

M. F. NEUTS (Tucson, AZ)
J.-M. LI (Adelaide, SA)
C. E. M. PEARCE (Adelaide, SA)

ON AN INTERVAL-PARTITIONING SCHEME

Abstract. In a recent paper, Neuts, Rauschenberg and Li [10] examined, by computer experimentation, four different procedures to randomly partition the interval $[0, 1]$ into m intervals. The present paper presents some new theoretical results on one of the partitioning schemes. That scheme is called *Random Interval (RI)*; it starts with a first random point in $[0, 1]$ and places the k th point at random in a subinterval randomly picked from the current k subintervals ($1 < k < m$).

1. Introduction. In a recent paper, Neuts, Rauschenberg and Li [10] examined, by computer experimentation, how different methods of randomly partitioning the interval $[0, 1]$ into m intervals can be identified from the resulting fragments. The four partitioning schemes treated in [10] are as follows.

1. **Random Partition (RP):** This is the familiar method where $m - 1$ points are chosen uniformly and independently on $[0, 1]$. An extensive literature review is given by Pyke [11]. See also, for example, Darling [3], Karlin and Taylor [6, pp. 100–107] and Mountford and Port [9].

2. **Longest Interval First (LI):** A first point is chosen at random in $[0, 1]$. For $k = 2, \dots, m - 1$, the k th point is uniformly distributed in the longest of the current k subintervals.

3. **Random Interval (RI):** Starting with a first random point in $[0, 1]$, to determine the k th point, $2 \leq k \leq m - 1$, one of the current k subintervals is chosen at random and a point is selected at random in it.

4. **Modified Random Interval (MRI):** Starting with a first random point in $[0, 1]$, the k th point, $2 \leq k \leq m - 1$, is placed at random in an interval

1991 *Mathematics Subject Classification*: Primary 60D05.

Key words and phrases: random partitioning, spacings.

chosen with equal probability from among those of the k current intervals that are at least of length $1/k$.

The procedures RP and LI have been the subject of considerable study. For RP, the Glivenko–Cantelli Theorem (see, e.g., Laha and Rohatgi [7, p. 114]) ensures that the empirical distribution of the subdivision points converges almost surely to the uniform distribution. Intuitively one expects convergence to uniformity but at an even faster rate if the longest interval is chosen at each stage, and Kakutani [5] conjectured that for LI the empirical distribution $F_m(x)$ of the m points is asymptotically uniform with probability one. This is rather more difficult to prove than for RP as the division process is more complicated. The conjecture was proved to be true by Lootgieter [8] and van Zwet [13]. See also Adler and Flatto [2], Slud [12] and Gutmann [4] for some related results.

In this paper we derive some results for RI. First, we obtain formulae for the mean length of each interval, cross-moment formulae of the first and last intervals and the distribution of the length of the first interval. Then a convergence result is derived.

2. Formulae for subintervals in RI. Denote by $Y_i(m)$ ($1 \leq i \leq m$) the length of the i th subinterval at that stage of the fragmentation procedure when there are $m \geq 2$ subintervals. We write $y_i(m) := E[Y_i(m)]$. To begin, we remark that by symmetry of the construction the joint density of $Y_1(m), \dots, Y_m(m)$ is unchanged if the order of $Y_1(m), \dots, Y_m(m)$ is reversed. In particular, the marginal densities for $Y_i(m)$ and $Y_{m-i+1}(m)$ are identical. Our basic result is as follows.

THEOREM 2.1. *For the procedure RI, the expected values $y_i(m)$ are*

$$(1) \quad y_1(m) = y_m(m) = \frac{1}{2^{2m-3}} \binom{2m-3}{m-1},$$

$$(2) \quad y_i(m) = \frac{1}{2^{2m-4}} \binom{2i-3}{i-1} \binom{2m-2i-1}{m-i} \quad \text{for } 2 \leq i \leq m-1.$$

PROOF. By considering the possible ways in which the index of the interval to be partitioned can be chosen, we see that the $\{y_i(m)\}$ satisfy the recurrence relation

$$y_i(m+1) = \frac{2i-3}{2m} y_{i-1}(m) + \frac{2m-2i+1}{2m} y_i(m) \quad \text{for } 1 \leq i \leq m+1,$$

with $y_1(1) = 1$, $y_0(m) = y_{m+1}(m) = 0$.

