The perimeter of uniform and geometric words: a probabilistic analysis

Guy Louchard*

August 24, 2017

Abstract

Let a word be a sequence of n i.i.d. integer random variables. The perimeter P of the word is the number of edges of the word, seen as a polyomino. In this paper, we present a probabilistic approach to the computation of the moments of P. This is applied to uniform and geometric random variables. We also show that, asymptotically, the distribution of P is Gaussian and, seen as a stochastic process, the perimeter converges in distribution to a Brownian motion

Keywords: Words, perimeter, moments, probabilistic approach, Gaussian distribution, Brownian motion

2010 Mathematics Subject Classification: 05A16, 05A05, 60C05, 60F05

1 Introduction

Our attention was recently attracted by a paper by Blecher et al. [4] on the perimeter of words: a word is a sequence of n i.i.d. integer random variables (RV) $\{x_0, x_1, \ldots, x_m\}, m := n-1$. In [4], the RV are distributed uniformly on $[1, k]$. The perimeter P of the word is the number of edges of the word, seen as a polyomino. A typical polyomino, based on the word $2, 3, 1, 3, n = 4, P = 18$ is given in Fig.1. The mean $M_P := \mathbb{E}(P)$ and variance $\mathbb{V}(P)$ of P are given in [4] by the following theorem:

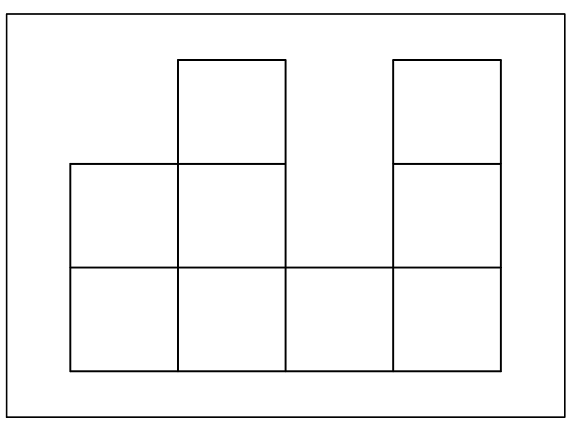

Figure 1: The polyomino based on the word $2, 3, 1, 3, n = 4, P = 18$

*Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, email: louchard@ulb.ac.be

Theorem 1.1 *In the uniform $[1,k]$ case, M_P and $\mathbb{V}(P)$ are given in [4] by*

$$M_P = (n-1)M + 2n + (k+1) = \frac{(3k+2k^2+1)+(k^2+6k-1)n}{3k}, \tag{1}$$
$$\mathbb{V}(P) = \frac{(-5k^2+4k^4+1)+(-3+3k^4)n}{45k^2}. \tag{2}$$

Some years ago, we had been interested in some uniformly distributed words: see [10]. Moreover, we had analyzed some polyominos, for instance in [7] and [8], where, in particular, we had derived some limiting Brownian motion (BM) processes for trajectories. Some recent papers on polyomino's perimeter are, for instance, [5], [6].

Another classical distribution is the classical geometric(p) one, with distribution $pq^{i-1}, i \geq 1, p \in (0,1), q := 1-p$. In several papers (some of them with H.Prodinger) we had analyzed related words parameters from a probabilistic point of view. Our last papers on this topic being [13], [12]. We again derived some limiting BM processes, for instance in [9], [11]. For other recent papers on geometric words, see [1], [2].

In the present paper, our motivation is to present a novel approach to the words perimeter problem:

- a probabilistic approach easily leads to the moments of P,
- the distribution of P is asymptotically shown to be Gaussian,
- seen as a stochastic process, the perimeter converges in distribution to a BM,
- our technique is applied to the geometric(p) case.

2 The mean and variance of P in the uniform case

In this section, we present a probabilistic approach to the mean and variance of P.

Set $P_m := \sum_1^m y_i, y_i := |x_i - x_{i-1}|$. Clearly, $P = P_m + x_0 + x_m + 2n$. For further use, we define $P_n := P_m + x_0 + x_m$. We see that the y_i are identically distributed, y_i is correlated with y_{i+1}, but *independent* of $y_k, k \geq i+2$.

The following notations and relations will be used throughout the paper:

$$\begin{aligned}
M &:= \mathbb{E}(y_i), \\
M_m &:= \mathbb{E}(P_m) = mM, \\
M_n &:= \mathbb{E}(P_n) = M_n + 2\mathbb{E}(x_0), \\
M_P &:= \mathbb{E}(P) = M_n + 2n, \\
T_{\alpha,\beta,\gamma,\delta} &:= \mathbb{E}\left(x_0^\alpha, y_1^\beta, y_2^\gamma, y_3^\gamma\right), \\
\overline{T_{0,\beta,\gamma,\delta}} &:= \mathbb{E}\left((y_1-M)^\beta, (y_2-M)^\gamma, (y_3-M)^\gamma\right).
\end{aligned}$$

