Statistics & Probability Letters 4 (1986) 183–186
North-Holland

June 1986

THE LIMIT BEHAVIOR OF AN INTERVAL SPLITTING SCHEME

Luc DEVROYE

School of Computer Science, McGill University, Montreal, Canada

Gerard LETAC

Département de Mathématiques, Université Paul Sabatier, Toulouse, France

Vanamamalai SESHADRI

Department of Mathematics and Statistics, McGill University, Montreal, Canada

Received December 1985
Revised February 1986

Abstract: We split [0,1] in a uniform manner, take the largest of the two intervals thus obtained, split this interval again uniformly, and continue in this fashion ad infinitum. We show that the extremes of this interval converge almost surely to a beta (2,2) random variable.

Keywords: spacings, uniform distribution, random processes, strong convergence, beta distribution, limit laws.

1. Introduction

An interval is split by generating a uniformly distributed random variable on that interval. Assume that after a split, we decide to split the larger of the two subintervals again. We start with [0,1], and continue this splitting scheme at infinitum. The asymptotic properties of the interval are studied. The main result of this paper is that the distribution of the eventual location of the extremes of the interval is beta (2,2) distributed. In other words, it is distributed as the median of three iid uniform [0,1] random variables.

The interval splitting process should not be confused with Kakutani's interval splitting procedure (Kakutani, 1975; Lootgieter, 1977; Pyke, 1980), in which a split is applied at every step to the largest subinterval (spacing) obtained thus far. In other words, after n splits, Kakutani splits the largest of the $n+1$ intervals induced by these splits.

Also, there is almost no connection with the spacings obtained in an iid sample of uniform [0,1] random variables on [0,1] (see Pyke (1965, 1972) or Devroye (1981) for surveys of properties of these spacings). In 1941, Kolmogorov introduced his rock-crushing process. A rock of size one is split. Each of the two rocks thus obtained is split. Thus, after k rounds of splitting, one obtains 2^k rocks. The distribution of the sizes of these rocks is studied by Fillippov (1961) and others. See also Athreya and Ney (1972). Our scheme is concerned with only one of these 2^k rocks. But aside from the size of the rock (which reduces to a trivial exercise), we are also interested in a location problem.

We will employ the following notation:

$$[A_0, B_0] = [0,1],$$

$$[A_{n+1}, B_{n+1}] = \begin{cases} [X_n, B_n] & \text{if } X_n \leqslant \frac{1}{2}(A_n + B_n), \\ [A_n, X_n] & \text{if } X_n > \frac{1}{2}(A_n + B_n), \end{cases}$$

$(n \geqslant 0)$.

Research of the authors was sponsored by NSERC and FCAC grants.

Here $X_n = A_n + (B_n - A_n)U_n$ and $U_0, U_1, U_2, \ldots$ is a sequence of iid uniform [0,1] random variables. In the first short subsection, we deal with the asymptotic size of the interval. In the next subsection, the limit distribution of the eventual location is obtained.

2. Size of the interval

Lemma 1. *$B_n - A_n$ is distributed as*

$$\Pi_{1 \leqslant i \leqslant n}((1 + U_i)/2)$$

where the U_i's are iid uniform [0,1] *random variables. In particular,*

A. $E(B_n - A_n) = (\frac{3}{4})^n$,

B. $B_n - A_n \to 0$ *almost surely as* $n \to \infty$.

C. $(e/2)^{\sqrt{n/(1-2\log^2(2))}}(B_n - A_n)^{1/\sqrt{n(1-2\log^2(2))}}$ *tends in distribution to the standard lognormal law. In other words we have that* $(\log(B_n - A_n) - n\log(2/e))/\sqrt{n(1 - 2\log^2(2))}$ *tends in distribution to a normal* (0,1) *random variable.*

Proof. The first statement follows from the fact that the interval selected after one split of [0,1] is distributed as $\max(U, 1 - U)$ where U is uniformly distributed on [0,1], and the observation that this maximum is in turn distributed as $(1 + U)/2$. Statement A follows immediately from this. Statement B follows from statement A and the Borel–Cantelli lemma. A small computation shows that $\log((1 + U)/2)$ has mean $\log(2/e)$, second moment $2\log(e/2) - \log^2(2)$ and variance $1 - 2\log^2(2)$. The central limit theorem gives us C. □

3. The asymptotic location

In this section, we look at the limit law of A_n, B_n and X_n. It is easy to see that such a limit law exists:

Theorem 1. Existence of a limit law. *There exists a random variable W such that*

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} X_n = \lim_{n\to\infty} B_n = W$$

almost surely.