We introduce the generating functions $\Phi_i(u) = \sum_{m=i}^{\infty} y_i(m) u^{m-i}$ for $i \geq 1$. After routine calculations, we obtain the difference-differential equation

$$2(1-u)\Phi'_i(u) = (2i-3)[\Phi_{i-1}(u) - \Phi_i(u)] \quad \text{for } i \geq 2.$$

Initial conditions are obtained by noting that $\Phi_i(0) = y_i(i) = y_1(i)$, and $\Phi_0(u) = 0$. The equation $2(1-u)\Phi_1'(u) = \Phi_1(u)$ leads to

$$\Phi_1(u) = (1-u)^{-1/2},$$

and recursively solving the equations for higher values of i , we obtain

$$\Phi_i(u) = \frac{1}{2^{2i-3}} \binom{2i-3}{i-1} (1-u)^{-1/2} \quad \text{for } i \geq 2.$$

The stated expressions follow by a routine series expansion and are readily seen to satisfy the original recurrence relations and boundary conditions. ■

Next we proceed to find the distribution of $Y_1(m)$ and $Y_m(m)$.

THEOREM 2.2. *For the procedure RI, the probability density $\phi_m(u)$ of the random variable $Y_1(m)$ is given by*

$$(3) \quad \phi_m(u) = \frac{1}{m-1} \left[1 + \sum_{j=1}^{m-2} \frac{\xi_{m-2}(j)}{(j!)^2} (-\log u)^j \right],$$

for $m \geq 3$, $0 < u < 1$, where for $i \geq k$, $\xi_i(k)$ is the coefficient of z^i in the power series of $(1-z)^{-1}[-\log(1-z)]^k$ about the origin.

Proof. The probability densities $\phi_m(u)$ satisfy the recurrence relation

$$(4) \quad \phi_m(u) = \frac{m-2}{m-1} \phi_{m-1}(u) + \frac{1}{m-1} \int_u^1 v^{-1} \phi_{m-1}(v) dv$$

for $m \geq 3$, $0 < u < 1$, and $\phi_2(u) = 1$ for $0 < u < 1$. An induction over m shows that the density is a polynomial of degree $m-2$ in $-\log u$.

The substitution $\phi_m(e^{-u}) = \theta_{m-2}(u)$ enables (4) to be simplified to

$$(5) \quad \theta_m(u) = \frac{m}{m+1} \theta_{m-1}(u) + \frac{1}{m+1} \int_0^u \theta_{m-1}(v) dv$$

for $m \geq 1$, with $\theta_0(u) = 1$ for $0 < u < 1$.

In order to identify $\theta_m(u)$, it is further useful to set

$$\theta_m(u) = (m+1)^{-1} [1 + \psi_m(u)] \quad \text{for } m \geq 1.$$

That leads to the recurrence relation

$$(6) \quad \psi_m(u) = \psi_{m-1}(u) + \frac{u}{m} + \frac{1}{m} \int_0^u \psi_{m-1}(v) dv \quad \text{for } m \geq 1,$$

with $\psi_0(u) = 0$, which defines $\psi_m(\cdot)$ uniquely by recursion. On defining the generating function

$$\Psi(u, z) := \sum_{m=1}^{\infty} \psi_m(u) z^m,$$

we readily obtain the integral equation

$$(1-z)\Psi(u, z) = -u \log(1-z) + \int_0^z dw \int_0^u \Psi(v, w) dv.$$

Successive substitutions suggest a solution of the form

$$\Psi(u, z) = (1-z)^{-1} g[-u \log(1-z)].$$

On substituting for a solution of that form, we find that $g(\cdot)$ must satisfy

$$g(z) = z + \int_0^z \frac{dt}{t} \int_0^t g(y) dy.$$

In fact, this equation has a unique solution which is analytic in a region about the origin. For if there is such a solution, then substituting $g(z) = \sum_{r=0}^{\infty} A_r z^r$ in the integral equation yields $A_0 = 0$, $A_1 = 1$ and $A_{r+1} = (r+1)^{-2} A_r$ for $r \geq 1$, so that

$$g(z) = \sum_{k=1}^{\infty} \frac{z^k}{(k!)^2} = I_0(2z^{1/2}) - 1,$$

where $I_0(\cdot)$ denotes the modified Bessel function of order zero (see, for example, Abramowitz and Stegun [1, p. 375]).