For instance, $M \equiv T_{0,1}, T_1 = \mathbb{E}(x_0) = \frac{k+1}{2}$ for the uniform case. Let us first compute the distribution of y_i: $f(u) := \mathbb{P}(y_i = u), u \in [0, k-1]$. Consider first the case $u > 0$. If $x_1 > x_0, u = x_1 - x_0, x_1 = u + x_0$. But $1 \leq x_1 \leq k$, hence $1 \leq x_0 \leq k-u$. So we first have

$$S_1 = \frac{1}{k}\sum_1^{k-u} \mathbb{P}(x_0 = i) = \frac{k-u}{k^2}.$$

Next, if $x_1 < x_0, u = x_0 - x_1, x_1 = x_0 - u$, But $1 \leq x_1 \leq k$, hence $1 + u \leq x_0 \leq k$. So

$$S_2 = \frac{1}{k}\sum_{1+u}^{k} \mathbb{P}(x_0 = i) = \frac{k-u}{k^2}.$$

Finally,

$$f(u) = S_1 + S_2 = \frac{2(k-u)}{k^2}, u > 0.$$

In the case $u = 0$, we simply have $f(0) = \frac{1}{k^2}\sum_1^k 1 = \frac{1}{k}$. A plot of $f(u), k = 6$, is given in Fig.2.

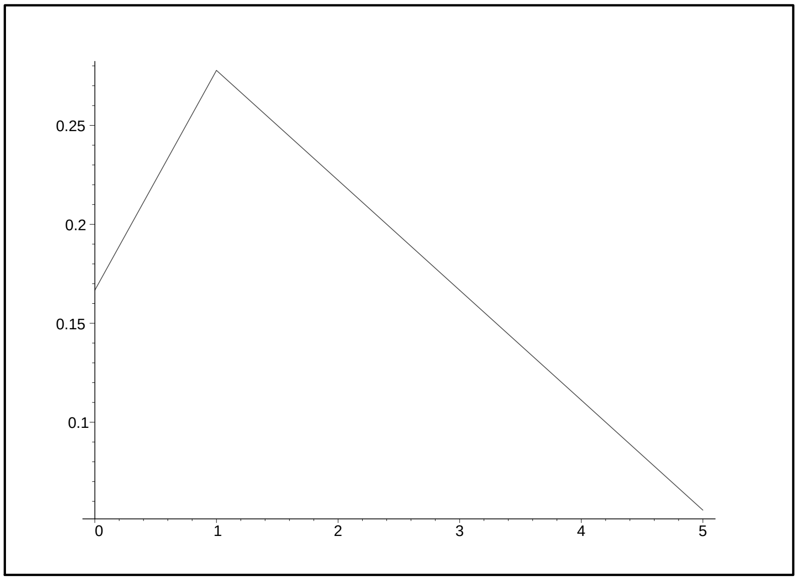

Figure 2: $f(u), k = 6$

Now we are ready to compute M. This is given either by

$$M := T_{0,1} := \sum_{i=1}^{k}\Big(\sum_{j=i}^{k}(j-i) + \sum_{j=1}^{i-1}(i-j)\Big) \Big/ k^2 = \frac{(k-1)(k+1)}{3k},$$

or $\sum_0^{k-1} f(u)u$, which of course leads to the same result.

Hence

$$M_n = (n-1)M + (k+1) = \frac{(3k+2k^2+1)+(k^2-1)n}{3k},$$
$$M_P = (n-1)M + 2n + (k+1) = \frac{(3k+2k^2+1)+(k^2+6k-1)n}{3k},$$

which fits with(1).

Some useful expressions will be used in this section. We collect them here.

$$T_1 = \sum_{i=1}^{k} i/k = \frac{k+1}{2},$$

$$T_2 = \sum_{i=1}^{k} i^2/k = \frac{(k+1)(1+2k)}{6},$$

$$T_{1,1} = \sum_{i=1}^{k} i\Big(\sum_{j=i}^{k}(j-i) + \sum_{j=1}^{i-1}(i-j)\Big) \Big/ k^2 = \frac{(k-1)(k+1)^2}{6k},$$

$$T_{0,2} = \sum_{i=1}^{k}\sum_{j=1}^{k}(j-i)^2 \Big/ k^2 = \sum_{u=1}^{k-1} f(u)u^2 = \frac{(k-1)(k+1)}{6},$$

$$\overline{T_{0,2}} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i-M)^2 + \sum_{1}^{i-1}(i-j-M)^2) \Big/ k^2 = \frac{(k-1)(k+1)(k^2+2)}{18k^2},$$

$$T_{0,1,1} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i)(\sum_{l=j}^{k}(l-j) + \sum_{l=1}^{j-1}(j-l)) + \sum_{j=1}^{i-1}(i-j)(\sum_{l=j}^{k}(l-j) + \sum_{l=1}^{j-1}(j-l))) \Big/ k^3$$
$$= \frac{(k-1)(k+1)(7k^2-8)}{60k^2},$$

$$\overline{T_{0,1,1}} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i-M)(\sum_{l=j}^{k}(l-j-M) + \sum_{l=1}^{j-1}(j-l-M))$$
$$+ \sum_{j=1}^{i-1}(i-j-M)(\sum_{l=j}^{k}(l-j-M) + \sum_{l=1}^{j-1}(j-l-M))) \Big/ k^3$$
$$= \frac{(k-1)(k-2)(k+2)(k+1)}{180k^2}.$$

These expressions are the only necessary ones in order to compute $\mathbb{V}(P)$.

