

Bounded discrete walks

C. Banderier¹ and P. Nicodème²

¹LIPN, CNRS-UMR 7030, Université Paris-Nord, 93430 Villetaneuse, France.

<http://www-lipn.univ-paris13.fr/~banderier/>

²LIX, CNRS-UMR 7161, École polytechnique, 91128 Palaiseau, France & AMIB project, INRIA Saclay.

<http://www.lix.polytechnique.fr/Labo/Pierre.Nicodeme/>

received 10 Mar 2010, revised 28th May 2010, accepted 10 May 2010.

This article tackles the enumeration and asymptotics of directed lattice paths (that are isomorphic to unidimensional paths) of bounded height (walks below one wall, or between two walls, for *any* finite set of jumps). Thus, for any lattice paths, we give the generating functions of bridges (“discrete” Brownian bridges) and reflected bridges (“discrete” reflected Brownian bridges) of a given height. It is a new success of the “kernel method” that the generating functions of such walks have some nice expressions as symmetric functions in terms of the roots of the kernel. These formulae also lead to fast algorithms for computing the n -th Taylor coefficients of the corresponding generating functions. For a large class of walks, we give the discrete distribution of the height of bridges, and show the convergence to a Rayleigh limit law. For the family of walks consisting of a -1 jump and many positive jumps, we give more precise bounds for the speed of convergence. We end our article with a heuristic application to bioinformatics that has a high speed-up relative to previous work.

Keywords: Random walks, analytic combinatorics, Brownian bridge, Rayleigh distribution, genes ranking statistics

1 Introduction

A lattice path is the drawing in \mathbb{Z}^2 of a sum of vectors from \mathbb{Z}^2 (where the vectors belongs to a finite fixed set, and where the origin of the path is usually taken as being the point $(0,0)$ from \mathbb{Z}^2). If all vectors are in $\mathbb{N} \times \mathbb{Z}$, the path is called directed (the path is going “to the right”). In this article, we consider the case for which all vectors are in $\{1\} \times \mathbb{Z}$; despite its natural representation as a drawing in \mathbb{Z}^2 , such a path is essentially a unidimensional object. If all vectors are in $\{1\} \times (\mathbb{N} \cup \{-1\})$, the path is called a “simple lattice path” or a “Łukasiewicz path” because there exists a bijection with simple families of trees. Unidimensional lattice paths pop up naturally in numerous fields (probability theory, combinatorics, algebra, economics, biology, analysis of algorithms, language theory, ...).

For lattice paths (or their probabilistic equivalent, random walks) which are space and time homogeneous, it was proven [4] that symbolic and analytic combinatorics [17] are quite powerful tools to study unidimensional lattice paths from an enumerative and asymptotic point of view. The authors of these articles developed a generating function approach to get exact enumeration of lattice paths (via the kernel method), and then used singularity analysis to study some basic parameters (like number of returns to

zero, final altitude) according to some constraints (drift/reflecting conditions). For random walks there are numerous works using an approach with a flavour of “generating function” and “analysis of singularity”, either with probabilistic or combinatorial methods, e.g. for interaction of random walks, for random walks on groups or for stationary distributions of 2-dimensional models in queueing theory.

In our simpler model of unidimensional lattice paths, for simple parameters which are in one sense “exactly solvable”, one can expect more than for the above difficult problems: not only can one get here critical exponents, but also we get fast computation schemes, for exact enumeration and for full asymptotics (to any order). It is then natural to ask what can be obtained for less trivial parameters, like the area under the walks or the height of the walk. The area was investigated in another article [6], and we concentrate here on the height. The height has already been investigated in combinatorics, mainly for Dyck paths (walks having just $+1$ or -1 jumps) which have a nice relationship with continued fractions and Chebyshev polynomials.

