log_2 n)^{-c/2}))$$
$$\sim 2(\log_2 n)^{1+c}.$$

PROOF. Note that u_n and k_n are monotone for $n > N$. Thus, for $n > N$, we have

$$P(K_n > u_n, K_{n+1} \le u_{n+1}) \le \begin{cases} P(K_n > u_n)2k_n u_{n+1}, & k_n = k_{n+1}, \\ P(K_n > u_n) & k_n < k_{n+1}. \end{cases}$$

By Lemma 1* of Barndorff-Nielsen (1961), it suffices to show that

$$P(K_n > u_n) \to 0 \quad \text{as} \quad n \to \infty, \tag{2.1}$$

and that

$$\textstyle\sum_{n=1}^{\infty} P(K_n > u_n, K_{n+1} \le u_{n+1}) < \infty. \tag{2.2}$$

By Lemma 2.5, $P(K_n > u_n) \le O(\exp(-n^{1/3})) + P(Z \ge k_n)$ where Z is binomial (p_n, n). By Lemma 2.3, $P(Z \ge k_n) \le P(Z \ge np_n + \delta_n) \le \exp(-\delta_n^2/(2np_n) + \delta_n^3/(2n^2p_n^2))$. Now,

$$\delta_n^3/(2n^2p_n^2) \sim \sqrt{8}\,(\log_2 n)^{(1-c/2)}.$$

Thus, if $b = e^{\sqrt{8}}$,

$$P(Z \ge k_n) \le (b + o(1))/((\log n)^2(\log_2 n)^{2+c+\theta}),$$

so that (2.1) holds. Furthermore,

$$P(K_n > u_n)k_n u_{n+1} \le \frac{(b + o(1))}{(\log n)^2(\log_2 n)^{2+c+\theta}} \cdot (2 + o(1))(\log_2 n)^{1+c} \cdot \frac{\log n}{n}$$
$$= \frac{2b + o(1)}{n \log n\ (\log_2 n)^{1+\theta}},$$

which is summable in n. To conclude the proof of (2.2), we need only show that

$$\sum_{n: k_n < k_{n+1}} P(K_n > u_n) < \infty.$$

Clearly, $k_n \le 3(\log_2 n)^{1+c}$ for all n large enough. For such n, we have $\log n \ge \exp((k_n/3)^{1/(1+c)})$. By our upper bounds for $P(K_n > u_n)$ obtained above it suffices to check that

$$\sum_{n: k_n < k_{n+1}} (\log n)^{-2}(\log_2 n)^{-(2+c+\theta)} \le \sum_{j=1}^{\infty} \exp(-2(j/3)^{1/(1+c)})(j/3)^{-\frac{2+c+\theta}{1+c}} < \infty.$$

This concludes the proof of Lemma 2.6.

3. Proof of the theorem. The proof is based upon the following implication:

$$\begin{aligned} &[M_n < (\log n - \log_3 n - \log 2)/n \text{ i.o.}] \\ &\supset [K_{n_j} > (\log n_j - (1 + c)\log_3 n_j - \log 2)/n_j \text{ f.o.}] \\ &\cap [A_{n_j} \text{ i.o.}] \cap [M_n > (\log n + 3 \log_2 n)/n \text{ f.o.}] \end{aligned} \tag{3.1}$$

where

(i) n_j is a monotone subsequence such that $n_{j+1} - n_j > \rho_{n_j}$, all j large enough, and

$$\rho_n = \lfloor cn \log_3 n/\log n \rfloor, \qquad \text{some} \quad c \ge 2$$

($\lfloor\ \rfloor$ is the floor function);

(ii) K_n, p_n, δ_n are defined as in Lemma 2.6;

(iii) A_n is defined as follows: let $m_n = (\log n - 4 \log_2 n)/n$. Let $B_1, \cdots, B_{k_n}$ disjoint sets of $[0, 1]$ with the property that each B_i is a finite union of intervals whose boundaries are measurable functions of $X_1, \cdots, X_n$ only; each B_i has Lebesgue measure m_n; and B_i either covers the ith largest spacing among $S_i(n)$, $1 \le i \le n + 1$, or covers the interval of length m_n centered at the middle of this spacing (when the spacing itself is larger than m_n). We let A_n be the event [all the B_i's, $1 \le i \le k_n$, are occupied by at least one X_i from $X_{n+1}, \cdots X_{n+\rho_n}$].

