The diminishing segment process

Gergely Ambrus [1]
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences,
PO Box 127, 1364 Budapest, Hungary; *e-mail:* `ambrus@renyi.hu`
Péter Kevei [2]
Centro de Investigación en Matemáticas, Jalisco S/N, Valenciana,
Guanajuato, GTO 36240, Mexico; *e-mail:* `kevei@cimat.mx`
Viktor Vígh [3]
Department of Mathematics and Statistics, University of Calgary
2500 University Drive NW Calgary, AB, Canada T2N 1N4
e-mail: `vigvik@gmail.com`

Abstract

Let $\Xi_0 = [-1, 1]$, and define the segments Ξ_n recursively in the following manner: for every $n = 0, 1, \ldots$, let $\Xi_{n+1} = \Xi_n \cap [a_{n+1} - 1, a_{n+1} + 1]$, where the point a_{n+1} is chosen randomly on the segment Ξ_n with uniform distribution. For the radius ρ_n of Ξ_n we prove that $n(\rho_n - 1/2)$ converges in distribution to an exponential law, and we show that the centre of the limiting unit interval has arcsine distribution.

Keywords: Arcsine law; Continuous state space Markov chain; Poisson–Dirichlet law; Intersection of convex discs.

1 Introduction

We consider the following stochastic process. Let $\Xi_0 = [-1, 1]$, and define the segments Ξ_n recursively in the following manner: for every $n = 0, 1, \ldots$, let

$$\Xi_{n+1} = \Xi_n \cap [a_{n+1} - 1, a_{n+1} + 1],$$

where the point a_{n+1} is chosen randomly on the segment Ξ_n with uniform distribution. After n steps, one obtains the segment

$$\Xi_n = [Z_n - \rho_n, Z_n + \rho_n].$$

The (centre, radius) process (Z_n, ρ_n) is a continuous state space Markov chain. The radius sequence (ρ_n) is monotonically decreasing, and it is easy

[1]Supported by the OTKA grants 75016 and 76099.
[2]Supported by the Analysis and Stochastics Research Group of the Hungarian Academy of Sciences.
[3]Supported by the OTKA grant 75016, and by NSERC of Canada.

to see that with probability 1, $\rho_n \to 1/2$. Moreover, (Z_n) is a convergent sequence, assuming values on $[-1/2, 1/2]$. Denote $\lim_{n\to\infty} Z_n$ by Z.

We are interested in the most straightforward questions:

(1) What is the distribution of Z_n and ρ_n for a given n?

(2) What is the asymptotic behaviour of the radius?

(3) What is the limit distribution of the centre?

Our work was motivated by the following problem formulated by Bálint Tóth (Tóth, 2010) some 20 years ago with K being the unit disc of the plane. Let $K = K_0$ be a convex body in the $\mathbb{R}^d$ that contains the origin, and define the process (K_n, p_n), $n \geqslant 1$, similarly to the above construction: let p_{n+1} be a uniform random point in K_n, and set $K_{n+1} = K_n \cap (p_{n+1} + K)$. Clearly, (K_n) is a nested sequence of convex bodies which converge to a non-empty limit object, again a convex body in $\mathbb{R}^d$. What can we say about the distribution of this limit body? When K is the unit disc the limit object is almost surely a convex disc of constant width 1. The present note deals with the 1-dimensional analogue of this problem. Intriguingly, apart from almost trivial results, nothing is known about the related questions even in the plane.

Another direction of generalising the problem treated here is to choose the subsequent centres according to a previously fixed distribution instead of the uniform one. Research on this version is currently in progress.

The paper is organised as follows. In Section 2 we give a recursion for the density function of ρ_n, which allows us to explicitly calculate the expectation for small n's. In Section 3 we show that $n(\rho_n - 1/2)$ converges in distribution to an exponential law, which actually shows the rapidness of the process. Section 4 contains the limit distribution of the centre, while in the last section we derive a somewhat unexpected connection between the process and the Poisson–Dirichlet distribution.