In our approach, one of the key tricks to solve the model for any set of jumps is the so-called “kernel method”, which is a way of solving a functional equation of the type $K(z, u)F(z, u) = A(z, u) + B(z, u)G(z)$ where F and G are the unknowns one wishes to determine. The kernel method consists in getting additional equations by plugging the roots $u(z)$ of the “kernel” $K(z, u)$ in the initial equation, which in general is enough to solve the system. The kernel method shares the spirit of the “quadratic method” of Tutte and Brown (for enumeration of maps). In combinatorics, only the simplest case of the kernel method (namely, when there is only one root) was used for 30 years, see Knuth [22] for sorting with stacks, Chung et al. [11] for a pebbling game, or [25] for generation of binary trees. During the same time period, and independently, difficult 2-dimensional generalisations of this method were well studied in queueing theory; the classification of the different cases for the nearest neighbour walks in \mathbb{N}^2 was already quite a challenge, see the book [15] or [24]. This last decade, there has been a revival in combinatorics for functional equations, and the full power of the kernel method was better put into evidence, both for enumeration and for asymptotics. Solving equations is not the only miracle that the kernel method offers, it also gives compact formulae [9, 1, 3], offers some nice interactions with computer algebra (see the recent solution of Gessel’s conjecture [20, 7]), and gives access to asymptotics [4, 2, 26]. Our article thus adds a new stone to the “kernel method” edifice, and gives more results on enumeration and asymptotics for the height of directed lattice paths, while other approaches (like e.g. transfer matrices or probability theory) would be either of much higher complexity, or would not give access to exact enumeration or to higher order asymptotics.

2 Preliminaries and previous results

As stated previously, we consider in this article directed walks with jumps in $\{1\} \times \mathbb{J}$ where \mathbb{J} is any finite subset of \mathbb{Z} . In this context, the altitude of the walk at time n is the sum of n increments taken in \mathbb{J} . Our main tool is the generating function $F(z, u) = \sum_{k=0}^{\infty} F_k(z)u^k$, where $F_k(z)$ is the generating function the n -th coefficient of which is the number of walks of length n ending at altitude k . To each walk, we associate a Laurent polynomial which encodes all the possible jumps $P(u) := \sum_{i=-c}^d p_i u^i$ (where c is the size of the largest backward jump and d is the size of the largest upward jump, and where the p_i ’s are some “weights”, “multiplicities”, or “probabilities”).

Figure 1 shows four drawings (for four different constraints) of lattice paths with jumps in $\mathbb{J} = \{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$; the associated Laurent polynomial is therefore $P(u) = p_{-3}u^{-3} + p_{-2}u^{-2} + p_{-1}u^{-1} + p_0 + p_1u + p_2u^2 + p_3u^3 + p_4u^4 + p_5u^5$.

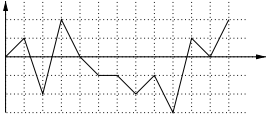
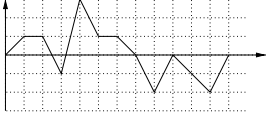
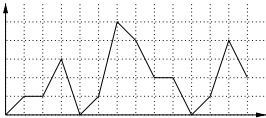
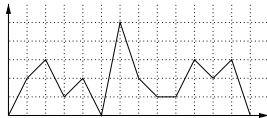
	ending anywhere	ending at 0
unconstrained (on \mathbb{Z})	 <p>walk/path (\mathcal{W})</p> $W(z) = \frac{1}{1 - zP(1)}$ $W_n = P(1)^n$	 <p>bridge (\mathcal{B})</p> $B(z) = z \sum_{i=1}^c \frac{u_i'(z)}{u_i(z)}$ $B_n \sim \beta_0 \frac{P(\tau)^n}{\sqrt{\pi n}}$
constrained (on $\mathbb{Z}_{\geq 0}$)	 <p>meander (\mathcal{M})</p> $M(z) = \frac{1}{1 - zP(1)} \prod_{i=1}^c (1 - u_i(z))$ $M_n \sim \mu_0 \frac{P(1)^n}{\sqrt{\pi n}} \text{ (zero drift)}$ $M_n \sim \mu_0^- \frac{P(\tau)^n}{\sqrt{\pi n^3}} \text{ (negative drift)}$ $M_n \sim \mu_0^+ P(1)^n + \mu_0^- \frac{P(\tau)^n}{\sqrt{\pi n^3}} \text{ (positive drift)}$	 <p>excursion (\mathcal{E})</p> $E(z) = \frac{(-1)^{c-1}}{p - cz} \prod_{i=1}^c u_i(z)$ $E_n \sim \varepsilon_0 \frac{P(\tau)^n}{\sqrt{\pi n^3}}$

Fig. 1: The four types of paths: walks, bridges, meanders, and excursions and the corresponding generating functions. The u_i 's are such that $1 - zP(u_i(z)) = 0$; the constants ε_0, β_0 , and the μ_0 's are algebraic numbers which are made explicit in [4]. In the rest of this article, we want to consider walks of bounded height (i.e., between one or two walls).