Proof. Since A_n and B_n are both monotone sequences, their almost sure limits exist. In view of Lemma 1, They must be equal. □

Our problem is solved if we can identify the law of W. It should be clear that W is symmetric about $\frac{1}{2}$. Furthermore, W satisfies the following distributional equality:

Lemma 2. The fundamental distributional equation. *Let (W, U) be independent random variables, where W is the random variable of Theorem* 1, *and U is a uniform* [0,1] *random variable. Then, define Z by*

$$Z = \begin{cases} WU, & U > \frac{1}{2}, \\ 1 - W(1 - U), & U \leqslant \frac{1}{2}. \end{cases}$$

Then Z is again distributed as W.

Proof. Consider U as the location of X_0 for the first split, and observe that conditioned on X_0, Z is the eventual location of the interval. But obviously, Z is distributed as W, by construction of the processes. □

In Theorem 1, we observed that there is one limit random variable W. This random variable satisfies the distributional equality of Lemma 2. The distributional property of Lemma 2 is satisfied for the beta (2,2) density:

Theorem 2. *Let W be a random variable with the beta* (2,2) *density. Then W satisfies the distributional equality of Lemma* 2.

Proof. The beta (2,2) density is $6x(1 - x)$, $0 \leqslant x \leqslant 1$. Let W have this density, and let Z be defined by

$$Z = \begin{cases} WU, & U > \frac{1}{2}, \\ 1 - W(1 - U), & U \leqslant \frac{1}{2}. \end{cases}$$

Here U is uniformly distributed on [0,1]. Then simple computations show that the density of Z is

$$6\int_{\max(z,\frac{1}{2})}^{1} \frac{z}{u}\left(1 - \frac{z}{u}\right)\frac{1}{u}\,du$$
$$+6\int_{\max(1-z,\frac{1}{2})}^{1} \frac{1-z}{u}\left(1 - \frac{1-z}{u}\right)\frac{1}{u}\,du$$

$$= 6z - 9z^2 - 6(1-z)\left(1 - \frac{1}{1-z}\right) + 3(1-z)^2\left(1 - \frac{1}{(1-z)^2}\right) \quad (z < \tfrac{1}{2})$$

$$= 6z - 9z^2 + 6z - 3(2z - z^2) = 6z - 6z^2.$$

This concludes the proof of Theorem 2. □

The fact that the distributional equality is satisfied for the beta (2,2) density does not imply that it cannot be satisfied for other distributions as well. We are done if we can show that the distributional equality has a unique solution.

Theorem 3. Main result. *W is beta (2,2) distributed.*

Proof. Lemma 2 provides us with a relationship for the moments of any distribution for which the distributional property is satisfied. As we will see below, any such distribution must share the same moments $\mu_1, \mu_2, \ldots$, so that by the fact that an infinite moment sequence uniquely determines a distribution when it has compact support, we conclude that only one distribution can satisfy the distributional property. But in view of Theorem 2, this then has to have beta (2,2) density. Since W of Theorem 1 also satisfies the distributional property (Lemma 2), W has beta (2,2) density.

To see how the moments are uniquely determined by our distributional equality, we use μ_r to denote $E(W^r)$, where r is a nonnegative integer. Then

$$\begin{aligned}\mu_r &= E\left(W^r U^r I_{U>1/2}\right) + E\left(I_{U\leqslant 1/2}(1 - W(1-U))^r\right) \\ &= E(W^r)E\left(U^r I_{U>1/2}\right) + \sum_{j=0}^{r}\binom{r}{j}(-1)^j E(W^j)E\left((1-U)^j I_{u\leqslant 1/2}\right) \\ &= \mu_r \frac{1}{r+1}\left(1 - 2^{-(r+1)}\right) + \sum_{j=0}^{r}\binom{r}{j}(-1)^j \mu_j \frac{1}{j+1}\left(1 - 2^{-(j+1)}\right).\end{aligned}$$

Thus,

$$\mu_r = \frac{\sum_{j=0}^{r-1}\binom{r}{j}(-1)^j\mu_j\frac{1}{j+1}\left(1 - 2^{-(j+1)}\right)}{1 - 2I_{r\,\text{odd}}\frac{1}{r+1}\left(1 - 2^{-(r+1)}\right)}$$

$(r = 1, 2, 3, \ldots)$.

