Simulation Analysis of Statistical Properties of Random Polygonal Lines Iteratively Generate on the Plane

Rui Santos
School of Technology and Management, Polytechnic Institute of Leiria
Morro do Lena – Alto do Vieiro, Apartado 4163, 2411-901 Leiria, Portugal
CEAUL — Center of Statistics and Applications of University of Lisbon
E-mail: rui.santos@ipleiria.pt

Abstract. *Using renormalization, we compute a scale factor such that the distance of the endpoints of a Pacheco d'Amorim random polygonal line in the plane is a non-degenerate random variable. We also analyze more general random polygonal lines and the corresponding limit curves, folded at random points of the segments at each iterative step of its construction, instead of the middle point.*

Keywords. Non-degenerate asymptotic laws, renormalization, Hausdorff dimension, simulation, plane random curves.

1. Introduction

At the dawn of modern Probability Theory, J. Bertrand, J. M. Keynes and many others looked at continuous probability with suspicion, and presented intriguing paradoxes, either connected to conditioning on events of zero probability, or due to the lack of important tools such as the concept of random variable and the manipulation of functions of random variables.

The early steps of Borel construction of continuous probability originated some misconceptions, and some results published at the time seem nowadays foolish. The results we present give a modern and correct solution of an ill-solved problem we came across in Pacheco d'Amorim's *Probability Calculus* [3].

According to Pacheco d'Amorim, the distance between the endpoints of a fixed length random curve in the plane is almost surely zero, a result that doesn't stand empirical confrontation. The mistake in Pacheco d'Amorim's conclusion is that, starting from a straight line L_0 of length ℓ_0, he considers an iterative procedure of breaking each side at the middle point and randomly folding the two resulting segments in the plane, considering that the limit is a random curve still of length ℓ_0. In fact, the limit would be a point, as reasoning clearly shows and the simulation we present corroborates.

Observe also that from self-similarity arguments [2], it is clear that the above procedure generates a fractal of Hausdorff dimension 1 in $\mathbb{R}^2$, and that the problem arises from the need of renormalizing so that the limit length will be neither zero nor infinity, an idea that we shall formally present in Section 3.

Renormalization has been used in Probability to define stable random variables (either in the additive or in the extreme value scheme), in Differential Geometry, via the Ricci-Hamilton flow [1] to tackle Poincaré's conjecture, in Physics to work out problems of phase transition [7]; in Time Series it is related to the Hurst exponent and long range dependence — in fact, re-scaling with eventually fractional exponents is a tool that permeates nowadays every branch of Mathematics and Physics.

2. Random polygonals and their renormalized limits

Starting from a straight segment L_0 of length ℓ_0, we define a random polygonal line as the result of an iterative procedure where in each iteration each polygonal segment is randomly (in what concerns the angle they form) folded up at its middle point, and rescaled by a factor 2^H, where $0 \leq H \leq 1$. Note that $H = 0$ corresponds to the Pacheco d'Amorim's definition and $H = 1$ to the case where we double the number of segments but

Proceedings of the *ITI 2011 33rd Int. Conf. on Information Technology Interfaces,* June 27-30, 2011, Cavtat, Croatia

the length of each segment is the original length ℓ_0.

After n iterations, the polygonal line L_n has 2^n sides each one with length $\ell_0 2^{n(H-1)}$ (and the polygonal line has total length $\ell_0 2^{nH}$). The angle formed by each side with the abscissas axis will be characterized by some random variable $\theta_i, i = 1, \ldots, 2^n$ with support in $[0, 2\pi)$. Fig. 1 illustrates some examples of a random polygonal line after n iterations (for $n \in \{1, 2, 3, 5, 7, 9, 11, 13, 15\}$).