2.1 Some asymptotics results for walks

In [4], Banderier and Flajolet showed that the kernel method was the key to get enumeration and asymptotics of directed lattice paths. The proofs rely on the following properties:

Property 1. There are c distinct roots u_1, \dots, u_c of the “kernel” $1 - zP(u) = 0$ which behave like $z^{1/c}$ at $z = 0$ (we call them “small roots”). The remaining d roots v_1, \dots, v_d behave like $z^{-1/d}$ at $z = 0$ (we call them “large roots”).

Property 2. A nice trick (the “kernel method”) allows to write all the generating functions counting walks as functions of the u_i 's (and/or of the v_j 's).

Property 3. There is a unique positive real number τ such that $P'(\tau) = 0$, and the radius of convergence of the generating functions is $\rho := 1/P(\tau)$.

Property 4. There is a dominant small root $u_1(z)$ and a dominant large root $v_1(z)$ such that for $0 < |z| < \rho$, one has $|u_i(z)| < u_1(|z|) < v_1(|z|) < |v_j(z)|$ for all non dominant roots $u_i(z)$ and $v_j(z)$.

Property 5. The asymptotics are coming from the dominant roots u_1 and v_1 , which are singular at ρ , the product of the other roots is analytical at ρ and therefore, these other roots only affect the multiplicative constant. (Some easy modifications have to be made here if the walk is “periodic”).

Property 6. The roots u_1 and v_1 share a square root behaviour at $z = \rho$:

$$u_1(z) \sim \tau + \sum_{i \geq 1} (-1)^i a_i (\rho - z)^{i/2} \quad v_1(z) \sim \tau + \sum_{i \geq 1} a_i (\rho - z)^{i/2}. \quad (1)$$

Property 7. The drift $\delta := P'(1)$ of the walk plays a rôle for the asymptotics of meanders ($\delta > 0$ when $\tau > 1$; $\delta = 0$ when $\tau = 1$; $\delta < 0$ when $\tau < 1$).

2.2 Walks on \mathbb{Z} ending at altitude k

In the next section, we will use a part of Theorem 1 in [4]:

Theorem 1 *The generating function⁽ⁱ⁾ $W_k(z) = [u^k] \frac{1}{1 - zP(u)}$ of paths starting at 0 having no bounding constraints and ending at altitude k is*

$$W_k(z) = z \sum_{j=1}^c \frac{u'_j(z)}{u_j(z)^{k+1}} \quad (\text{for } k < c) \quad \text{and} \quad W_k(z) = -z \sum_{j=1}^d \frac{v'_j(z)}{v_j(z)^{k+1}} \quad (\text{for } k > -d), \quad (2)$$

where u_1, \dots, u_c are the small roots and v_1, \dots, v_d are the large roots.

3 Generating functions of bounded walks

This section is dedicated to finding the explicit generating functions of numerous variants of bounded lattice paths: below one wall, between two walls.

Theorem 2 *Walks of bounded height (i.e. in the strip $[-\infty, +h]$) The bivariate generating function of walks starting at 0, and remaining below the wall $y = h$ is algebraic:*

$$F^{[-\infty, h]}(z, u) = \frac{1 - \sum_{i=1}^d \left(\frac{u}{v_i}\right)^{h+1} \prod_{1 \leq j \leq d, j \neq i} \frac{u - v_j}{v_i - v_j}}{1 - zP(u)}. \quad (3)$$

Proof: Using a step by step decomposition of the walks gives the following functional equation:

$$F^{[-\infty, h]}(z, u) = 1 + zP(u)F^{[-\infty, h]}(z, u) - z\{u > h\}P(u)F^{[-\infty, h]}(z, u).$$

This theorem is then a consequence of the kernel method applied to

$$(1 - zP(u))F^{[-\infty, h]}(z, u) = 1 - z \sum_{k=0}^{d-1} r_{h-k}(u) u^{h-k} F_{h-k}^{[-\infty, h]}(z)$$

where $r_{h-k}(u) := \sum_{i=k+1}^d p_i u^i$.

⁽ⁱ⁾ Throughout this article, we use the notation $[u^k]F(z, u)$ for the coefficient of u^k in $F(z, u)$, and $\{u^k\}F(z, u)$ stands for the monomial in u^k of $F(z, u)$. We will write $F_k(z)$ for $[u^k]F(z, u)$, and, for any interval $[a, b]$, $F^{[a, b]}$ will correspond to a generating function of walks restricted to this interval.