Now we turn to the computation of variance $\mathbb{V}(P)$. Of course, only P_n must be used here. The dominant term of $\mathbb{V}(P_n)$ is immediate: this is given by

$$n[(T_{0,2} - M^2) + 2(T_{0,1,1} - M^2)] = nV^*,$$
$$V^* = \frac{(k-1)(k+1)(k^2+1)}{15k^2},$$

which is of course also given by $\overline{T_{0,2}} + 2\overline{T_{0,1,1}}$.

To compute $\mathbb{V}(P)$, we must collect all necessary terms. We symbolically expand

$$(x_0 + x_m + y_1 + y_i + y_{i+1} + y_{i+2} + y_m)^2.$$

We collect the relevant contributions, with their weights:

$$y_i^2 \to mT_{0,2},$$
$$y_i y_{i+1} \to 2(m-1)T_{0,1,1},$$
$$x_0^2, x_m^2 \to 2T_2,$$
$$x_0 y_1, y_m x_m \to 4T_{1,1},$$
$$y_i y_k, k \geq i+2 \to (m-2)(m-3)M^2,$$
$$y_1 y_k, k \geq 3, y_m y_k, k \leq m-2 \to 2(m-2)M^2,$$
$$x_0 x_m \to 2T_1^2,$$
$$x_0 y_i, x_m y_i \to 4T_1 M(m-1).$$

This gives

$$\mathbb{V}(P_n) = (n-1)T_{0,2} + 2T_2 + (n-2)2T_{0,1,1} + 4T_{1,1} + ((n-1)(n-4)+2)M^2 + 2T_1^2 + 4T_1M(n-2) - M_n^2$$
$$= \frac{(-5k^2 + 4k^4 + 1) + (-3 + 3k^4)n}{45k^2},$$

which fits with (2).

3 The third centered moment $\mu_3(P)$ of P in the uniform case

In this section, we apply our probabilistic technique to the third centered moment computation. We will only compute the $n-$dominant term of $\mu_3(P)$, the complete analysis goes as in Sec. 2, only with elementary but tedious algebra, we omit the details.

The necessary expressions are as follows:

$$\overline{T_{0,3}} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i-M)^3 + \sum_{j=1}^{i-1}(i-j-M)^3) \Big/ k^2 = \frac{(k-1)(k-2)(k+2)(k+1)(2k^2-5)}{270k^3},$$

$$T_{0,3} = \sum_{u=1}^{k-1} f(u)u^3 = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i)^3 + \sum_{1}^{i-1}(i-j)^3) \Big/ k^2 = \frac{(k-1)(k+1)(3k^2-2)}{30k},$$

$$\begin{aligned}\overline{T_{0,1,1,1}} = &\sum_{i=1}^{k}(\sum_{j=i}^{k}((j-i-M)(\sum_{l=j}^{k}(l-j-M)(\sum_{r=l}^{k}(r-l-M)\\ &+\sum_{r=1}^{l-1}(l-r-M)) + \sum_{l=1}^{j-1}(j-l-M)(\sum_{r=l}^{k}(r-l-M) + \sum_{r=1}^{l-1}(l-r-M)))\\ &+\sum_{j=1}^{i-1}(i-j-M)(\sum_{l=j}^{k}(l-j-M)(\sum_{r=l}^{k}(r-l-M) + \sum_{r=1}^{l-1}(l-r-M))\\ &+\sum_{l=1}^{j-1}(j-l-M)(\sum_{r=l}^{k}(r-l-M) + \sum_{r=1}^{l-1}(l-r-M)))) \Big/ k^4 = -\frac{(k-1)(k-2)(k+2)(k+1)(k^2+5)}{3780k^3},\end{aligned}$$

$$\begin{aligned}T_{0,1,1,1} = &\sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i)(\sum_{l=j}^{k}(l-j)(\sum_{r=l}^{k}(r-l)\\ &+\sum_{r=1}^{l-1}(l-r)) + \sum_{l=1}^{j-1}(j-l)(\sum_{r=l}^{k}(r-l) + \sum_{r=1}^{l-1}(l-r)))\\ &+\sum_{j=1}^{i-1}(i-j)(\sum_{l=j}^{k}(l-j)(\sum_{r=l}^{k}(r-l) + \sum_{r=1}^{l-1}(l-r)) + \sum_{l=1}^{j-1}(j-l)(\sum_{r=l}^{k}(r-l) + \sum_{r=1}^{l-1}(l-r)))) \Big/ k^4\\ &= \frac{(k-1)(k+1)(17k^4-39k^2+24)}{420k^3},\end{aligned}$$

$$\begin{aligned}\overline{T_{0,1,2}} = &\sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i-M)(\sum_{l=j}^{k}(l-j-M)^2\\ &+\sum_{l=1}^{j-1}(j-l-M)^2) + \sum_{j=1}^{i-1}(i-j-M)(\sum_{l=j}^{k}(l-j-M)^2 + \sum_{l=1}^{j-1}(j-l-M)^2)) \Big/ k^3\\ &= \frac{(k-1)(k-2)(k+2)(k+1)(k^2+2)}{540k^3},\end{aligned}$$

