# THE LARGEST EXPONENTIAL SPACING

Luc Devroye*

ABSTRACT. We consider the largest spacing $M_n$ defined by $n$ independent exponentially distributed random variables. We give its limit law, obtain some large deviation probabilities and derive some laws of the iterated logarithm. For example, it is shown that

$$\liminf_{n\to\infty} M_n \log\log n = \pi^2/6 \text{ almost surely,}$$

and that if $x_n \uparrow \infty$, $P(M_n > x_n \text{ i.o.}) = 0$ or $1$ according to $\sum_{n=1}^{\infty} \frac{1}{n} \exp(-x_n) < \infty$ or $= \infty$.

## 0. *Introduction.*

The maximal spacing defined by a sample of size $n$ drawn from the uniform distribution on $[0,1]$ is close to $\frac{\log n}{n}$. Its exact distribution (Whitworth (1897)), asymptotic distribution (Levy (1939)), large deviation properties (Devroye (1981)) and almost sure behavior (Devroye (1981,1982) and Deheuvels (1982)) are well-known. When the data are not uniformly distributed, but have a density $f$ on $[0,1]$ that stays bounded away from 0 and satisfies some smoothness conditions, the maximal spacing will behave as for a properly normalized uniform distribution. But if one considers for example a density $f$ with support $[0,\infty)$ and nonincreasing hazard rate, the spacings in the tail tend to be larger than the other spacings, and the problem becomes asymmetric.

---

*Research of the author was sponsored in part by National Research Council of Canada Grant No. A3456.

Partially because it has a constant hazard rate, and partially because it occupies an important place in probability and statistics, we will take a close look at the exponential distribution. In deriving the properties of the maximal spacing, a good understanding of the asymmetry inherent in the problem is needed.

In what follows, we let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent exponentially distributed random variables. Let $X_{(1,n)} < \cdots < X_{(n,n)}$ be the order statistics of $X_1, \ldots, X_n$, and let $S_{(i,n)}$ be the spacings $X_{(i,n)} - X_{(i-1,n)}$, $1 \le i \le n$, where $X_{(0,n)} = 0$ by convention. Finally, we set $M_n = \max_{1 \le i \le n} S_{(i,n)}$.

## 1. *The Limit Distribution.*

One of the crucial properties needed in this note is due to Sukhatme (1937) (see also the survey paper by Pyke (1965)):

LEMMA 1.1. $S_{(n,n)}, \ldots, S_{(1,n)}$ *are distributed as* $Y_1/1, Y_2/2, \ldots, Y_n/n$ *where* $Y_1, \ldots, Y_n$ *are independent exponentially distributed random variables.*

Because

$$P(M_n < x) = P\left(\bigcap_{i=1}^{n} S_{(i,n)} < x\right) = \prod_{i=1}^{n} P(Y_i < ix)$$

$$= \prod_{i=1}^{n} (1-e^{-ix}) \downarrow \prod_{i=1}^{\infty} (1-e^{-ix}), \text{ all } x > 0,$$

we have proved

LEMMA 1.2. [Limit distribution.]

$$\lim_{n\to\infty} P(M_n < x) = F(x) = \prod_{i=1}^{\infty} (1-e^{-ix}), \ x > 0. \quad (1)$$

It is easy to check that $F$ has a density. Thus, the maximal spacing, unnormalized, tends in distribution to a nondegenerate random variable. Since this limit random variable has support $[0,\infty)$, with a little work one can show that $M_n$ oscillates wildly:

$\limsup_{n\to\infty} M_n = \infty$ a.s. and $\liminf_{n\to\infty} M_n = 0$ a.s. . In order to be able to be more specific about how fast $M_n$ oscillates, we need a few large deviation results.

LEMMA 1.3. [Large deviations.]