This functional equation follows from the fact that a walk is either an empty walk, either it makes a new jump (i.e., we multiply by $zP(u)$ and, since the walk is constrained to remain under altitude h , we remove all jumps which sent us above the border h ; these are monomials in u^{h+1}, \dots, u^{h+d}); plugging the d large roots v_i in this equation (this is a legitimate operation as the v_i behaves like $z^{-1/d}$ and as $F^{[-\infty, h]}$ is a Laurent power series in $1/u$) leads to a linear system of d equations with d unknowns. Solving it with the Cramer formula provides expressions involving Vandermonde-like determinants

$$\mathbb{V}_i = \prod_{r=1}^{d-1} \prod_{s=r+1}^d (v_r(z) - v_s(z)) \Big|_{v_i=u}, \quad (4)$$

or more precisely, quotients of the type $\mathbb{V}_i/\mathbb{V}_0 = Q(u)/Q(v_i)$ where

$$Q(u) = \prod_{2 \leq j \leq d} (u - v_j(z)) = \sum_{i=0}^{d-1} q_i(z) u^i. \quad (5)$$

Further simplifications lead to Theorem 2. \square

Theorem 3 *The generating function of walks of bounded height ending at altitude k is*

$$[u^k] F^{[-\infty, h]}(z, u) = W_k(z) - \sum_{m=0}^{d-1} W_{k-h-1-m}(z) \sum_{i=1}^d \frac{(-1)^{m+d-1} e_{d-1-m}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_d)}{v_i^{h+1} \prod_{j=1, j \neq i}^d (v_i - v_j)}$$

where the e_i 's are elementary symmetric functions defined by Equation 7 and the functions $W_k(z)$ are defined in Theorem 1.

Proof: Extracting the coefficient of u^k in Equation (3) of Theorem 2 gives

$$\begin{aligned} [u^k] F^{[-\infty, h]}(z, u) &= [u^k] \frac{1}{1 - zP(u)} \left(1 - \sum_{i=1}^d \left(\frac{u}{v_i} \right)^{h+1} \prod_{j=1, j \neq i}^d \left(\frac{u - v_j}{v_i - v_j} \right) \right) \\ &= W_k(z) - [u^{k-h-1}] \frac{1}{1 - zP(u)} \sum_{i=1}^d \left(\left(\frac{1}{v_i} \right)^{h+1} \sum_{m=0}^{d-1} u^m \frac{[u^m] \prod_{j=1, j \neq i}^d (u - v_j)}{\prod_{j=1, j \neq i}^d (v_i - v_j)} \right). \end{aligned} \quad (6)$$

For any set of variables \mathcal{X} , let $\mathcal{X}_m := \{\text{subsets of size } m \text{ of } \mathcal{X}\}$; we then introduce the elementary symmetric function e_m :

$$e_m(\mathcal{X}) := \sum_{X \in \mathcal{X}_m} \prod_{x \in X} x.$$

E.g., $e_0 = 1$, $e_2(v_1, v_3, v_4) = v_1 v_3 + v_1 v_4 + v_3 v_4$, $e_3(v_1, v_3, v_4) = v_1 v_3 v_4$, etc. We thus have

$$[u^m] \prod_{j=1, j \neq i}^d (u - v_j) = (-1)^{m+d-1} e_{d-1-m}(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_d). \quad (7)$$

Using these e_i 's and inverting the two sums in Equation (6) leads to the theorem. (Note that using Theorem 1 allows to express all the functions $W_k(z)$ in terms of the roots.) \square

Theorem 4 Walks in a strip $[-h_2, +h]$. The bivariate generating function of walks starting at 0, and remaining between the walls $y = -h_2$ and $y = h$ is rational:

$$F^{[-h_2, h]}(z, u) = \frac{1 - z \sum_{k=0}^{c-1} r_{-h_2+k}(u) u^{-h_2+k} F_{-h_2+k}^{[-h_2, h]}(z) - z \sum_{k=0}^{d-1} r_{h-k}(u) u^{h-k} F_{h-k}^{[-h_2, h]}(z)}{1 - zP(u)}$$

with

$$F_{-h_2+k}^{[-h_2, h]}(z) = \frac{1}{p_d z} \frac{s_{\lambda'_k}}{s_{\lambda}} \quad \text{and} \quad F_{h-k}^{[-h_2, h]}(z) = \frac{1}{p_d z} \frac{s_{\lambda''_k}}{s_{\lambda}}$$

where the s_{λ} 's are defined in Equations (9,10,11,12) and where the r_j 's are defined in Equation (8).