In (3.1) we are using the fact that if A_{n_j} occurs, $M_{n_j} \le (\log n_j + 3 \log_2 n_j)/n_j$, and $K_{n_j} \le (\log n_j - (1 + c)\log_3 n_j - \log 2)/n_j$, then

$$\begin{aligned} M_{n_j+\rho_{n_j}} &\le K_{n_j} \le (\log n_j - (1 + c)\log_3 n_j - \log 2)/n_j \\ &\le (\log(n_j + \rho_{n_j}) - \log_3(n_j + \rho_{n_j}) - \log 2)/(n_j + \rho_{n_j}). \end{aligned} \tag{3.2}$$

The last inequality in (3.2) follows from our choice of ρ_n because

$$\frac{n + \rho_n}{n}(\log n - (1 + c)\log_3 n - \log 2) - (\log(n + \rho_n) - \log_3(n + \rho_n) - \log 2)$$
$$\le \frac{\rho_n}{n}\log n - c\frac{n + \rho_n}{n}\log_3 n \le c \log_3 n(1 - 1) - c\frac{\rho_n \log_3 n}{n} \le 0.$$

The first inequality in (3.2) is valid because each of the k_{n_j} largest intervals among $S_i(n_j)$, $1 \le i \le n_j + 1$, is either smaller than m_{n_j} or is split into two intervals of length at most $(1/2)(m_{n_j} + (\log n_j + 3 \log_2 n_j)/n_j) = (\log n_j - (1/2)\log_2 n_j)/n_j$. In either case, for n_j large enough, all the new intervals at time $n_j + \rho_{n_j}$ are smaller than $(\log n_j - (1/2)\log_2 n_j)/n_j \le K_{n_j}$.

We have to show that the three events on the right-hand side of (3.1) have probability one. By Lemma 2.6,

$$P(K_{n_j} > (\log n_j - (1 + c)\log_3 n_j - \log 2)/n_j \quad \text{f.o.}) = 1.$$

By (1.1),

$$P(M_n > (\log n + 3 \log_2 n)/n \quad \text{f.o.}) = 1.$$

The Theorem follows if $P(A_{n_j}$ i.o.$) = 1$. Let $\mathscr{F}_j$ be the σ-algebra generated by $A_{n_1}, \cdots, A_{n_j}$ (i.e., it is the σ-algebra generated by $X_1, X_2, \cdots, X_{n_j+\rho_{n_j}}$). *Since* $n_{j+1} - n_j > \rho_{n_j}$ for j large enough, we have

$$P(A_{n_j} \mid \mathscr{F}_{j-1}) = P(A_{n_j}) \quad \text{a.s.}$$

for all large j. Thus, $P(A_{n_j}$ i.o.$) = 1$ when

$$\sum_{j=1}^{\infty} P(A_{n_j}) = \infty \tag{3.3}$$

(see for example Serfling (1975), Theorem 2 or Iosifescu and Theodorescu (1969), page 2, for a more general statement of this type). We are still free to choose n_j within condition (i). Let us define

$$n_j = \exp(\sqrt{2c'j \log_2 j}), \quad \text{some } c' > c.$$

Let us first check that $n_{j+1} - n_j > \rho_{n_j}$ for all j large enough. A trivial analysis shows that

$$\rho_{n_j} \sim c n_j \log_3 n_j / \log n_j \sim n_j \sqrt{(\log_2 j/2j)}\, c/\sqrt{c'}$$