*Let* F *be the limit distribution function of Lemma* 1.2. *Then*

$$F(x) \sim \exp\left(-\frac{1}{x}\left(\frac{\pi^2}{6} + o(1)\right)\right) \quad \textit{as} \quad x \downarrow 0, \quad \textit{and} \tag{2}$$

$$1-F(x) \sim \exp(-x) \quad \textit{as} \quad x \to \infty . \tag{3}$$

*Let* $x_n$ *be a sequence of positive numbers. Then*

$$P(M_n < x_n) \sim \exp\left(-\frac{1}{x_n}\left(\frac{\pi^2}{6} + o(1)\right)\right) \textit{ when } \lim_{n\to\infty} x_n = 0,\ \lim_{n\to\infty} \frac{nx_n}{\log n} = \infty \tag{4}$$

*and*

$$P(M_n > x_n) \sim \exp(-x_n) \quad \textit{when} \quad \lim_{n\to\infty} x_n = \infty . \tag{5}$$

*Proof of Lemma 1.3.*

Let $x > 0$. Then

$$e^{-x} = 1-(1-e^{-x}) \le 1-F(x) \le 1-\Big(1-\sum_{i=1}^{\infty} e^{-ix}\Big) = \sum_{i=1}^{\infty} e^{-ix}$$

$$= e^{-x}/(1-e^{-x}) \sim e^{-x} \ (\text{as} \ \ x \to \infty).$$

This proves (3). To prove (2), we let $x \downarrow 0$, and choose integers $a = a(x)$ such that $a \to \infty$ and $ax \to 0$ as $x \downarrow 0$. Now,

$\log F(x) = \sum_{i=1}^{\infty} \log(1-e^{-ix}) = \sum_{i\le a} + \sum_{i>a}$ . By the inequality $n! > (\frac{n}{e})^n$, valid for $n \ge 1$ (this is a special form of Stirling's inequality and can for example be deduced from Buchner's inequality (Buchner (1951); Mitrinović (1970, pp. 181-185)), we have

$$0 \ge \sum_{i\le a} \log(1-e^{-ix}) \ge \sum_{i\le a} \log(ix) = a \log x + \log a! \ge a \log\left(\frac{ax}{e}\right)$$

$$= \frac{1}{x}\left(ax \log\left(\frac{ax}{e}\right)\right) = o\left(\frac{1}{x}\right).$$

Also,

$$0 \le \sum_{i>a} \log(1-e^{-ix}) - \int_a^\infty \log(1-e^{-ux})du \le -\log(1-e^{-ax})$$

$$= o(\frac{1}{ax}) = o(\frac{1}{x}) .$$

Next, by the integral $\int_0^1 (\log y)/(1-y)dy = -\pi^2/6$ (Gradshteyn and Ryzhik (1980, formula 4.231.2)), we have

$$\int_a^\infty \log(1-e^{-ux})du = \frac{1}{x}\int_{1-\exp(-ax)}^{1} (\log y)/(1-y)dy = \frac{1}{x}(-\frac{\pi^2}{6} + o(1)).$$

Combining all the estimates gives us $\log F(x) = \frac{1}{x}(-\pi^2/6 + o(1))$ as $x \downarrow 0$.

Results (4) and (5) follow with little extra work. Let $x_n$ be a positive number sequence. Clearly,

$\prod_{i=n}^{\infty} (1-e^{-ix_n}) \le F(x_n)/P(M_n < x_n) \le 1$. But $\prod_{i=n}^{\infty} (1-e^{-ix_n})$

$\ge 1- \sum_{i=n}^{\infty} \exp(-ix_n) = 1 - \exp(-nx_n)/(1-\exp(-x_n)) = 1 - (1+o(1))\frac{e^{-nx_n}}{x_n}$

$= 1 - o(1)$ when $nx_n + \log x_n \to \infty$, $x_n \downarrow 0$. The former condition is implied by the condition $nx_n/\log n \to \infty$ . Thus, (4) follows from (2). Finally, (5) follows from

$$e^{-x_n} = 1-(1-e^{-x_n}) \le P(M_n > x_n) \le \sum_{i=1}^{n} e^{-ix_n} = \frac{e^{-x_n}-e^{-(n+1)x_n}}{1-\exp(-x_n)}$$

$$\sim e^{-x_n} .$$

## 2. *The Limit Supremum.*

THEOREM 2.1. *Let* $x_n \uparrow \infty$. *Then*

$$P(M_n > x_n \text{ i.o.}) = \begin{matrix} 0 \\ 1 \end{matrix} \quad \textit{according to} \quad \sum_{n=1}^{\infty} \frac{1}{n} e^{-x_n} \begin{matrix} < \infty \\ \infty \end{matrix}$$