Proof: The polynomials r_k just encode the forbidden jumps when we are close to the walls:

$$r_{h-k}(u) := \sum_{j=k+1}^d p_j u^j \quad \text{for } 0 \leq k \leq d-1 \quad \text{and} \quad r_{-h_2+k}(u) := \sum_{j=k+1}^c p_{-j} u^{-j} \quad \text{for } 0 \leq k \leq c-1. \quad (8)$$

We have therefore the following functional equation:

$$F^{[-h_2, h]}(z, u) = 1 + zP(u)F^{[-h_2, h]}(z, u) - z \sum_{k=0}^{d-1} r_{h-k}(u) u^{h-k} F_{h-k}^{[-h_2, h]}(z) - z \sum_{k=0}^{c-1} r_{-h_2+k}(u) u^{-h_2+k} F_{-h_2+k}^{[-h_2, h]}(z).$$

Substituting u by the roots u_i 's and v_i 's in this equation (this is legitimate since F is here a Laurent polynomial in u , and not a Laurent series) leads to a system with $c + d$ unknowns; the Cramer formula gives the solutions of the system in terms of more complicated Vandermonde-like determinants, for which we need here to introduce a few new notations. Let S_{c+d} be the group of permutations of $c + d$ elements, and let $|\sigma|$ be the signature of any permutation σ . For any integer partition $\lambda = (\lambda_1, \dots, \lambda_{c+d})$ with $c + d$ (possibly empty) parts, let the antisymmetric associated Schur function be (we note $v_i = u_{i+c}$):

$$s_{\lambda} := \sum_{\sigma \in S_{c+d}} (-1)^{|\sigma|} \prod_{i=1}^{c+d} u_i^{\lambda_i}. \quad (9)$$

In the Vandermonde-like determinants, after suitable multiplications to get polynomials instead of Laurent polynomials, the exponents of the u_i 's are:

$$\lambda := (0, 1, \dots, c-1, h+h_2+c+1, \dots, h+h_2+c+d), \quad (10)$$

$$\lambda'_k := (0, 1, \dots, c-1, h_2+c, h+h_2+c+1, \dots, h+h_2+c+d) \setminus \{k\}, \quad (11)$$

$$\lambda''_k := (1, \dots, c-1, h_2+c, h+h_2+c+1, \dots, h+h_2+c+d) \setminus \{h+h_2+c+d-k\}. \quad (12)$$

A simplification of the Cramer formula then leads to the expressions for the $F_{-h_2+k}^{[-h_2,h],s}$ and $F_{h-k}^{[-h_2,h],s}$ given in the Theorem. \square

Using Theorem 4, one can obtain the generating function of walks ending at altitude k (and therefore the generating function of bridges between 2 walls). Theorem 4 can be specified to the case $h_2 = 0$. This gives a formula for bounded excursions, already studied with this approach in [1]. Further links with Schur functions, the Jacobi–Trudi identity, and the platypus algorithm⁽ⁱⁱ⁾ were investigated in [8].

In the full version of this article, we will use our formula to answer the following question (which comes from bioinformatics): what is the distribution of the maximum of the absolute value of some random discrete bridges? This case can be seen as a discrete equivalent of “reflected Brownian bridges” (or absolute value of the Brownian), continuous objects for which results exist [14]. In the discrete case, this corresponds to another specialisation of Theorem 4 above, via the case $h_2 = h$.

4 Asymptotic limit laws

Due to numerous works making the link between discrete random walks and Brownian motion variants [13, 19, 18], and in particular the Komlós–Major–Tusnády strong embedding result [23] (see Chatterjee [10] for a modern approach), it is expected that the limit law for the height of bridges satisfies a Rayleigh distribution. Although we do not investigate the case here, the limit law for the height of discrete excursions should satisfy a Theta distribution (note that the height of excursions is related to the height of some trees [12, 16]).

From a probabilistic point of view, our work on bridges brings new results on the speed of convergence (exact error terms). From an analytic point of view, it also describes the exact nature of the corresponding singularities.