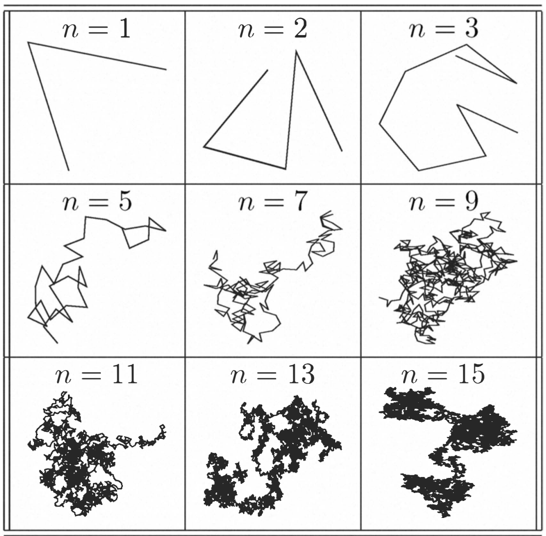

Figure 1. Random polygonal lines

We will also analyze the case where the folding is chosen at random in each former segment. Our main goal is to investigate the limiting behavior of L_n for various exponents $H \in [0, 1]$, searching for a value of H such that the square of the distance between the two endpoints of the polygonal line $\left(D_n^2\right)$ converges to a non-degenerate random variable D_∞^2 with finite mean.

3. Simulation of D_n^2 with independent angles

In this section our purpose is to analyze the expected value of D_n^2 when $n \to \infty$. Table 1 exhibits the results of the simulations of D_n^2 performed with *Mathematica 7.0*, using $\ell_0 = 1$, $n \in \{5, 10, 15, 20\}$, $H \in \{.25, .49, .5, .51, .75\}$ and 10000 replicas.

We can infer that, for $H < 0.5$, D_n^2 converges to a degenerate random variable at zero (as in Pacheco d'Amorim's case) and for $H > 0.5$ the expected value of D_n^2 diverges to infinity. For $H = 0.5$ the random variable D_n^2 has expected value and standard deviation which seems to be near unity (the situation where $\ell_0 = 1$). Simulations, for $H \in \{0, 0.5, 1\}$, are illustrated in Fig. 2. Observe that this is what should be expected from self-similarity considerations, i.e. the need of renormalization so that the limit will be neither zero nor infinity.

Table 1. Simulated values of D_n^2

$n = 5 \Leftrightarrow 2^n = 32$ segments				
.25	.49	$H = .5$	.51	.75
0.189	0.901	0.967	1.083	5.582
(.185)	(.902)	(.946)	(1.045)	(5.495)
$n = 10 \Leftrightarrow 2^n = 1\ 024$ segments				
.25	.49	$H = .5$	.51	.75
0.033	0.876	1.022	1.151	30.99
(.033)	(.868)	(.984)	(1.147)	(30.29)
$n = 15 \Leftrightarrow 2^n = 32\ 768$ segments				
.25	.49	$H = .5$	.51	.75
0.006	0.820	0.995	1.228	178.5
(.006)	(.828)	(1.014)	(1.225)	(177.8)
$n = 20 \Leftrightarrow 2^n = 1\ 048\ 576$ segments				
.25	.49	$H = .5$	.51	.75
0.001	0.766	0.998	1.288	1024
(.001)	(.806)	(1.026)	(1.357)	(1007)

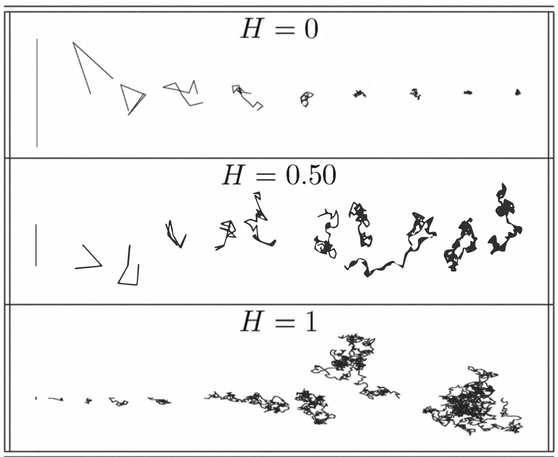

Figure 2. Iterative procedure without scale adjustments

4. D_n^2 distribution with independent angles

Assume now that one of the endpoints of the polygonal line is fixed in $A_0 = (x_0, y_0)$.