2 First observations

At the $(n+1)$st step, we separate two cases according to the location of a_{n+1}. First, if $a_{n+1} \in [Z_n + \rho_n - 1, Z_n - \rho_n + 1]$, then no change occurs to the segment, and thus $\Xi_{n+1} = \Xi_n$. In the second case, when a_{n+1} is close to one of the endpoints of Ξ_n, the centre moves and the length decreases. Introduce yet two new random processes measuring the change of the location of the centre by

$$\varepsilon_n X_n = Z_n - Z_{n-1}, \ n \geqslant 1,$$

with $\varepsilon_n = \pm 1$ and $X_n \geqslant 0$ (if $Z_n = Z_{n-1}$, then of no consequence, let $\varepsilon_n = 1$). Thus, for $n \geqslant 1$,

$$X_n = \begin{cases} 0 & \text{w.prob. } (1-\rho_{n-1})/\rho_{n-1} \\ \text{uniform on } [0, \rho_{n-1} - 1/2] & \text{w.prob. } (2\rho_{n-1} - 1)/\rho_{n-1}, \end{cases} \tag{1}$$

moreover,

$$\mathbb{P}(\varepsilon_n = 1 | X_n \neq 0) = \frac{1}{2}\,.$$

By definition,

$$Z_n = \sum_{i=1}^{n} \varepsilon_i X_i, \tag{2}$$

and it is easy to see that

$$\rho_n = 1 - \sum_{i=1}^{n} X_i =: 1 - S_n.$$

Thus, with probability 1, $\sum_1^\infty X_i = 1/2$. By an inductive argument, it follows that S_n has a continuous distribution. Denote by $f_n(x)$ the probability density function of S_n. Using the Markov property and (1) it is easy to express f_{n+1} in terms of f_n: for $0 \leqslant x \leqslant 1/2$,

$$\begin{aligned} \mathbb{P}(S_{n+1} > x) &= \int_0^{1/2} \mathbb{P}(S_{n+1} > x | S_n = y) f_n(y)\,\mathrm{d}y \\ &= \mathbb{P}(S_n > x) + \int_0^x f_n(y) \frac{1-2x}{1-y}\,\mathrm{d}y, \end{aligned}$$

whence differentiating

$$f_{n+1}(x) = f_n(x) \frac{x}{1-x} + 2\int_0^x \frac{f_n(y)}{1-y}\,\mathrm{d}y. \tag{3}$$

The first few examples are, for $0 \leqslant x \leqslant 1/2$,

$$\begin{aligned} f_1(x) &= 2 \\ f_2(x) &= \frac{2x}{1-x} - 4\ln(1-x) \\ f_3(x) &= \frac{2x(2-x)}{(1-x)^2} + \frac{4(1-2x)}{1-x}\ln(1-x) + 4(\ln(1-x))^2. \end{aligned}$$

In order to calculate the expectation of S_n (and thus $\mathbb{E}\rho_n$), we consider the Taylor series expansion of f_n about 0:

$$f_n(x) = \sum_{k=0}^{\infty} c_{n,k} x^k.$$

Based on (3), one readily obtains the formula

$$c_{n+1,k} = \frac{k+2}{k} \sum_{j=0}^{k-1} c_{n,j} \,, \tag{4}$$

and hence

$$\mathbb{E}S_n = \int_0^{1/2} x f_n(x)\mathrm{d}x = \sum_{k=0}^{\infty} c_{n,k} \frac{2^{-(k+2)}}{k+2} = \frac{1}{4} \sum_{j=0}^{\infty} c_{n-1,j} \sum_{k=j+1}^{\infty} \frac{2^{-k}}{k}, \tag{5}$$

which may be useful for obtaining a recursive expression involving only the expectations. These formulas also enable us to efficiently compute the expectations for relatively small n's, see the figure below.

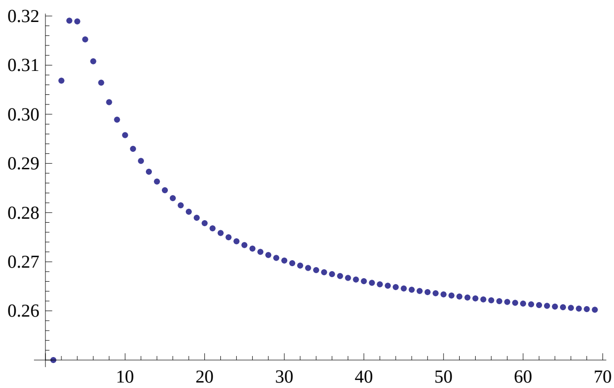

Figure 1: The first 70 values of $n(1/2 - \mathbb{E}S_n)$

3 Asymptotics of the length

We determine the asymptotic behavior of ρ_n, the radius of the nth segment.