*Proof of the first half of Theorem 2.1.*

We know that $P(M_n > x_n \text{ i.o.}) = 0$ when $P(M_n > x_n) \to 0$ and $\sum_n P(M_n \le x_n, M_{n+1} > x_{n+1}) < \infty$ (Barndorff-Nielsen (1961)). The first condition follows from (5). Also, by the monotonicity of $x_n$,

$$P(M_n \le x_n, M_{n+1} > x_{n+1}) \le P(X_{n+1} \ge \max_{1\le i\le n} X_i + x_{n+1})$$

$$= E(\exp(-\max_{1\le i\le n} X_i - x_{n+1})) = \exp(-x_{n+1})\, E(\min_{1\le i\le n} U_i)$$

$$= (n+1)^{-1} \exp(-x_{n+1})$$

where $U_1,\ldots,U_n$ are independent uniform [0,1] random variables. This shows the first half of Theorem 2.1.

LEMMA 2.1. *Let* $A_n$ *be the event* $[X_n > \max_{i<n} X_i + x_n]$. *Then* $P(A_n) = \frac{1}{n}\exp(-x_n)$ *and* $P(A_n A_j) \le 2P(A_n)P(A_j)$, $n \ne j$, *whenever* $x_n > 0$.

*Proof of Lemma 2.1.*

Define $T_n = X_{(n,n)} = \max_{i\le n} X_i$. Because $e^{-T_n}$ is distributed as $\min(U_1,\ldots,U_n)$ where $U_1,\ldots,U_n$ are independent uniform [0,1] random variables, we have

$$P(A_n) = E(\exp(-T_{n-1}-x_n)) = \frac{1}{n}\exp(-x_n) .$$

Next, if $j < n$, we have

$$P(A_n \cap A_j) = P(X_j > T_{j-1} + x_j,\ X_n > T_{n-1} + x_n)$$

$$= E(I_{X_j > T_{j-1}+x_n} \exp(-T_{n-1}-x_n))$$

$$= \exp(-x_n) . E(I_{X_j > T_{j-1}+x_j} \exp(-T_{n-1}))$$

where I is the indicator function of an event. We note that on $A_j$, $T_{n-1} = \max(X_j,\ldots,X_{n-1})$. In particular, on $A_j$, $T_{n-1}$ is distributed as $X_j + \max(X_1',\ldots,X_N')$ where $X_1',\ldots,X_N'$ are independent exponential random variables and N is binomial $(n-j-1, \exp(-X_j))$. We will now need the fact that for a binomial (n,p) random variable B,

$E(1/(1+B)) = (1-(1-p)^{n+1})/(p(n+1))$. Thus,

$$E(I_{A_j}\exp(-T_{n-1})|X_j,X_{j-1},\dots,X_1) = I_{A_j}\exp(-X_j)\,E(\tfrac{1}{N+1}|X_j)$$
$$= \frac{1}{n-j}(1-(1-\exp(-X_j))^{n-j})\,I_{A_j}\exp(-X_j)\,.$$

Thus,

$$P(A_n\cap A_j) = \exp(-x_n).E(I_{A_j}\frac{1}{n-j}(1-(1-\exp(-X_j))^{n-j}))$$
$$= \frac{\exp(-x_n)}{n-j}\,E\Big(\int_{T_{j-1}+x_j}^{\infty} e^{-y}(1-(1-e^{-y})^{n-j})\,dy\Big)$$
$$= \frac{\exp(-x_n)}{n-j}\,E\Big(\int_{1-\exp(-T_{j-1}-x_j)}^{1}(1-u^{n-j})\,du\Big) \quad \text{(by a change of variables)}$$
$$= \frac{\exp(-x_n)}{n-j}\,E(H(\exp(-T_{j-1}-x_j)))$$

where $H(x) = x - \frac{1}{n-j+1}(1-(1-x)^{n-j+1})$, $0 \le x \le 1$. Using the obvious inequality $H(x) \le x$, we obtain

(6) $$P(A_n\cap A_j) \le \frac{\exp(-x_n)}{n-j}\cdot\exp(-x_j).\ E(\exp(-T_{j-1})) = \frac{e^{-x_n}e^{-x_j}}{(n-j)j}\,.$$