As is immediate from an application of the ratio test, this power series for g has infinite radius of convergence and so represents a function which is analytic in the whole of the complex plane. We readily verify that our solution does indeed satisfy the integral equation for g .

Our solution yields

$$\begin{aligned} \sum_{r=1}^{\infty} z^r \psi_r(u) &= \Psi(u, z) = \sum_{k=1}^{\infty} \frac{u^k}{(k!)^2} \{(1-z)^{-1} [-\log(1-z)]^k\} \\ &= \sum_{k=1}^{\infty} \frac{u^k}{(k!)^2} \sum_{r=k}^{\infty} \xi_r(k) z^r \\ &= \sum_{r=1}^{\infty} z^r \sum_{k=1}^r \frac{u^k}{(k!)^2} \xi_r(k). \end{aligned}$$

The stated expression for $\phi_m(u)$ follows. We can verify that this satisfies (6) with $\psi_0(u) = 0$. As the stated recurrence relation defines $\psi_m(u)$ uniquely, we are done. ■

REMARK. Since for $|z| < 1$,

$$[-\log(1-z)]^k = \sum_{j=k}^{\infty} (-1)^{j+k} \frac{k!}{j!} S_j^{(k)} z^j,$$

where $S_j^{(k)}$ is the Stirling number of the first kind, we see that for $m \geq 3$, $\phi_m(u)$ may also be written as

$$(7) \quad \phi_m(u) = \frac{1}{m-1} \left[1 + \sum_{j=1}^{m-2} \frac{1}{j!} \sum_{\nu=j}^{m-2} \frac{(-1)^{\nu+j}}{\nu!} S_\nu^{(j)} (-\log u)^j \right].$$

COROLLARY 2.3. *The random variable $-\log Y_1(m)$ has the density with Laplace–Stieltjes transform*

$$(8) \quad \theta_m^*(s) = \prod_{r=0}^{m-2} \left(\frac{r}{r+1} + \frac{1}{r+1} \frac{1}{s+1} \right),$$

and $-(2 \log m)^{-1/2} \log[mY_1(m)]$ has a standard normal asymptotic distribution.

PROOF. The density of $-\log Y_1(m)$ is given by $e^{-u}\theta_{m-2}(u)$ on $[0, \infty)$. By evaluating the transform

$$\theta_m^*(s) = \int_0^\infty e^{-su} e^{-u} \theta_{m-2}(u) du$$

from the recurrence relation (5), the stated expression is obtained. The corresponding density is a mixture of the Erlang distributions of order 1 to $m-2$ and scale parameter 1. Its mean and variance are respectively given by

$$\sum_{r=0}^{m-2} \frac{1}{r+1} \quad \text{and} \quad \sum_{r=0}^{m-2} \frac{2r+1}{(r+1)^2},$$

from which it follows that $-(2 \log m)^{-1/2} \log[mY_1(m)]$ asymptotically has a standard normal distribution. ■

The joint density $\phi_m(u_1, u_2)$ of the random variables $Y_1(m)$ and $Y_m(m)$ satisfies the recurrence relation

$$(9) \quad \begin{aligned} \phi_m(u_1, u_2) &= \frac{m-3}{m-1} \phi_{m-1}(u_1, u_2) \\ &+ \frac{1}{m-1} \int_{u_1}^{1-u_2} v^{-1} \phi_{m-1}(v, u_2) dv \\ &+ \frac{1}{m-1} \int_{u_2}^{1-u_1} v^{-1} \phi_{m-1}(u_1, v) dv, \end{aligned}$$

for $m \geq 4$ and $u_1, u_2 \geq 0$, $0 \leq u_1 + u_2 \leq 1$, where

$$\phi_3(u_1, u_2) = \frac{1}{2} \left[\frac{1}{1-u_1} + \frac{1}{1-u_2} \right].$$

It appears rather difficult to extract an expression for $\phi_m(u_1, u_2)$ for general m . However, all moments of that density can be found explicitly. Define $M_{r_1, r_2}(m)$ by

$$M_{r_1, r_2}(m) = E[Y_1^{r_1}(m)Y_m^{r_2}(m)],$$

for $r_1 \geq 0$ and $r_2 \geq 0$. Then we have the following.