This concludes the proof of Theorem 3. □

It is perhaps interesting to note that the limit law for our interval splitting process is the same as that of the median of three iid uniform [0,1] random variables. Another by-product is that the median of three iid uniform [0,1] random variables is distributed as the median of three uniform $[A, 'B]$ random variables where $[A, B]$ is the largest of $[0, U]$, $[U, 1]$, and U is a uniform [0,1] random variable.

4. Other splitting schemes

The splitting scheme can be generalized by taking the largest spacing with probability p and the smallest spacing with probability $1-p$. Let W be the asymptotic location (which exists with probability one). Let W, U, B be independent random variables where U is uniform [0,1], B is Bernoulli (p), and W is our asymptotic location. Then the random variable Z defined by

$$Z = \begin{cases} WU, & U > \tfrac{1}{2},\ Z = 1 \text{ or } U \leqslant \tfrac{1}{2}, \\ & Z = 0, \\ 1 - W(1-U), & \text{otherwise}, \end{cases}$$

is again distributed as W.

Theorem 4. *In the general interval splitting scheme, W is beta distributed if and only if $p = 1$ or $p = \frac{1}{2}$. When $p = 1$, it is beta (2,2), and when $p = \frac{1}{2}$ it is beta $(\frac{1}{2}, \frac{1}{2})$* (arcsin).

Proof. The uniqueness of the solution of the distributed equation is easily established (see e.g. the proof of Theorem 3). Also, in view of the fact that W is distributed as a uniform scale mixture, W has a density, which we shall call f. Furthermore, by symmetry, $f(1-w) = f(w)$ for all $w \in [0,1]$. For $w \in [0, \frac{1}{2}]$, we have

$$f(w) = (1-p)\int_{2w}^{1} \frac{f(v)}{v}\,\mathrm{d}v + p\left[\int_{1-w}^{1} \frac{f(v)}{v}\,\mathrm{d}v + \int_{w}^{2w} \frac{f(v)}{v}\,\mathrm{d}v\right],$$
$$0 \leqslant w \leqslant \tfrac{1}{2}.$$

if we replace $f(w)$ by $(w(1-w))^a$ in the integral equation, where $a > -1$ is a constant. Then the following relation is obtained:

$$a(w(1-w))^{a-1}(1-2w) \equiv (4p-2)\frac{(2w(1-2))^a}{2w} + p(w(1-w))^a\left(\frac{1}{1-w} - \frac{1}{w}\right).$$

This should be an identity in w. This is the case when $p = 1$, $a = 1$ and $p = \frac{1}{2}$, $a = -\frac{1}{2}$. To see that there are no other solutions, rewrite the required identity as follows:

$$(a+p)(1-w)^{a-1} \equiv 2^{a-1}(1-2w)^{a-1}(4p-2).$$

When a is not equal to 1, we need to require that $a + p = 4p - 2 = 0$. When $a = 1$ it is necessary that $a + p = 4p - 2$ (i.e., $p = 1$). Thus there are no other beta solutions. □

The integral equation in the proof of Theorem 4 can be used to determine the shape of the density f in the general case.

Acknowledgement

We thank Jean-Pierre Dion and Luc Lefebvre for their helpful simulations which suggested that the limit distribution might be beta.

References

Athreya, K.B. and P.E. Ney (1972), *Branching Processes* (Springer Verlag, Berlin).

Devroye, L. (1981), Laws of the iterated logarithm for order statistics of uniform spacings, *Annals of Probability* **9**, 860–867.

Filippov, A.F. (1961), On the distribution of the sizes of particles which undergo splitting. *Theory of Probability and its Applications* **6**, 275–294.

Kakutani, S. (1975), A problem in equidistribution, *Lecture Notes in Mathematics* **541**, 369–376 (Springer-Verlag, Berlin).

Kolmogorov, A.N. (1941), Uber das logarithmisch normale Vertelungsgesetz der Dimensionen der Teilchen bei Zerstuckelung, *Doklady Akad. Nauk SSSR* **31**, 99–101.

Lootgieter, J.C. (1977), Sur la repartition des suites de Kakutani, *Comptes Rendus de l'Academie des Sciences de Paris* **285A**, 403–406.

Pyke, R. (1965), Spacing, *Journal of the Royal Statistical Society Series B* **7**, 395–445.

Pyke, R. (1972) Spacings revisited, *Proceedings of the Sixth Berkeley Symposium* **1**, 417–427.

Pyke, R. (1980), The asymptotic behavior of spacings under Kakutani's model for interval subdivision, *Annals of Probability* **8**, 157–163.