But we also have $H(x) \le \frac{n-j}{2}x^2$ (this follows from $(1-x)^{n-j+1} \le 1-(n-j+1)x + \binom{n-j+1}{2}x^2$). We will need the value $E(\exp(-2T_{j-1}))$:

$$E(\exp(-2T_{j-1})) = E(\min{}^2(U_1,\dots,U_{j-1})) = \int_0^1 2y(1-y)^{j-1}dy = \frac{2}{j(j+1)}\,.$$

Thus,

(7) $$P(A_n\cap A_j) \le \frac{\exp(-x_n)}{n-j}\cdot\frac{n-j}{2}\cdot\exp(-2x_j)\cdot\frac{2}{j(j+1)} \le \frac{e^{-x_n}e^{-x_j}}{j(j+1)}\,.$$

We combine (6) and (7) and note that $\min_j \max(n-j,j+1) \ge \frac{n+1}{2} \ge \frac{n}{2}$. From this and the first part of the Lemma, we conclude that $P(A_n\cap A_j) \le 2\exp(-x_n)\exp(-x_j)/(nj) = 2P(A_n)P(A_j)$, which was to be shown.

*Proof of the second half of Theorem 2.1.*

Clearly, $[M_n > x_n \text{ i.o.}] \supseteq [A_n \text{ i.o.}]$. Since $P(A_n \text{ i.o.})$ is a tail event, we need only show that $P(A_n \text{ i.o.}) > 0$. This follows if we can show that there exists a constant $c > 0$ such that

$$P(\bigcup_{n=m}^{\infty} A_n) \geq c, \quad \text{for all } m.$$

From Lemma 2.1 we remember that $P(A_n) = \exp(-x_n)/n$, and thus that $\sum_{n=1}^{\infty} P(A_n) = \infty$. Lemma 2.1 also implies that $P(\bigcup_{n=m}^{\infty} A_n) \geq \frac{1}{2}$, all $m$: to see this, just apply the Chung-Erdös inequality (Chung and Erdös (1952))

$$P(\bigcup_{n=m}^{\infty} A_n) = \sup_{M>m} P(\bigcup_{n=m}^{M} A_n) \geq \sup_{M>m} \frac{(\sum_{n=m}^{M} P(A_n))^2}{\sum_{n=m}^{M}\sum_{n'=m}^{M} P(A_n A_{n'}) + \sum_{n=m}^{M} P(A_n)}$$

(8)

$$\geq \sup_{M>m} \frac{(\sum_{n=m}^{M} P(A_n))^2}{2(\sum_{n=m}^{M} P(A_n))^2 + \sum_{n=m}^{M} P(A_n)} = \frac{1}{2}.$$

This concludes the proof of Theorem 2.1.

*Remark.* We have shown that

$$\limsup_{n\to\infty} M_n/\log\log n = 1 \quad \text{almost surely.}$$

We have also shown that if $\log_j$ is the $j$ times iterated logarithm, then

$$P(M_n > \log_2 n + \log_3 n + \ldots + \log_{j-1} n + (1+\varepsilon)\log_j n \text{ i.o.}) = 0 \ (1)$$

according to $\varepsilon > 0$ $(\varepsilon \leq 0)$, all $j \geq 2$.

## 3. *The Limit Infimum.*

THEOREM 3.1.

$$\liminf_{n\to\infty} M_n \log\log n = \frac{\pi^2}{6}\text{, almost surely.}$$

*Proof of Theorem 3.1.*

For $\varepsilon > 0$, we define $b_n = \frac{1}{6}\pi^2(1+\varepsilon)/\log\log n$. Let $n_i = [\exp(i \log i)]$ where $[.]$ denotes the largest integer contained in a given real number. Let $M_i^*$ be the largest spacing defined by the subsample $X_j$, $n_{i-1} < j \le n_i$ only! We know that

$$P(\text{There exists } i_0 \text{ such that for all } i \ge i_0 (n_{i-1}, n_i] \text{ contains a record, i.e. } X_j = \max(X_1,\ldots,X_j) \text{ for some } j \in (n_{i-1}, n_i]) = 1$$