Theorem 1. *We have the distributional convergence*

$$n\left(\rho_n - \frac{1}{2}\right) \xrightarrow{\mathcal{D}} \mathrm{Exp}(4).$$

Accordingly, for any $k \geqslant 1$,

$$\lim_{n\to\infty} \mathbb{E}n^k \left(\rho_n - \frac{1}{2}\right)^k = \frac{k!}{4^k} \,.$$

Proof. Since $\rho_n = 1 - S_n$, it is equivalent to prove the corresponding limit theorems for $1/2 - S_n$.

It follows from (1) that

$$S_{n+1} = \begin{cases} S_n & \text{w.prob. } S_n/(1-S_n) \\ \text{uniform on } [S_n, 1/2] & \text{w.prob. } (1-2S_n)/(1-S_n). \end{cases}$$

Observe that given S_n, S_{n+1} has the same distribution as $\max\{S_n, U\}$, where U is a uniform random variable on $[S_n/2, 1/2]$, independent of S_n. Let $0 \leqslant \alpha \leqslant 1/2$, and $U_1^\alpha, U_2^\alpha, \ldots$ be independent, uniform random variables on $[1/4 - \alpha/2, 1/2]$. Define $M_n^\alpha = \max_{1 \leqslant i \leqslant n} U_i^\alpha$. It is well-known in extreme value theory (Billingsley, 1995, p.192), and easy to check that

$$n\left(\frac{1}{2} - M_n^\alpha\right) \xrightarrow{\mathcal{D}} \operatorname{Exp}(4/(1+2\alpha)). \tag{6}$$

We say that X *stochastically dominates* (or just dominates) Y if $\mathbb{P}(X > x) \geqslant \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$.

To obtain a lower estimate, we use that $S_n \leqslant 1/2$, and thus an easy induction argument shows that M_n^0 dominates S_n. From this and (6) we obtain

$$\mathbb{P}\left(n\left(\frac{1}{2} - S_n\right) > x\right) \geqslant \mathbb{P}\left(n\left(\frac{1}{2} - M_n^0\right) > x\right) \to \mathrm{e}^{-4x},$$

i.e.

$$\liminf_{n\to\infty} \mathbb{P}\left(n\left(\frac{1}{2} - S_n\right) > x\right) \geqslant \mathrm{e}^{-4x}, \quad x \geqslant 0.$$

The almost sure convergence $S_n \to 1/2$ heuristically means that with overwhelming probability, $1/2 - S_n/2 \approx 1/4$, hence for n sufficiently large, S_n behaves approximately like M_n^0 does. This heuristic idea is made precise as follows. Fix the small positive numbers $\beta > 0$ and $\varepsilon > 0$. Since $S_n \to 1/2$ a.s., for n sufficiently large, $\mathbb{P}(S_{\beta n} < 1/2 - \varepsilon) \leqslant \varepsilon$. Moreover, if $S_{\beta n} \geqslant 1/2 - \varepsilon$, then S_n is minored by $M_{n-\beta n}^\varepsilon$. This can be shown again by induction. Thus, by (6),

$$\begin{aligned} \mathbb{P}\left(n\left(\frac{1}{2} - S_n\right) > x\right) &= \mathbb{P}\left(S_n < \frac{1}{2} - \frac{x}{n}\right) \\ &\leqslant \mathbb{P}\left(S_n < \frac{1}{2} - \frac{x}{n} \,\middle|\, S_{\beta n} \geqslant \frac{1}{2} - \varepsilon\right) + \varepsilon \\ &\leqslant \mathbb{P}\left(M_{(1-\beta)n}^\varepsilon < \frac{1}{2} - \frac{x}{n}\right) + \varepsilon \\ &\to \mathrm{e}^{-\frac{4x(1-\beta)}{1+2\varepsilon}} + \varepsilon\,, \end{aligned}$$

i.e. for $x \geqslant 0$

$$\limsup_{n\to\infty} \mathbb{P}\left(n\left(\frac{1}{2}-S_n\right)>x\right) \leqslant \mathrm{e}^{-\frac{4x(1-\beta)}{1+2\varepsilon}}+\varepsilon.$$

Since ε and β are arbitrary, this gives the distributional convergence.