As the first side makes an angle θ_1 with the abscissas axis, the other endpoint of the first side l_1 will have coordinates $A_1 = (x_1, y_1)$ given by $(x_0 + \ell_1 \cos(\theta_1), y_0 + \ell_1 \sin(\theta_1))$. The second side l_2 will have one of its endpoint in A_1 (shared with l_1) and the other endpoint will depend of the angle θ_2 that l_2 formed with the abscissas axis. Therefore, the other endpoint of l_2 will be $A_2 = (x_2, y_2)$ given by $(x_1 + \ell_2 \cos(\theta_2), y_1 + \ell_2 \sin(\theta_2))$. Hence, the point $A_k = (x_k, y_k)$ will have coordinates $x_k = x_0 + \sum_{j=1}^{k} \ell_j \cos(\theta_j)$ and $y_k = y_0 + \sum_{j=1}^{k} \ell_j \sin(\theta_j)$. Thus, the square of the distance between the two endpoints, after n iterations, is $D_n^2 = X_n^2 + Y_n^2$ where $X_n = \sum_{j=1}^{2^n} \ell_j \cos(\theta_j)$ and $Y_n = \sum_{j=1}^{2^n} \ell_j \sin(\theta_j)$. Let X_∞ and Y_∞ represent the limit, as $n \to \infty$, of those random variables.

4.1. Fold up at middle point

Assuming that in each iteration all segments are fold up at middle point, then all the sides l_j will have the same length $\ell_j = \ell_0 2^{n(H-1)}$, $\forall j$. With independent uniform random variables θ_i, the Central Limit Theorem implies that in the limit $n \to \infty$ the random variables $\sum_{j=1}^{2^n} \cos(\theta_j)$ and $\sum_{j=1}^{2^n} \sin(\theta_j)$ will be Gaussian with zero mean and variance 2^{n-1}. Thus, X_∞ and Y_∞ have Gaussian distribution with zero mean and variance $0.5\,\ell_0^2\, 2^{n(2H-1)}$, which diverges to infinity when $H > 0.5$ and goes to zero when $H < 0.5$. For $H = 0.5$, $\sigma_{X_\infty}^2 = \sigma_{Y_\infty}^2 = 0.5\,\ell_0^2$. In fact, the limit random variables X_∞ and Y_∞ are independent because they are uncorrelated (as a consequence of $\mathbb{E}[\cos(\theta_i)\sin(\theta_j)] = 0, \forall i, j$) and they are both Gaussian. So, $D_\infty^2 = 0.5\ \ell_0^2\ W$ where $W \sim \chi_2^2$ (chi-square distribution with two degrees of freedom) for $H = 0.5$, thus $\mathbb{E}\left[D_\infty^2\right] = \ell_0^2$ and $\sigma_{D_\infty^2}^2 = \ell_0^4$. In Fig. 3 we exhibit the theoretical limit distribution function (in orange) and the empirical distribution function (in gray) obtained by simulation with 1000 observations. We can observe that, even with few iterations, the approximation is strikingly good.

In 1905, Lord Rayleigh [4] deduced the

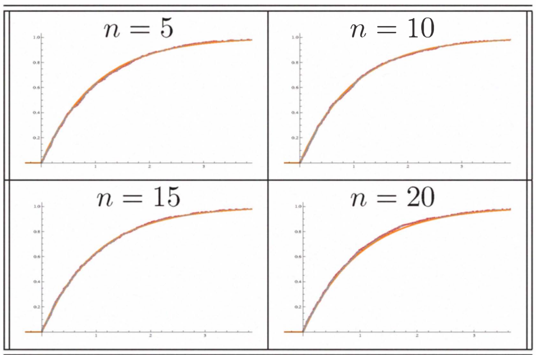

Figure 3. Theoretical versus simulated distribution function

probability density

$$\mathbb{P}(r \leq D_N \leq r + \delta r) = \frac{2}{N\ \ell^2} e^{-\frac{r^2}{N\ell^2}}\ r\delta r \quad (1)$$

of the distance between the two endpoints of a polygonal line with N sides, with length ℓ and random angles. If we fold up in middle point, after n iterations we have $N = 2^n$ sides with length $\ell = \ell_0 2^{n(H-1)}$ that, for $H = 0.5$, we find $\ell = \ell_0 2^{-0.5\frac{\ln(N)}{\ln(2)}} = \frac{\ell_0}{\sqrt{N}}$. If we use $\ell = \frac{\ell_0}{\sqrt{N}}$ in the random flight of Lord Rayleigh, when $N \to \infty$, then $\mathbb{E}(D_\infty) = \ell_0 \Gamma\left(\frac{3}{2}\right) = \frac{\ell_0\sqrt{\pi}}{2}$ and $\mathbb{E}\left(D_\infty^2\right) = \ell_0^2 \Gamma(1) = \ell_0^2$. This way, in the random flight of Lord Rayleigh, with H=0.5, we get finite variance $\left(\sigma^2 = \ell_0^2\left(1 - \frac{\pi}{4}\right)\right)$ when the number of sides goes to infinity.