$$T_{0,1,2} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i)(\sum_{l=j}^{k}(l-j)^2 + \sum_{l=1}^{j-1}(j-l)^2) + \sum_{j=1}^{i-1}(i-j)(\sum_{l=j}^{k}(l-j)^2 + \sum_{l=1}^{j-1}(j-l)^2)) \Big/ k^3$$

$$= \frac{(k-1)(k+1)(11k^2-14)}{180k},$$

$$\overline{T_{0,2,1}} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i-M)^2(\sum_{l=j}^{k}(l-j-M)+\sum_{l=1}^{j-1}(j-l-M))$$
$$+\sum_{j=1}^{i-1}(i-j-M)^2(\sum_{l=j}^{k}(l-j-M)+\sum_{j=1}^{i-1}(j-l-M)))\Big/ k^3 = \frac{(k-1)(k-2)(k+2)(k+1)(k^2+2)}{540k^3},$$

$$T_{0,2,1} = \sum_{i=1}^{k}(\sum_{j=i}^{k}(j-i)^2(\sum_{l=j}^{k}(l-j)+\sum_{l=1}^{j-1}((j-l))+\sum_{j=1}^{i-1}(i-j)^2(\sum_{l=j}^{k}(l-j)+\sum_{l=1}^{j-1}(j-l)))\Big/ k^3$$
$$= \frac{(k-1)(k+1)(11k^2-14)}{180k}.$$

Now we symbolically expand (recall that y_i is independent of y_{i+2})

$$(\overline{y_i}+\overline{y_{i+1}}+\overline{y_{i+2}})^3, \text{ with } \overline{y_i} := y_i - M.$$

We must only retain the terms

$$S := \overline{y_i}^3 + 3\overline{y_i y_{i+1}}^2 + 3\overline{y_i}^2\overline{y_{i+1}} + 6\overline{y_i y_{i+1} y_{i+2}}.$$

We expand, this leads to

$$(y_i^3 + 3y_iy_{i+1}^2 + 6y_iy_{i+1}y_{i+2} + 3y_i^2y_{i+1}) + (-6y_i^2 - 3y_{i+1}^2 - 18y_iy_{i+1} - 6y_{i+1}y_{i+2} - 6y_iy_{i+2})M$$
$$+ (15y_{i+1} + 6y_{i+2} + 18y_i)M^2 - 13M^3.$$

We make a three steps substitution, *in this order*

- $$y_i^3 = T_{0,3}, y_i^2y_{i+1} = T_{0,2,1}, y_iy_{i+1}^2 = T_{0,1,2}, y_iy_{i+1}y_{i+1} = T_{0,1,1,1},$$
- $$y_i^2 = T_{0,2}, y_{i+1}^2 = T_{0,2}, y_iy_{i+1} = T_{0,1,1}, y_{i+1}y_{i+2} = T_{0,1,1}, y_iy_{i+2} = M^2,$$
- $$y_i = M, y_{i+1} = M, y_{i+2} = M.$$

This leads to the dominant term of $\mu_3(P)$

Theorem 3.1 *In the uniform $[1,k]$ case, the dominant term of $\mu_3(P)$ given by*

$$\mu_3(P) = n\mu_3^* + \mathcal{O}(1),$$
$$\mu_3^* = (T_{0,3} + 3T_{0,1,2} + 6T_{0,1,1,1} + 3T_{0,2,1}) + (-9T_{0,2} - 24T_{0,1,1} - 6M^2)M - 26M^3$$
$$= \frac{4(k-2)(1+2k)(2k-1)(k+2)(k-1)(k+1)}{945k^3}.$$

Of course, this can also be obtained as

$$n[\overline{T_{0,3}} + 3\overline{T_{0,1,2}} + 6\overline{T_{0,1,1,1}} + 3\overline{T_{0,2,1}}],$$

but we also gave the first approach, which will be used in the next section.

The fourth centered moment $\mu_4(P)$ can be similarly mechanically computed. Note that the dominant term is there of order n^2: we have contribution of type $\overline{y_i}^2, \overline{y_k}^2, k \geq i+2$.

4 The geometric(p) case

We will now consider the geometric(p) case, with distribution $pq^{i-1}, i \geq 1, p \in (0,1), q := 1-p$. The computation of the centered cross-moments $\overline{T}_{.}$ is rather intricate (in particular with many indices), even for Maple. So we will only use the ordinary cross-moments $T_{.}$.

The distribution $f(u) := \mathbb{P}(y_i = u), u \in [0, \infty]$ is given as follows:

$$f(u) = \sum_{i=1}^{\infty} pq^{i-1}pq^{i+u-1} + \sum_{i=u+1}^{\infty} pq^{i-1}pq^{i-u-1} = \frac{2p(1-p)^u}{2-p}, u > 0,$$