(this follows from Renyi (1962) or Strawderman and Holmes (1970)). Thus, $P(M_n < b_n \text{ i.o.}) = 1$ when $P(M_i^* < b_{n_i} \text{ i.o.}) = 1$. By the independent version of the Borel-Cantelli lemma, it suffices to verify that $\sum_i P(M_i^* < b_{n_i}) = \infty$. But, by Lemma 1.3,

$$P(M_i^* < b_{n_i}) = \exp\left(-\frac{1+o(1)}{1+\varepsilon}\log\log n_i\right) = (i \log i)^{\frac{1+o(1)}{1+\varepsilon}},$$

and this is not summable in $i$. Thus, $\liminf M_n \log\log n \le \frac{1}{6}\pi^2$ almost surely.

For the second half, we take $0 < \varepsilon < 1$, and set $a_n$ equal to $\frac{1}{6}\pi^2(1-\varepsilon)/\log\log n$. Now, $P(M_n < a_n \text{ i.o.}) = 0$ when $P(M_n < a_n) \to 0$ (a consequence of Lemma 1.2) and $\sum_n P(M_n < a_n, M_{n+1} \ge a_{n+1}) < \infty$. We note first that

$$\begin{aligned}
&P(M_n < a_n, M_{n+1} \ge a_{n+1})\\
&\le P(M_n < a_{n+1}, M_{n+1} \ge a_{n+1}) + P(a_{n+1} \le M_n < a_n)\\
&\le P(X_{n+1} = \max(X_1,\ldots,X_{n+1}), M_n < a_{n+1}) + P(a_{n+1} \le M_n < a_n)\\
&\le E\left(e^{-Z_{n+1}} I_{[Z_{n+1} \le (1-\delta)\log n]}\right) + e^{-(1-\delta)\log n} P(M_n < a_{n+1})\\
&\qquad + P(a_{n+1} \le M_n < a_n)\\
&= I + II + III,
\end{aligned}$$

where $Z_n = \max(X_1,\ldots,X_n)$, and $\delta = \varepsilon \frac{\log\log n}{\log n}$. We verify that I, II and III are all summable in $n$. The easiest term is II : by Lemma 1.3, we have

$$II = n^{-(1-\delta)}(\log n)^{-\frac{1+o(1)}{1-\varepsilon}} = \frac{(\log n)^{\varepsilon}}{n\,(\log n)^{\frac{1+o(1)}{1-\varepsilon}}} ,$$

which is summable in $n$ in view of $\frac{1}{1-\varepsilon} > 1+\varepsilon$. To bound I, we can replace $Z_{n+1}$ by $Z_n$. Thus,

$$I \le E\left(e^{-Z_n} I_{[Z_n<(1-\delta)\log n]}\right) = \int_0^1 P\left(e^{-Z_n} I_{[Z_n<(1-\delta)\log n]} > t\right)dt$$

$$= \int_0^1 P(Z_n < \min(\log(\tfrac{1}{t}),(1-\delta)\log n))dt$$

$$= n^{-(1-\delta)}P(Z_n<(1-\delta)\log n) + \int_{n^{-(1-\delta)}}^1 P(Z_n < \log(\tfrac{1}{t}))dt.$$

For $a$ $0$, we have $P(Z_n$ $a) = (1-e^{-a})^n$ $\exp(-ne^{-a})$. Using this twice, the right-hand-side of the last chain of inequalities is further bounded from above by

$$n^{-(1-\delta)}e^{-n^{\delta}} + \int_{n^{-(1-\delta)}}^{\infty} e^{-nt}dt = e^{-n^{\delta}}(n^{-(1-\delta)} + \tfrac{1}{n})$$

$$\sim e^{-n^{\delta}}n^{-(1-\delta)} = \frac{(\log n)^{\varepsilon}}{n\,e^{(\log n)^{\varepsilon}}} ,$$

which is summable in $n$. To handle III, we have the identity

$$P(M_n \in [a_{n+1},a_n]) = \prod_{i=1}^{n}\left(1-e^{-ia_n}\right) - \prod_{i=1}^{n}\left(1-e^{-ia_{n+1}}\right)$$

$$= \prod_{i=1}^{n}\left(1-e^{-ia_n}\right)\left(1-\prod_{i=1}^{n}\left(\frac{1-e^{-ia_{n+1}}}{1-e^{-ia_n}}\right)\right).$$