THEOREM 2.4. *For the procedure RI, the moment $M_{r_1, r_2}(m)$ is given by*

$$(10) \quad M_{r_1, r_2}(m) = \frac{r_1! r_2!}{(r_1 + r_2 + 1)!} \frac{\Gamma[m - 2 + \gamma(r_1, r_2)]}{(m - 1)! \Gamma[\gamma(r_1, r_2)]} \quad \text{for } m \geq 2,$$

where $\gamma(r_1, r_2) = (r_1 + 1)^{-1} + (r_2 + 1)^{-1}$.

Proof. From equation (9) we obtain

$$\begin{aligned} M_{r_1, r_2}(m) &= \frac{m - 3}{m - 1} M_{r_1, r_2}(m - 1) \\ &+ \frac{1}{m - 1} \int_0^1 du_2 u_2^{r_2} \int_0^{1 - u_2} u_1^{r_1} du_1 \int_{u_1}^{1 - u_2} \phi_{m - 1}(v, u_2) v^{-1} dv \\ &+ \frac{1}{m - 1} \int_0^1 du_1 u_1^{r_1} \int_0^{1 - u_1} u_2^{r_2} du_2 \int_{u_2}^{1 - u_1} \phi_{m - 1}(u_1, v) v^{-1} dv. \end{aligned}$$

The first integral may be rewritten as

$$\int_0^1 du_2 u_2^{r_2} \int_0^{1 - u_2} \phi_{m - 1}(v, u_2) v^{-1} dv \int_0^v u_1^{r_1} du_1 = \frac{1}{r_1 + 1} M_{r_1, r_2}(m - 1),$$

and a corresponding expression holds for the second integral. It follows that

$$M_{r_1, r_2}(m) = \frac{1}{m - 1} \left[m - 3 + \frac{1}{r_1 + 1} + \frac{1}{r_2 + 1} \right] M_{r_1, r_2}(m - 1),$$

with $M_{r_1, r_2}(2) = B(r_1 + 1, r_2 + 1)$, a Beta function. Routine calculations lead to the stated formula. ■

For $r_1 = 1$ and $r_2 = 0$, we find that

$$M_{1, 0}(m) = \frac{\Gamma(m - 1/2)}{\Gamma(1/2) (m - 1)!} = \frac{1}{2^{2m - 3}} \binom{2m - 3}{m - 1},$$

which agrees with the expression in Theorem 2.1. The second moment of $Y_1(m)$ is given by

$$M_{2, 0}(m) = \frac{\Gamma(m - 2/3)}{\Gamma(1/3) (m - 1)!},$$

and the cross-moment $M_{11}(m)$ is found to be

$$M_{11}(m) = \frac{1}{6(m - 1)}.$$

For $m \geq 3$, the joint density of the random variables $-\log Y_1(m)$ and $-\log Y_m(m)$ is given by $\phi_m(e^{-v_1}, e^{-v_2})e^{-v_1-v_2}$ on the region where $v_1 > 0$, $v_2 > 0$, and $e^{-v_1} + e^{-v_2} \leq 1$. By a routine calculation using the recurrence relation (9), we find that its joint Laplace–Stieltjes transform is given by

$$\theta_m^*(s_1, s_2) = M_{s_1, s_2}(m).$$

3. A limit theorem and a conjecture. As $m \rightarrow \infty$, the expected values $y_i(m)$ investigated at the beginning of the previous section have an interesting limiting behavior. To examine it, we form a discrete probability density by placing the masses $y_i(m)$ at the points $i(m+1)^{-1}$. That distribution is symmetric about $1/2$. We also form the absolutely continuous distributions with densities $f(u; m)$ defined by

$$f(u; m) = (m+1)y_i(m) \quad \text{for } \frac{2i-1}{2(m+1)} < u \leq \frac{2i+1}{2(m+1)}$$

for $1 \leq i \leq m$ and zero elsewhere. Then we have the following result.