$$f(0) = \sum_{i=1}^{\infty}(pq^{i-1}pq^{i-1}) = \frac{p}{2-p}.$$

A plot of $f(u), p = 1/2$ is given in Fig.3

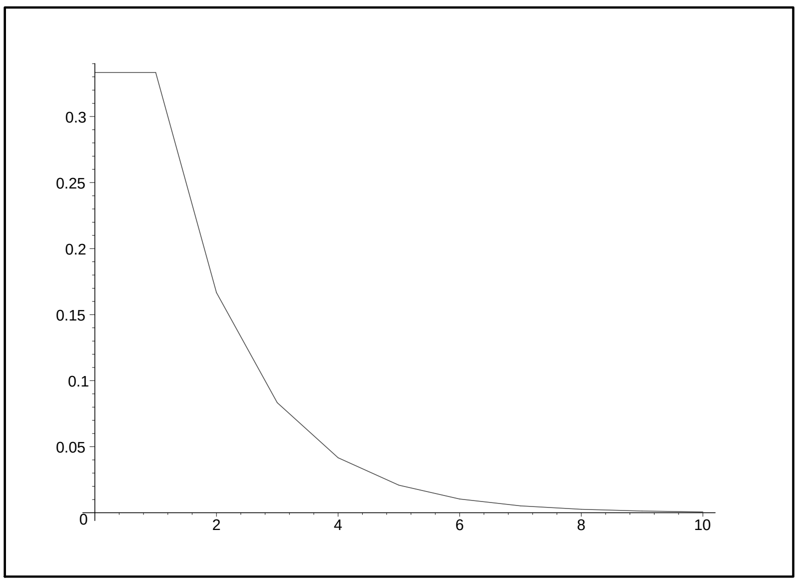

Figure 3: $f(u), p = 1/2$

The first expressions are given as follows

$$T_1 = \sum_{i=1}^{\infty} pq^{i-1}i = \frac{1}{p},$$

$$M = T_{0,1} = \sum_{u=1}^{\infty} f(u)u = \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i) + \sum_{j=1}^{i-1} pq^{j-1}(i-j)) = \frac{21-p)}{p(2-p)},$$

hence

$$M_n = (n-1)M + 2T_1 = \frac{2 + (2-2p)n}{p(2-p)},$$

$$M_P = (n-1)M + 2n + 2T_1 = \frac{2 + (2+2p-2p^2)n}{p(2-p)}.$$

The next necessary expressions are given as follows:

$$T_{0,2} = \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)^2 + \sum_{j=1}^{i-1} pq^{j-1}(i-j)^2) = \sum_{u=1}^{\infty} f(u)u^2 = \frac{2(1-p)}{p^2},$$

$$\begin{aligned} T_{0,3} &= \sum_{u=1}^{\infty} f(u)u^3 = \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)^3 + \sum_{j=1}^{i-1} pq^{j-1}(i-j)^3) \\ &= \frac{2(1-p)(p^2-6p+6)}{p^3(2-p)}, \end{aligned}$$

$$\begin{aligned} T_{0,1,1} &= \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)(\sum_{l=j}^{\infty} pq^{l-1}(l-j) + \sum_{l=1}^{j-1} pq^{l-1}(j-l)) \\ &+ \sum_{j=1}^{i-1} pq^{j-1}(i-j)(\sum_{l=j}^{\infty} pq^{l-1}(l-j) + \sum_{l=1}^{j-1} pq^{l-1}(j-l))) \\ &= \frac{(1-p)(p^4-7p^3+23p^2-32p+16)}{p^2(2-p)^2(p^2+3-3p)}, \end{aligned}$$

$$\begin{aligned} T_{0,1,1,1} &= \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)(\sum_{l=j}^{\infty} pq^{l-1}(l-j)(\sum_{r=l}^{\infty} pq^{r-1}(r-l) + \sum_{r=1}^{l-1} pq^{r-1}(l-r)) \\ &+ \sum_{l=1}^{j-1} pq^{l-1}(j-l)(\sum_{r=l}^{\infty} pq^{r-1}(r-l) + \sum_{r=1}^{l-1} pq^{r-1}(l-r))) \\ &+ \sum_{j=1}^{i-1} pq^{j-1}(i-j)(\sum_{l=j}^{\infty} pq^{l-1}(l-j)(\sum_{r=l}^{\infty} pq^{r-1}(r-l) \\ &+ \sum_{r=1}^{l-1} pq^{r-1}(l-r)) + \sum_{l=1}^{j-1} pq^{l-1}(j-l)(\sum_{r=l}^{\infty} pq^{r-1}(r-l) \\ &+ \sum_{r=1}^{l-1} pq^{r-1}(l-r)))) \\ &= \frac{2(28-84p+113p^2-86p^3+39p^4-10p^5+p^6)(1-p)^2}{p^3(p^2-2p+2)(2-p)(p^2+3-3p)^2}, \end{aligned}$$

$$\begin{aligned} T_{0,1,2} &= \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)(\sum_{l=j}^{\infty} pq^{l-1}(l-j)^2 \\ &+ \sum_{l=1}^{j-1} pq^{l-1}(j-l)^2) + \sum_{j=1}^{i-1} pq^{j-1}(i-j)(\sum_{l=j}^{\infty} pq^{l-1}(l-j)^2) + \sum_{l=1}^{j-1} pq^{l-1}(j-l)^2)) \\ &= \frac{(28-56p+38p^2-10p^3+p^4)(1-p)}{p^3(2-p)^3}, \end{aligned}$$