4.2. Fold up at random point

If we do not use the middle point, choosing instead a point at random (uniformly) in each side in the former iteration step to fold up, we start with one segment of length ℓ_0, and in the first iteration we get two segments with lengths $U_1 \ell_0 2^H$ and $(1 - U_1)\,\ell_0 2^H$; after two iterations we have four segments with lengths $U_1 U_2 \ell_0 2^{2H}$, $U_1(1 - U_2)\,\ell_0 2^{2H}$, $(1 - U_1)\,U_3 \ell_0 2^{2H}$ and $(1 - U_1)(1 - U_3)\,\ell_0 2^{2H}$; $\ldots$; after n iterations we have 2^n segments with lengths $\ell_j = \ell_0\, 2^{nH} \prod_{k=1}^{n} U_k^{[j]}$, where $U_k^{[j]}$, $k = 1, \ldots, n$ and $j = 1, \ldots, 2^n$, are independent random variables with uniform distribution over $(0, 1)$ (except in some cases where $U_k^{[j_1]} = U_k^{[j_2]}$ or $U_k^{[j_1]} = 1 - U_k^{[j_2]}$, $j_1 \neq j_2$, as the previous exemplification of the second iteration illustrates). However, $\mathbb{E}[\ell_j \cos(\theta_j)\, \ell_k \cos(\theta_k)] = 0$ for $j \neq k$ as a consequence of $\mathbb{E}[\cos(\theta_i)\cos(\theta_j)] = 0, \forall i \neq j$

(and the independence of ℓ_j and θ_i, $\forall i, j$) and the variables $\ell_j \cos(\theta_j)$ and $\ell_k \cos(\theta_k)$ are uncorrelated. So, $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = \sigma^2_{X_n}$ is given by

$$\ell_0^2 2^{2nH} \sum_{j=1}^{2^n} \prod_{k=1}^{n} \mathbb{E}\left[\left(U_k^{[j]}\right)^2\right] \mathbb{E}\left[\cos^2(\theta_j)\right] =$$

$$= \ell_0^2 \left(\frac{2^{2H}}{3}\right)^n \frac{2^n}{2} = 0.5\ell_0^2 \left(\frac{2^{2H+1}}{3}\right)^n. \quad (2)$$

And, analogous, we can conclude that $\mathbb{E}[Y_n] = 0$ and $\sigma^2_{Y_n} = 0.5\ell_0^2\left(\frac{2^{2H+1}}{3}\right)^n$. For $H > \frac{\ln(1.5)}{\ln(4)} \approx 0.29248$ the variances $\sigma^2_{X_n}$ and $\sigma^2_{Y_n}$ diverge to infinity; for $H < \frac{\ln(1.5)}{\ln(4)}$ converge to zero and for $H = \frac{\ln(1.5)}{\ln(4)}$ we get $\sigma^2_{X_\infty} = \sigma^2_{Y_\infty} = 0.5\,\ell_0^2$ and, therefore, D^2_∞ has finite mean. However, in this case, we cannot use the classical Central Limit Theorem because $\ell_j \cos(\theta_j)$ and $\ell_k \cos(\theta_k)$ are not independent (although uncorrelated). In Fig. 4 we compare the empirical distribution function of D_n^2 (by simulation with $\ell_0 = 1$, $H = \frac{\ln(1.5)}{\ln(4)}$ and 1000 observations) with the distribution function in the situation that we fold up in middle point ($D^2_\infty = 0.5\,\ell_0^2\,W$ where $W \sim \chi^2_2$). We can easily deduce their dissimilitude and infer that, in this situation, the mean value will be lesser than unity.