4.1 Convergence to the Rayleigh law

Definition 1 *The Rayleigh distribution of parameter α is defined as a distribution which has support $[0, +\infty]$ with a density given by $\mathcal{R}_\alpha(x) = \frac{x}{\alpha} \exp(-\frac{x^2}{2\alpha^2})$. Accordingly, its cumulative distribution function is $\int_0^x \mathcal{R}_\alpha(t)dt = 1 - \exp(-\frac{x^2}{2\alpha^2})$.*

Theorem 5 *The height of discrete bridges satisfies a Rayleigh limit law of parameter $\frac{\tau\sqrt{P''(\tau)}}{2\sqrt{\rho}}$:*

$$\Pr\left(\max_{0 \leq k \leq n} B_k > x\sqrt{n}\right) = e^{-\frac{2\rho x^2}{P''(\tau)\tau^2}} + O\left(\frac{1}{\sqrt{n}}\right).$$

This shows that whatever the drift is, we still get a Rayleigh limit law. This result is also coherent with the symmetry induced by “reversing the time”, i.e., $P(u) \leftrightarrow P(1/u)$.

The proof follows the same lines as the proof of Theorem 6 hereafter; indeed, to any set of jumps, one can associate a probabilistic model with drift 0 which can be done by Cramér’s method of shifting the mean:

⁽ⁱⁱ⁾ The Platypus algorithm presented in [4] allows to get the minimal polynomial of our algebraic generating functions, and therefore, going to the associated differential equation, it allows to compute the number of walks of length n with $O(\sqrt{n})$ operations. In the case of two boundaries, we are in the rational case; using a variant of binary exponentiation when taking powers of matrices leads to a complexity $O((h+h_2)^\varpi \log n)$, where $\varpi \approx 2.38$ is an improvement of Strassen’s constant for matrix multiplication. This can be improved using D-finiteness of our generating functions, and then recurrences of order less than $h+h_2$.

define $\tilde{P}(u) := P(ux)/P(x)$ for a x such that $\tilde{P}'(1) = 0$. An elementary computation shows that x is then exactly τ , the unique real positive saddle point of $P(u)$. We therefore consider hereafter the case of drift 0 without loss of generality.

We prove in the next theorem that under the usual normalisation, the height of the discrete bridge (built upon jumps with expectation zero, corresponding to walks without drift) converges asymptotically weakly to the height of the standard Brownian motion. We also prove that the speed of convergence is $O(1/\sqrt{n})$ and provide an algorithm computing asymptotically the distribution function of the height of a discrete bridge at any order in the case of Łukasiewicz walks (one downward jump -1).

Theorem 6 *Considering a discrete bridge $(B_k)_{k=0,\dots,n}$ with zero drift (i.e. $P'(1) = 0$) and variance σ^2 (i.e. $P''(1) = \sigma^2$), we have the following properties:*

1. *For $x \in]0, +\infty[$, we have a Rayleigh limit law of parameter $1/2$:*

$$\lim_{n \rightarrow \infty} \Pr \left(\max_{0 \leq k \leq n} B_k > x \times \sigma \sqrt{n} \right) = e^{-2x^2} \times \left(1 + O \left(\frac{1}{\sqrt{n}} \right) \right). \quad (13)$$

2. *The density of the height of the bridge is concentrated around $\sigma \sqrt{n}$:*

$$\lim_{n \rightarrow \infty} \Pr \left(\sigma \sqrt{n} \times \frac{1}{n^\varepsilon} < \max_{0 \leq k \leq n} B_k < \sigma \sqrt{n}(1 + \eta_n) \right) = 1, \quad (14)$$

for any $\varepsilon > 0$ and any sequence $\eta_n = o(\sqrt{n})$ tending to infinity with n .

Proof (sketch): We consider the non-periodic case (the proof for the periodic case follows the same lines) for which we compute the probability that a bridge goes beyond an upper barrier at level $x\sigma\sqrt{n}$ for a fixed $x \in]0, +\infty[$. Using the product Q defined in Equation 5, Theorem 2 can be rewritten as:

$$F^{[>h]}(z, u) = \frac{\sum_{i=1}^d \left(\frac{u}{v_i} \right)^{h+1} \frac{Q(u)}{Q(v_i)}}{1 - zP(u)}.$$