To prove the convergence of the moments, it is enough to show that for any $k \geqslant 1$, the sequence $\{n^k\left(\frac{1}{2}-S_n\right)^k\}_{n=1}^{\infty}$ is uniformly integrable (Billingsley, 1995, Theorem 25.12). The fact that S_n dominates $M_n^{1/2}$ readily implies

$$\mathbb{P}\left(n\left(\frac{1}{2}-S_n\right)>x\right) \leqslant \mathrm{e}^{-2x},$$

which shows uniform integrability. □

In particular, $\mathbb{E}\rho_n = 1/2 - 1/(4n) + o(n^{-1})$ and $\mathrm{Var}(\rho_n) \sim 1/(16n^2)$. The question of obtaining an exact or asymptotic formula for $\mathbb{E}S_n$ using only the recursive relations (4) and (5) remains open.

4 Limit distribution

In this section, we determine the limit behaviour of Z_n. Let $F(x)$ denote the cumulative distribution function of Z; based on the definition of Z, it follows that $F(x) = 0$ if $x \leqslant -1/2$ and $F(x) = 1$ if $x \geqslant 1/2$.

Theorem 2. *The distribution of Z is a translated arcsine law: for $-1/2 \leqslant x \leqslant 1/2$,*

$$F(x) = \frac{2}{\pi}\arcsin\sqrt{x+1/2}\,.$$

Proof. By (2), we have to determine the limit distribution of $\sum \varepsilon_i X_i$. Clearly, this is not affected by the steps where $X_i = 0$. We introduce the thinned process $(\widetilde{Z}_n, \tilde{\rho}_n)$ as follows: for $n \geqslant 1$, let ξ_n be independent Bernoulli(1/2) random variables, and $\widetilde{X}_n$ be a uniform random variable on $[0, 1/2 - \sum_1^{n-1}\widetilde{X}_i]$. The centre of the segment after the nth step of the thinned process is given by $\widetilde{Z}_n = \sum_1^n \xi_i \widetilde{X}_i$, and the radius is $\tilde{\rho}_n = 1 - \sum_1^n \widetilde{X}_i$. Plainly,

$$Z \stackrel{\mathcal{D}}{=} \sum_{i=1}^{\infty} \xi_i \widetilde{X}_i.$$

Introduce $r_n = 2\tilde{\rho}_n - 1$. If $U_1, U_2, \ldots$ are i.i.d. Uniform(0, 1) random variables, then setting $(\widetilde{Z}_0, r_0) = (0, 1)$,

$$
\begin{aligned}
\widetilde{Z}_{n+1} &= \widetilde{Z}_n + \frac{1}{2}\,\xi_{n+1}(1-U_{n+1})r_n\,, \\
r_{n+1} &= U_{n+1}r_n\,.
\end{aligned}
\tag{7}
$$

Notice that after choosing $(\widetilde{Z}_1, r_1)$, the process $(\widetilde{Z}_2, r_2), (\widetilde{Z}_3, r_3), \ldots$ is a scaled and translated copy of the original one, which implies the distributional equation

$$
Z \overset{\mathcal{D}}{=} r_1\, Z' + \widetilde{Z}_1, \tag{8}
$$

where Z' is independent from $(\widetilde{Z}_1, r_1)$, and has the same distribution as Z. Thus, for every $x \in [-1/2, 1/2]$,

$$
\begin{aligned}
F(x) &= \int_0^{1/2} F\left(\frac{x-y}{1-2y}\right)\,\mathrm{d}y + \int_0^{1/2} F\left(\frac{x+y}{1-2y}\right)\,\mathrm{d}y \\
&= (1-2x)\int_{-\infty}^{x} \frac{F(z)}{(1-2z)^2}\,\mathrm{d}z + (1+2x)\int_x^{\infty} \frac{F(z)}{(1+2z)^2}\,\mathrm{d}z \\
&= (1-2x)\int_{-1/2}^{x} \frac{F(z)}{(1-2z)^2}\,\mathrm{d}z + (1+2x)\int_x^{1/2} \frac{F(z)}{(1+2z)^2}\,\mathrm{d}z + \frac{1+2x}{4}\,.
\end{aligned}
$$