Also,

$$\begin{aligned} n_{j+1} - n_j &\geq n_j[\exp(\sqrt{2c'(j+1)\log_2(j+1)} - \sqrt{2c'j \log_2 j}) - 1] - 1 \\ &\sim n_j[\sqrt{2c'(j+1)\log_2(j+1)} - \sqrt{2c'j \log_2 j}] - 1 \\ &\geq n_j \log n_j [1 + o(1)][\sqrt{1 + 1/j} - 1] - 1 \\ &\sim n_j \log n_j / 2j \\ &\sim n_j \sqrt{(\log_2 j/2j)}\sqrt{c'}. \end{aligned}$$

Thus, (i) holds in view of $\sqrt{c'} > c/\sqrt{c'}$.

We conclude the proof by showing that for this choice of n_j, (3.3) holds. A helpful lower bound for $P(A_n)$ is provided in Lemma 2.4 if we set $\varepsilon := n^{-1/4}$, $a := (\log n - 4 \log_2 n)/n$, $n := \rho_n$ and $k := k_n$ in the formal inequality obtained there. This gives

$$P(A_n) \geq \left(1 - \exp\left(-\left(\frac{\log n - 4 \log_2 n}{n}\right)\rho_n(1 - n^{-1/4})\right)\right)^{k_n} - \exp(-\rho_n/2\sqrt{n}).$$

We note that

$$\left(\frac{\log n - 4 \log_2 n}{n}\right)\rho_n(1 - n^{-1/4}) \geq c \log_3 n - \frac{5c \log_2 n \log_3 n}{\log n} \geq \frac{c}{2} \log_3 n,$$

all n large enough.

Also, $\exp(-\rho_n/2\sqrt{n}) \leq \exp(-c\sqrt{n} \log_3 n/2 \log n) \leq \exp(-n^{1/3})$ for n large enough. By combining these estimates, and using the inequality $\log(1 - u) \geq -u/(1 - u)$, $u \in (0, 1)$, we have

$$\begin{aligned} P(A_n) &\geq \exp(-k_n \exp(-[c \log_3 n - 5c \log_2 n \log_3 n/\log n])/(1 - \exp(-(c/2) \log_3 n))) \\ &\quad - \exp(-n^{1/3}) \\ &\geq \exp(-2 \log_2 n(1 + O((\log_2 n)^{-c/2}))) - \exp(-n^{1/3}). \end{aligned}$$

We used the asymptotic estimate for k_n given in the Note following Lemma 2.6. Replacing n by n_j gives

$$P(A_{n_j}) \geq \exp(-2\log\sqrt{2c'j\log_2 j}(1 + O((\log j)^{-c/2}))) - \exp(-n_j^{1/3})$$

$$= \left[\frac{1}{2c'j\log_2 j}\right]^{1+O((\log j)^{-c/2})} - O(e^{-j}).$$

The last expression is not summable in j when $c' > 0$, $c \geq 2$. This concludes the proof of (3.3) and the Theorem.

REFERENCES

[1] BARNDORFF-NIELSEN, O. (1961). On the rate of growth of the partial maxima of a sequence of independent identically distributed random variables. *Math. Scand.* **9** 383–394.

[2] DEHEUVELS, P. (1982). Strong limiting bounds for maximal uniform spacings. *Ann. Probability* **10**, to appear.

[3] DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Probability* **9** 860–867.

[4] DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probability* **6** 899–929.

[5] IOFISESCU, M. and THEODORESCU, R. (1969). *Random Processes and Learning.* Springer-Verlag, New York.

[6] MALLOWS, C. L. (1968). An inequality involving multinomial probabilities. *Biometrika* **55** 422–424.

[7] OKAMOTO, M. (1958). Some inequalities related to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** 29–35.

[8] SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probability* **3** 726–731.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTREAL, CANADA H3A2K6