$$T_{0,2,1} = \sum_{i=1}^{\infty} pq^{i-1}(\sum_{j=i}^{\infty} pq^{j-1}(j-i)^2(\sum_{l=j}^{\infty} pq^{l-1}(l-j),$$

$$+\sum_{l=1}^{j-1} pq^{l-1}(j-l)) + \sum_{j=1}^{i-1} pq^{j-1}(i-j)^2(\sum_{l=j}^{\infty} pq^{l-1}(l-j) + \sum_{l=1}^{j-1} pq^{l-1}(j-l)))$$
$$= \frac{(28-56p+38p^2-10p^3+p^4)(1-p)}{p^3(2-p)^3},$$

$$T_{1,1} = \sum_{i=1}^{\infty} pq^{i-1} i(\sum_{j=i}^{\infty} pq^{j-1}(j-i) + \sum_{j=1}^{i-1} pq^{j-1}(i-j)) = \frac{(1-p)(p^2-4p+6)}{p^2(2-p)^2}.$$

The dominant term of $\mathbb{V}(P_n)$is given by

$$n[(T_{0,2}-M^2) + 2(T_{0,1,1}-M^2)] = nV^*,$$
$$V^* = \frac{4(1-p)(p^4+9p^2-4p^3-10p+5)}{p^2(2-p)^2(p^2+3-3p)}.$$

The exact value of $\mathbb{V}(P)$ is given by

$$\mathbb{V}(P_n) = (n-1)T_{0,2} + 2T_2 + (n-2)2T_{0,1,1} + 4T_{1,1} + ((n-1)(n-4)+2)M^2 + 2T_1^2 + 4T_1M(n-2) - M_n^2$$
$$= \frac{n[4(1-p)(p^4+9p^2-4p^3-10p+5)] + 4(3p^2-5p+5)(1-p)^2}{p^2(2-p)^2(p^2+3-3p)}.$$

The third centered moment $\mu_3(P)$ (dominant term) is given by

$$\mu_3(P) = n[(T_{0,3} + 3T_{0,1,2} + 6T_{0,1,1,1} + 3T_{0,2,1}) + (-9T_{0,2} - 24T_{0,1,1} - 6M^2)M - 26M^3] + \mathcal{O}(1)$$
$$= n\frac{8(1-p)(114-570p+1332p^2-1908p^3+1849p^4-1263p^5+616p^6-213p^7+52p^8-9p^9+p^{10})}{(2-p)^3p^3(p^2-2p+2)(p^2+3-3p)^2}.$$
$$+\mathcal{O}(1).$$

We summarize our results in the following theorem

Theorem 4.1 *The first three moments of P in the geometric(p) case are given by*

$$M_P = (n-1)M + 2n + 2T_1 = \frac{-2+(-2-2p+2p^2)n}{p(-2+p)},$$
$$\mathbb{V}(P) = \frac{n[4(1-p)(p^4+9p^2-4p^3-10p+5)] + 4(3p^2-5p+5)(1-p)^2}{p^2(2-p)^2(p^2+3-3p)},$$
$$\mu_3(P) = n\frac{8(1-p)(114-570p+1332p^2-1908p^3+1849p^4-1263p^5+616p^6-213p^7+52p^8-9p^9+p^{10})}{(2-p)^3p^3(p^2-2p+2)(p^2+3-3p)^2}$$
$$+\mathcal{O}(1).$$

5 The stochastic processes

In this section, we analyze the stochastic processes related to P. Seen as a stochastic process, the random part of the perimeter is asymptotically given by $P_m(j) := \sum_i^j y_i$: we can ignore x_0, x_m and the contribution $2n$ is a constant. By the functional central limit theorem ([3, p. 174, Thm. 20.1]), we obtain the following result, where $B(t)$ is the standard Brownian Motion and $\Rightarrow$ denotes the weak convergence of random functions in the space of all right-continuous functions that have right limits and are endowed with the Skorohod metric (the φ-mixing property is immediate here: see ([3, p. 167, example 1])). This gives the limiting trajectories corresponding $P_m(j)$.

Theorem 5.1

$$\frac{P_m(\lfloor mt \rfloor) - Mmt}{\sigma\sqrt{m}} \Rightarrow B(t), \quad m \to \infty,\ t \in [0,1], \sigma = \sqrt{V^*}.$$

As a corollary, we have

Theorem 5.2

$$\frac{P_m - mM}{\sigma\sqrt{m}} \sim \mathcal{N}(0,1), m \to \infty,$$

where $\mathcal{N}$ is a Gaussian (normal) random variable.