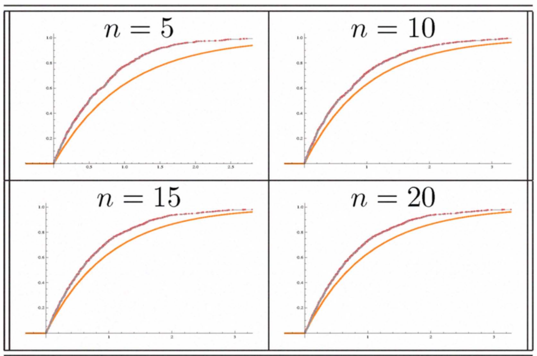

Figure 4. Theoretical versus simulated distribution function

5. D_n^2 simulation with dependent angles

For simplicity, let θ'_i denote the angle of the side l_i with the side l_{i-1} (not with the abscissas axis as in the previous sections). This way we can limit the variation of θ'_i in order to *smoothen* the polygonal line obtained in the limit (this can be done by diminishing the support or using a distribution where we have high probability of getting values in the neighborhood of zero).

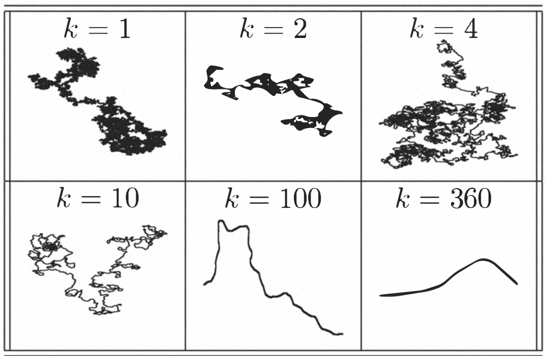

Figure 5. Random polygonal with angle dependence

First, we assume that the θ'_i are independent random variables with uniform distribution with support $\left[-\frac{\pi}{k}, \frac{\pi}{k}\right]$ (but, for $k \neq 1$, θ_i are dependent — θ_i depends of previous angle θ_{i-1}). In Fig. 5 we can see the results of simulation without using re-scaling factor ($H = 0$, $n = 15$, $\ell_0 = 1$). In Table 4 we register the simulation results (10000 replicas).

Table 2. D_n^2 simulation — dependent angles

$n = 15$ iterations			
$k = 1$	$k = 2$	$k = 4$	$k = 10$
0.00010	0.00045	0.00194	0.01213
(.00010)	(.00045)	(.00198)	(.01197)
$k = 50$	$k = 100$	$k = 180$	$k = 360$
0.25411	0.62003	0.84931	0.95930
(.18807)	(.21925)	(.11543)	(.03489)
$n = 20$ iterations			
$k = 4$	$k = 10$	$k = 180$	$k = 360$
0.00019	0.00121	0.31797	0.68480
(.00019)	(.00122)	(.21279)	(.19795)

As expected, dependence decreases the speed of convergence of D_n^2 to a degenerate random variable (nonetheless, it still converges to zero). In Table 2 we register the results of simulation, $k \in \{4, 10, 180, 360\}$, using $n = 20$ iterations to stress this point. We can notice that D_n^2 decreases when we increase the number of iterations.

We can apply another continuous and symmetric distribution with mode at 0 for the angles θ'_i, and therefore produces a smoother

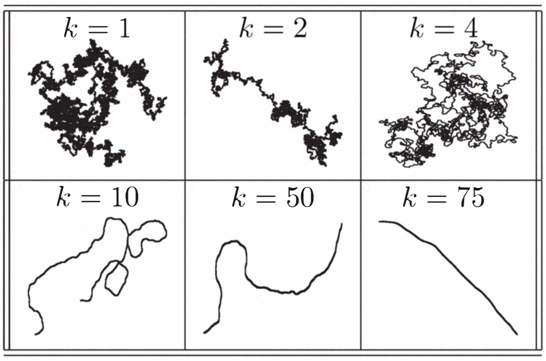

Figure 6. Random polygonal, Beta angles

limiting curve. In Fig. 6 and Table 3 we show the simulation results, $\theta_i' = (\theta_i'' - 0.5)\frac{2\pi}{k}$, where θ_i'' are independent Beta$(2,2)$ random variables. Even with $k = 360$, when we increase the number of iterations the values of D_n^2 decrease.