The i-th term in the second product of the last expression is

$$\left(1-e^{-ia_{n+1}}\right)\left(1-e^{-ia_{n+1}}e^{-i\Delta}\right) \quad (\text{where } \Delta = a_n - a_{n+1})$$

$$\geq \left(1-e^{-ia_{n+1}}\right)\left(1-e^{-ia_{n+1}} + i\Delta e^{-ia_{n+1}}\right) \geq 1 - i\Delta \frac{e^{-ia_{n+1}}}{1-e^{-ia_{n+1}}} \; .$$

Using the obvious inequality $\prod_{i=1}^{n} (1-u_i) \geq 1-\sum_{i=1}^{n} u_i$ for $u_i \geq 0$ , we obtain

$$P(M_n \in [a_{n+1}, a_n]) \leq \prod_{i=1}^{n} \left(1-e^{-ia_n}\right) \cdot (a_n - a_{n+1}) \cdot \sum_{i=1}^{n} \frac{ie^{-ia_{n+1}}}{1-e^{-ia_{n+1}}}$$

$= \text{IV. V. VI.}$

By Lemma 1.3, $\text{IV} = (\log n)^{-\frac{1+o(1)}{1-\varepsilon}}$ . Also,

$$V = \frac{1}{6}\pi^2(1-\varepsilon)\left(\frac{1}{\log\log n} - \frac{1}{\log\log(n+1)}\right)$$

$$\sim \frac{1}{6}\pi^2(1-\varepsilon)\frac{1}{n \log n\, (\log\log n)^2} \; .$$

Finally,

$$\text{VI} \leq {1-e^{-a_{n+1}}}^{-1} \sum_{i=1}^{\infty} ie^{-ia_{n+1}} \leq {1-e^{-a_{n+1}}}^{-3} \sim a_n^{-3}$$

$$\sim {\frac{6 \log\log n}{\pi^2 (1-\varepsilon)}}^{3} \; .$$

Clearly, IV. V. VI does not exceed a constant times

$$\frac{\log\log n}{n\, (\log n)^{1+\frac{1+o(1)}{1-\varepsilon}}}$$

which is summable in $n$. This conclude the proof.

## REFERENCES

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

[ 2] P. Buchner, *Bemerkungen zur Stirlingschen Formel*, Elem. Math. *6* (1951), 8-11.

[ 3] K. L. Chung, P. Erdös, *On the application of the Borel-Cantelli lemma*, Trans. Am. Math. Soc. *72* (1952), 179-186.

[ 4] P. Deheuvels, *Strong limiting bounds for maximal uniform spacings*, Ann. Probability *10* to appear, 1982.

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

[ 6] L. Devroye, *A log log law for maximal uniform spacings*, Ann. Probability *10* (1982), 863-868.

[ 7] I. S. Gradshteyn, I. M. Ryzhik, *Table of integrals, series and products*, Academic Press, New York, 1980.

[ 8] P. Lévy, *Sur la division d'un segment par des points choisis au hasard*, C.R. Acad. Sci. Paris *208* (1939), 147-149.

[ 9] D. S. Mitrinovic, *Analytic Inequalities*, Springer-Verlag, 1970.

[10] R. Pyke, *Spacings*, J. Roy. Stat. Soc. *27* (1965), 395-436.

[11] A. Renyi, *Théorie des elements saillants d'une sutie d'observations*, Coll. Aarhus, (1962), pp. 104-115.

[12] W. E. Strawderman, P. T. Holmes, *On the law of the iterated logarithm for inter-record times*, Journal of Applied Probability *7* (1970), pp. 432-439.

[13] P. V. Sukhatme, *Tests of significance for samples of the chi-square population with two degrees of freedom*, Ann. Eugen. London *8* (1937), 52-56.

[14] W. A. Whitworth, *Choice and chance*, Cambridge University Press, 1897.

School of Computer Science
McGill University
805 Sherbrooke Street West
Montreal, Canada H3A 2K6

*Received June 27, 1983.*