In the uniform case, $k = 6$, we have made a simulation of $N = 100000$ trajectories $P_m(j), m = 500$. A typical trajectory is given in Fig.4, together with the drift jM. In Fig.5, we show a typical normalized

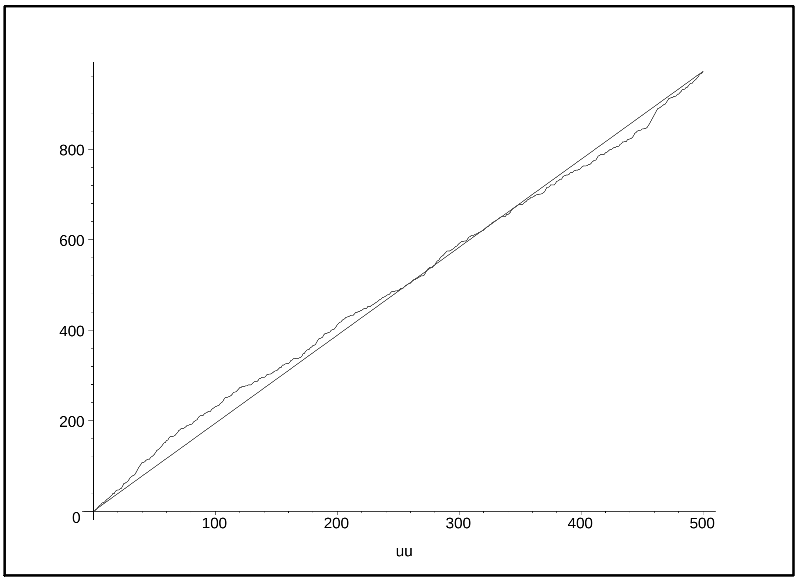

Figure 4: $P_m(j), m = 500$, drift$= jM$

trajectory

$$\frac{P_m(\lfloor mt \rfloor) - Mmt}{\sigma\sqrt{m}},$$

with the classical strongly irregular BM behaviour. We have also computed the observed moments: set z_ℓ:=the ℓth simulated value of $P_m(m) - mM$. We obtain

$$\left(\sum_1^N \frac{z_\ell}{\sigma\sqrt{m}}\right) \Big/ N = -0.0038\ldots, \left(\sum_1^N \left[\frac{z_\ell}{\sigma\sqrt{m}}\right]^2\right) \Big/ N = 0.991\ldots, \sum_1^N z_\ell^3 = 1287.47,$$

to be compared with the theoretical values $\{0, 1, m\mu_3^* = 1569.272976\ldots\}$. About the third moment, another simulation gives $1911.44\ldots$: m is not large enough to give a really good fit.

To illustrate Thm 5.2, we have build an histogram as follows: we construct a set of intervals $I(i) := [i\Delta - 3 - 3\Delta/2, i\Delta - 3 - \Delta/2], i = 0..6/\Delta + 2$, centered on $i\Delta - 3 - \Delta$ and covering the interval $[-3 - \Delta, 3 + \Delta]$. We choose here $\Delta = 1/2$. We define cells such that *cell*(i) corresponds to interval $I(i)$. We compute the number $N[i]$ of values of $\frac{z_\ell}{\sigma\sqrt{m}}$ falling into interval $I(i)$ and put $N(i)/N$ into *cell*(i) (values < 3.5 are attributed to *cell*(0) and similarly for values > 3.5). This gives the empirical

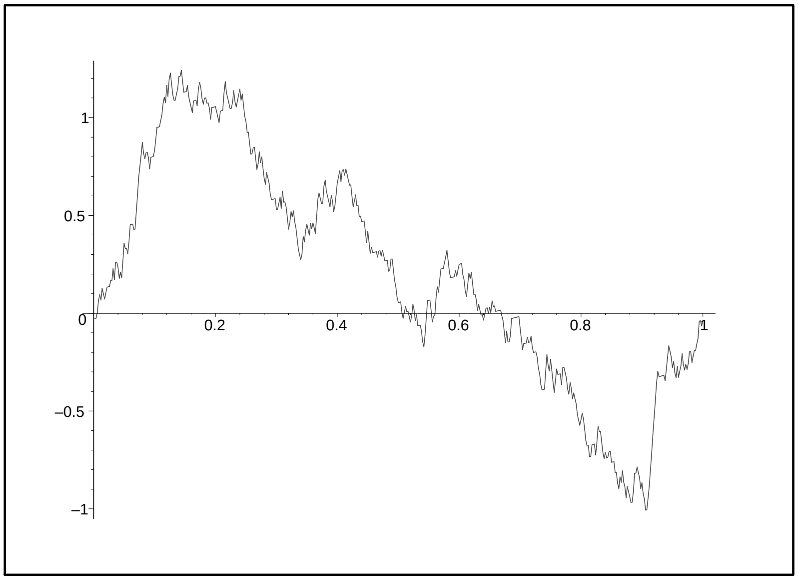

Figure 5: a typical normalized trajectory

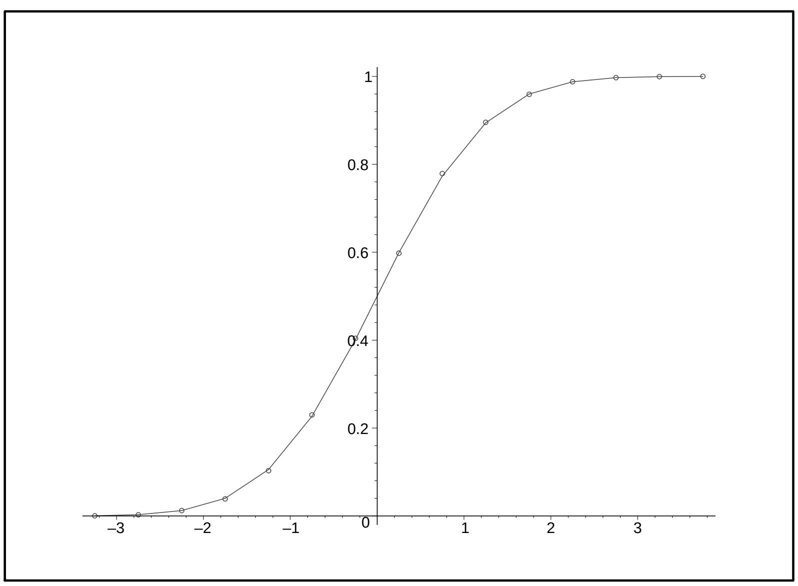

Figure 6: the cumulative histogram (circle) and the Gaussian distribution function (line)