Table 3. D_n^2 **simulation — Beta(2,2)**

k	2	4	360	360	360
n	15	15	15	17	20
$\overline{x}$	.0032	.0130	.9937	.9392	.5838
s	.0032	.0129	.0057	.0515	.2299

6. D_n^2 distribution with dependent angles

In this case the point $A_k' = (x_k', y_k')$, as we denote the endpoint of side l_k, will have coordinates $x_k' = x_0' + \sum_{i=1}^{k} \ell_i \cos\left(\sum_{j=1}^{i} \theta_j'\right)$ and $y_k' = y_0' + \sum_{i=1}^{k} \ell_i \sin\left(\sum_{j=1}^{i} \theta_j'\right)$. The square of the distance between the two endpoints (D_n^2) is

$$\sum_{i=1}^{2^n} \ell_i^2 + 2\sum_{i=1}^{2^n-1} \sum_{j=i+1}^{2^n} \ell_i\, \ell_j \cos\left(S_{j-i}\right), \quad (3)$$

where $S_{j-i}=\sum_{k=i+1}^{j} \theta_k'$. Assuming that the θ_i' are symmetric, the corresponding characteristic function is $\varphi_{\theta_i'}(t)=\mathbb{E}\left[e^{it\theta_i'}\right]=\mathbb{E}\left[\cos\left(t\theta_i'\right)\right]$. Hence, $\mathbb{E}\left[\cos\left(S_m\right)\right]$, $m \in \mathbb{N}$, is given by

$$\varphi_{S_m}(1) = \left(\varphi_{\theta'}(1)\right)^m = \left(\mathbb{E}\left[\cos\left(\theta'\right)\right]\right)^m, \quad (4)$$

so using $\rho = \mathbb{E}\left(\cos\left(\theta_i'\right)\right) \in [0,1)$ (we did not consider the random variable θ_i' equal to zero with probability one), then

$$\mathbb{E}\left(D_n^2\right) = \sum_{i=1}^{2^n} \ell_i^2 + 2\sum_{i=1}^{2^n-1} \sum_{j=i+1}^{2^n} \ell_i\, \ell_j\, \rho^{j-i}. \quad (5)$$

If the two sides have the same length (i.e. folding up at the middle point), then $\mathbb{E}\left(D_n^2\right)$ can be simplified to:

$$\ell^2\left[2^n + 2\sum_{i=1}^{2^n-1} \rho \frac{1-\rho^{2^n-i}}{1-\rho}\right] = \ell^2 \xi\left(n,\rho\right), \quad (6)$$

where $\ell = \ell_0 2^{n(H-1)}$ and $\xi\left(n,\rho\right)$ is given by

$$2^n + \frac{2\rho(2^n-1)}{(1-\rho)} - \frac{2\rho^2\left(1-\rho^{2^n-1}\right)}{(1-\rho)^2}. \quad (7)$$

We further observe that for the limit of $\mathbb{E}\left(D_n^2\right)$ being finite and not null we need to use H=0.5 and $\mathbb{E}\left(D_\infty^2\right)=\ell_0^2\,\frac{1+\rho}{1-\rho}$. So, in order to get $\mathbb{E}\left(D_\infty^2\right)=l$, we must use H=0.5 and start with a $\ell_0 = \sqrt{l\,\frac{1-\rho}{1+\rho}}$.

Table 4. D_n^2 **simulation,** $H = 0.5$

	$n = 5$	$n = 10$	$n = 15$
Uniform $k = 1$	1.048 (.996)	1.028 (1.033)	.975 (.983)
	The limit is 1		
Uniform $k = 45$	31.73 (.227)	790.0 (162.7)	2271 (2172)
	The limit is 2461.7		
Uniform $k = 360$	32.00 (.004)	1019 (3.705)	28705 (3180)
	The limit is 157574		
Beta $k = 1$	1.809 (1.736)	1.840 (1.784)	1.850 (1.902)
	The limit is 1.873		
Beta $k = 45$	31.83 (.147)	867.4 (118.9)	3856 (3529)
	The limit is 4103.2		
Beta $k = 360$	31.99 (.002)	1021. (2.254)	30216. (2086)
	The limit is 262624		

For a finite number of iterations, if we use the scaling factor $H = 0.5$ (as in the previous section) we shall not be able to avoid the alteration of D_n^2 when we use different numbers of iterations, converging to the limit $\ell_0^2\frac{1+\rho}{1-\rho}$ (as the simulations presented in Table 4 illustrates).