This also shows that F is continuously differentiable, and by differentiating we arrive at

$$
F'(x) = -2\int_{-1/2}^{x} \frac{F(z)}{(1-2z)^2}\,\mathrm{d}z + 2\int_x^{1/2} \frac{F(z)}{(1+2z)^2}\,\mathrm{d}z + \frac{4xF(x)}{1-4x^2} + \frac{1}{2}\,.
$$

Once again, we derive that F is twice differentiable (being the reason for starting with the distribution function rather than the density function), whence

$$
F''(x) = F'(x)\frac{4x}{1-4x^2}\,.
$$

Taking into account that $F'(x)$ is a density function on $[-1/2, 1/2]$ yields that the solution is

$$
F'(x) = \frac{1}{\pi\sqrt{(1/2+x)(1/2-x)}}\,,
$$

the desired density function. □

5 Further remarks

Setting $v_i = U_1U_2\ldots U_i(1-U_{i+1})$, (7) implies that $\widetilde{Z}_n = 1/2\sum_0^{n-1} v_i\xi_{i+1}$, hence the limit Z has the infinite series representation

$$
Z = \frac{1}{2}\sum_{i=0}^{\infty} v_i\xi_{i+1}. \tag{9}
$$

It is easy to check that $\sum_0^\infty v_i = 1$ almost surely, thus $v := (v_0, v_1, \ldots) \in \Delta$, where

$$\Delta = \left\{ x = (x_0, x_1, \ldots) : \sum_{i=0}^{\infty} x_i = 1, x_i \geqslant 0, i = 0, 1, \ldots \right\}$$

is the infinite dimensional simplex. The construction of the random vector v implies that it has the so-called GEM (Griffiths–Engen–McCloskey) distribution with parameter 1 (see the residual allocation model in Bertoin, 2006, p.89). This distribution appears in various contexts, such as prime factorisation of a random integer (Hirth, 1997); and in particular, the decreasing reordering of the GEM distribution is the so-called Poisson–Dirichlet distribution, which is one of the most important distributions in fragmentation theory, see Bertoin, 2006.

Using this terminology and (9), Theorem 2 can be reformulated as: if $v = (v_0, v_1, \ldots)$ is a GEM distributed random vector, and $\xi, \xi_1, \xi_2, \ldots$ is an iid sequence of Bernoulli random variables such that $\mathbb{P}(\xi = 1) = 1/2 = \mathbb{P}(\xi = -1)$, which is independent of v, then $\sum_{i=0}^\infty v_i \xi_{i+1}$ has arcsine distribution. This theorem was first proved by Donnelly and Tavaré (1987), using the construction of the Poisson–Dirichlet distribution by means of an inhomogeneous Poisson process. Later Hirth (1997) also gave a proof by using the method of moments. As far as we know, our proof, solving an integral equation based on the distributional equality (8), is new.

Acknowledgements. Our thanks are due to Imre Bárány for introducing the problem to us and for illuminating discussions, and to Juan Carlos Pardo for bringing the connection with the Poisson–Dirichlet law to our attention. The authors are also grateful to the unknown referee for a number of comments and suggestions that improved the paper. While carrying out the research, the first named author held a pleasant visiting scholarship at the Isaac Newton Institute for Mathematical Sciences, University of Cambridge.

References

[1] Bertoin, J. (2006) *Random Fragmentation and Coagulation Processes*, Cambridge University Press.

[2] Billingsley, P. (1995) *Probability and Measure*, Wiley–Interscience Publication, New York.

[3] Donnelly, P. and Tavaré, S. (1987) The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.* **25**, 381–391.

[4] Hirth, U.M. (1997) Probabilistic number theory, the GEM/Poisson–Dirichlet distribution and the arc-sine law. *Combinatorics, Probability and Computing* 6, 57–77.

[5] Tóth, B., Private communication, 2010.