histogram. In Fig.6, we compare the cumulative histogram (circle) with the Gaussian distribution function(line): the fit is quite good.

But it it still more precise to compare, in Fig.7 the histogram itself (circle) with the Gaussian probability mass in interval $I(i)$: $\int_{i\Delta-3-3\Delta/2}^{i\Delta-3-\Delta/2}(exp(-x^2/2)/\sqrt{2\pi}dx$ (line). The fit is quite satisfactory.

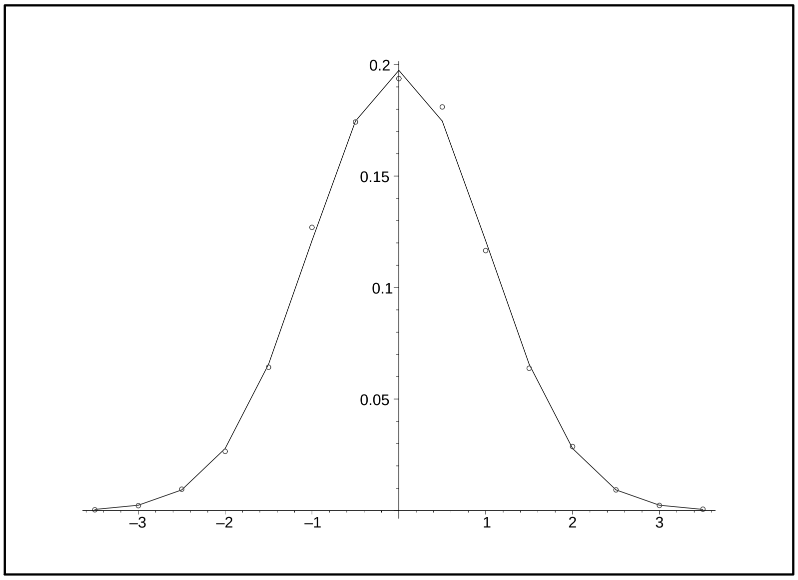

Figure 7: the histogram (circle) and the Gaussian probability mass in each interval $I(i)$ (line)

We have also made he same kind of simulations for he geometric(p) case. The results are quite similar.

6 Conclusion

We have shown that a probabilistic approach leads, almost mechanically, to the first three moments of P and its asymptotic Brownian and Gaussian properties. This technique can be applied to other moments and to other initial probability distributions.

References

[1] M. Archibald, A. Blecher, C. Brennan, A. Knopfmacher, and Helmut Prodinger. Geometric random variables: Descents following maxima. *Statistics and Probability Letters*, 124:140–147, 2017.

[2] M. Archibald and A. Knopfmacher. The largest missing value in a sample of geometric random variables. *Combinatorics, Probability and Computing*, 23(05):670–685, 2014.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[4] A. Blecher, C. Brennan, A. Knopmacher, and T. Mansour. The perimeter of words. *Discrete Mathematics*, 340:2456–2017, 2017.

[5] M. Bousquet-Mélou and A. Rechnitzer. The site-perimeter of bargraphs. *Advances in Applied Mathematics*, 31(1):86–112, 2003.

[6] Svjetlan Fereti. A perimeter enumeration of column-convex polyominoes. *Discrete Mathematics and Theoretical Computer Science*, 9:57–84, 2007.

[7] G. Louchard. Probabilistic analysis of some (un)directed animals. *Theoretical Computer Science*, 159(1):65–79, 1996.

[8] G. Louchard. Probabilistic analysis of column-convex and directed diagonally-convex animals. II: Trajectories and shapes. *Random Structures and Algorithms*, 15:1–23, 1999.

[9] G. Louchard. Runs of geometrically distributed random variables: a probabilistic analysis. *Journal of Computational and Applied Mathematics*, 142(1):137–153, 2002.

[10] G. Louchard. Monotone runs of uniformly distributed integer random variables: a probabilistic analysis. *Theoretical Computer Science*, 346(2-3):358–387, 2005.

[11] G. Louchard and H. Prodinger. Ascending runs of geometrically distributed random variables: a probabilistic analysis. *Theoretical Computer Science*, 304:59–86, 2003.

[12] G. Louchard and H. Prodinger. A combinatorial and probabilistic study of initial and end heights of descents in samples of geometrically distributed random variables and in permutations. *Discrete Mathematics and Theoretical Computer Science*, 9,1:137–170, 2005.

[13] G. Louchard and H. Prodinger. The number of elements close to near-records in geometric samples. *Quaestiones Mathematicae*, 29,4:447–470, 2006.