THEOREM 3.1. *The discrete distribution with density $\{y_i(m)\}$ and the continuous distribution with density $f(u; m)$ converge weakly to the arc-sine law with density $\pi^{-1}[u(1-u)]^{-1/2}$ on $(0, 1)$.*

Proof. Fix x with $0 < x < 1$. Then by using Stirling's formula, we easily verify that $f(x; m)$ converges to $\pi^{-1}[x(1-x)]^{-1/2}$ as $m \rightarrow \infty$. Moreover, that convergence is uniform in x in any interval $[a, b]$ with $0 < a < b < 1$, so that

$$\int_a^b f(u; m) du \rightarrow \int_a^b \pi^{-1}[u(1-u)]^{-1/2} du,$$

which is equivalent to convergence in law. Since $f(u; m)$ is piecewise constant, the first integral can be written as a finite sum. The stated convergence therefore also implies that of the sequence of the discrete distributions. ■

REMARK. The large magnitude of the $Y_1(m)$ and $Y_m(m)$ compared to the lengths of intervals in the middle is an outstanding feature of the intervals generated by the RI procedure. One may easily verify that as $m \rightarrow \infty$, the ratio of $y_1(m)$ to the length of the middle interval tends to infinity as $m^{1/2}$.

Intuitive insight into the reasons for that behavior is gained by wrapping the interval $[0, 1]$ around a circle. When there are m intervals, there is probability $2/m$ that the arc of length $Y_1(m) + Y_m(m)$ about the fixed point corresponding to 0 and 1 is reduced. For successive m , reductions in size of that interval occur according to Bernoulli trials with probabilities $2/m$. The number of such reductions in m steps behaves as $\log m$, which accounts for the greater relative length of that interval.

We now conjecture that the almost sure convergence may be established for the empirical distribution in the RI context, as was done for procedures RP and LI.

CONJECTURE 3.2. *For the procedure RI, the empirical distribution of the points of divisions asymptotically converges to the arc-sine law almost surely.*

Acknowledgements. Research of Marcel F. Neuts and Jian-Min Li was supported in part by NSF Grant Nr. DMI-9306828. One of the authors, Jian-Min Li, now with the Teletraffic Research Centre, the Department of Applied Mathematics at the University of Adelaide, also thanks the support of the Australian Research Council through grant A69602762 since August 1996.

References

- [1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1965.
- [2] R. L. Adler and L. Flatto, *Uniform distribution of Kakutani's interval splitting procedure*, Z. Wahrsch. Verw. Gebiete 38 (1977), 253–259.
- [3] D. A. Darling, *On a class of problems related to the random division of an interval*, Ann. Math. Statist. 24 (1953), 239–253.
- [4] S. Gutmann, *Interval-dividing processes*, Z. Wahrsch. Verw. Gebiete 57 (1981), 339–347.
- [5] S. Kakutani, *A problem of equidistribution on the unit interval $[0, 1]$* , in: Proc. Oberwolfach Conf. on Measure Theory (1975), Lecture Notes in Math. 541, Springer, Berlin, 1976, 369–376.
- [6] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.
- [7] R. G. Laha and V. K. Rohatgi, *Probability Theory*, John Wiley & Sons, New York, 1979.
- [8] J. C. Lootgieter, *Sur la répartition des suites de Kakutani*, C. R. Acad. Sci. Paris 285A (1977), 403–406.
- [9] T. S. Mountford and S. C. Port, *Random splittings of an interval*, J. Appl. Probab. 30 (1993), 131–152.
- [10] M. F. Neuts, D. E. Rauschenberg and J.-M. Li, *How did the cookie crumble? Identifying fragmentation procedures*, Statist. Neerlandica 51 (1997), 238–251.
- [11] R. Pyke, *Spacings*, J. Roy. Statist. Soc. Ser. B 27 (1965), 395–449.
- [12] E. Slud, *Entropy and maximal spacings for random partitions*, Z. Wahrsch. Verw. Gebiete 41 (1978), 341–352.

- [13] W. R. van Zwet, *A proof of Kakutani's conjecture on random subdivision of longest intervals*, Ann. Probab. 6 (1978), 133–137.

Marcel F. Neuts
Department of Systems and
Industrial Engineering
The University of Arizona
Tucson, AZ 85721, U.S.A.
E-mail: marcel@sie.arizona.edu

Jian-Min Li and Charles E. M. Pearce
Department of Applied Mathematics
The University of Adelaide
Adelaide, SA 5005, Australia
E-mail: jli@maths.adelaide.edu.au
cpearce@maths.adelaide.edu.au

Received on 16.3.1999;
revised version on 31.5.1999