In Table 5 we show the simulation results when the initial segment is $\ell_0 = \sqrt{\frac{1-\rho}{1+\rho}}$, so that the expected value of D_n^2 converges to 1.

Table 5. D_n^2 **simulation,** $\ell_0 = \sqrt{\frac{1-\rho}{1+\rho}}$

	$n=5$	$n=10$	$n=15$
Uniform $k=1$	.988 (.913)	1.002 (1.008)	.971 (.933)
Uniform $k=45$	.013 (.0001)	0.322 (.065)	.945 (.890)
Uniform $k=360$	.0002 (.0000)	.006 (.00002)	.181 (.021)
Beta $k=1$	.963 (.913)	0.965 (.920)	.999 (1.025)
Beta $k=45$	.008 (.00003)	0.212 (.030)	.913 (.824)
Beta $k=360$	00012. (.0000)	0.0039 (.0000)	.116 (.007)

For those results we used $\rho = \varphi_{\theta_i'}(1)$, and hence for the uniform distribution over $\left[-\frac{\pi}{k}, \frac{\pi}{k}\right]$ we get $\rho = \frac{\sin(\pi/k)}{\pi/k}$; on the other hand, for the Beta distribution case we have $\rho = 3\left[\frac{\sin(\pi/k)}{(\pi/k)^3} - \frac{\cos(\pi/k)}{(\pi/k)^2}\right]$. The quotient $R_n = \frac{\mathbb{E}(D_n^2)}{\mathbb{E}(D_\infty^2)}$ can be used to grasp the convergence rate as shown in Table 6.

Table 6. Convergence rates

	Uniform		
n	$k=1$	$k=45$	$k=360$
5	1	.0129	.0002
10	1	.3211	.0065
15	1	.9624	.1819
20	1	.9988	.9249
25	1	1	.9977
30	1	1	.9999
	Beta		
n	$k=1$	$k=45$	$k=360$
5	.9791	.0078	.0001
10	.9993	.2127	.0039
15	.9999	.9374	.1150
20	1	.9980	.8748
25	1	.9999	.9961
30	1	1	.9999

7. Conclusion

Our results show that the choice of an appropriate exponent for re-scaling achieves the convenient result: the contraction of D due to folding is counteracted by the stretching using an appropriate factor 2^H, so that at the end we have non-degenerate random variable, and an almost surely open curve, as empirical facts imply.

Acknowledgements: *The author wishes to thank the referees for suggesting several improvements and corrections and express his gratitude to Professor Dinis Pestana, for his guidance, encouragement and continuous support. This research was partially supported by FCT/OE.*

References

[1] Chow, B., Lu, P., and Li, N. (2006). *Hamilton's Ricci Flow.* American Mathematical Society, Providence, RI.

[2] Falconer, K. *Fractal Geometry, Mathematical Foundations and Applications.* Chichester: Wiley; 1990.

[3] Mendonça, S., Pestana, D., and Sousa Santos, R. (2007). *Diogo Pacheco d'Amorim's The Elements of Probability Calculus — diplomatic bilingual edition of Pacheco d'Amorim's 1914 thesis on the construction of Probability*; English translation and comments, `http://www.estg.ipleiria.pt/~rsantos/Elements_of_Probability_Calculus.pdf`.

[4] Rayleigh, L. (1905). The problem of the random walk, *Nature*, **72**, 318.

[5] Santos, R. S. (2008). *Probability Circa 1914 and the Construction of Pacheco d'Amorim* (*Probabilidade Circa 1914 e a Construção de Pacheco d'Amorim*), Universidade de Lisboa.

[6] Santos, R., and Pestana, D. (2009). Distance between the endpoints of a random polygonal line (Distância entre os extremos de uma linha poligonal "ao acaso"), In Oliveira, I., Correia, E., Ferreira, F., Dias, S., e Braumann, C. (eds.) Estatística: Arte de Explicar o Acaso, SPE, 591-602.

[7] Wilson, K. G. (1982). The renormalization group and critical phenomena, Nobel Lecture, `http://nobelprize.org/nobel_prizes/physics/laureates/1982/wilson-lecture.pdf`.