As $\left(\frac{u}{v_j} \right)^h = \left(\frac{u}{v_1} \right)^h \times \left(\frac{v_1}{v_j} \right)^h = \left(\frac{u}{v_1} \right)^h \times O(A^h)$ where $A := \max_{2 \leq j \leq d} \max_{|z| \leq \rho} \frac{|v_1(z)|}{|v_j(z)|}$ and $A < 1$ by the domination property (see Section 2), the main asymptotics contribution comes from the summand for $i = 1$:

$$\begin{aligned} F^{[>h]}(z, u) &= \frac{1}{1 - zP(u)} \left(\frac{u}{v_1} \right)^{h+1} \frac{Q(u)}{Q(v_1)} \times (1 + O(A^h)) \\ &= \frac{1}{Q(v_1)v_1^{h+1}} \sum_{i=0}^{d-1} q_i \frac{u^{h+1}u^i}{1 - zP(u)} \times (1 + O(A^h)). \end{aligned}$$

The last line just uses the rewriting $Q(u) = \sum q_i u^i$. Taking the coefficient $[u^0]$ of $F^{[>h]}$ in order to get

bridges and using the functions $W_k(z)$ introduced in Theorem 1 for walks ending at altitude k leads to

$$\begin{aligned} [u^0]F^{[>h]}(z, u) &= \frac{1}{Q(v_1)v_1^{h+1}} \sum_{i=0}^{d-1} q_i W_{-h-1-i}(z) \times (1 + O(A^h)) \\ &= \frac{1}{Q(v_1)v_1^{h+1}} \sum_{i=0}^{d-1} q_i z \sum_{k=1}^c u_k^h \frac{u'_k}{u_k^{-i}} \times (1 + O(A^h)) \\ &= \frac{zu_1^h}{Q(v_1)v_1^{h+1}} \sum_{i=0}^{d-1} q_i \sum_{k=1}^c \left(\frac{u_k}{u_1} \right)^h \frac{u'_k}{u_k^{-i}} \times (1 + O(A^h)). \end{aligned}$$

Using now $B := \max_{2 \leq j \leq c} \max_{|z| \leq \rho} \frac{|u_j(z)|}{|u_1(z)|} < 1$, the main contribution comes from the summand for $k = 1$:

$$[u^0]F^{[>h]}(z, u) = \frac{zu_1^h}{Q(v_1)v_1^{h+1}} \sum_{i=0}^{d-1} q_i u_1^i \times (1 + O(A^h)) \times (1 + O(B^h)).$$

Using the product Q defined in Equation 5, we have for $C := \max(A, B) < 1$:

$$[u^0]F^{[>h]}(z, u) = z \left(\frac{u_1(z)}{v_1(z)} \right)^h \times \frac{u'_1(z)Q(u_1(z))}{v_1(z)Q(v_1(z))} \times (1 + O(C^h)). \quad (15)$$

Singularity analysis. In our zero drift probabilistic setup, Property 6 from Section 2 can be rewritten using $\rho = \tau = 1$, $P(1) = 1$, $a_1 = \frac{\sqrt{2}}{\sqrt{P''(1)}}$; this leads to

$$\left(\frac{u_1}{v_1} \right)^h = \left(1 - 2 \frac{\sqrt{2}}{\sqrt{P''(1)}} \sqrt{1-z} \right)^h \times (1 + O(1-z))^h \quad \text{for } z \sim 1^-. \quad (16)$$

Due to the organisation in cycles of the roots (in conjugated tuples of branches at their singularities), the sum and product of the v_i 's ($i > 1$) is regular at 1, and therefore so are all the q_i 's, thus one has:

$$\frac{Q(u_1(z))}{Q(v_1(z))} = \frac{Q(1) + O(\sqrt{1-z})}{Q(1) + O(\sqrt{1-z})} = 1 + O(\sqrt{1-z}). \quad (17)$$

Asymptotic expansions. We compute now asymptotically $b_n^{>h} = [z^n]F_0^{[>h]}(z)$ when $h = x\sigma\sqrt{n}$. The singularity of $F_0^{[>h]}(z)$ is $z = 1$, as seen in Equation (15) and (16). We compute $b_n^{>h}$ by a Cauchy integral,

$$b_n^{>h} = \frac{1}{2i\pi} \oint_{\Gamma} \frac{F_0^{[>h]}(z)}{z^{n+1}} dz, \quad (18)$$

where Γ is a Hankel contour winding around the point $z = 1$.

We follow here the proof from [5] for the limit distribution of semi-large powers and apply also the Big-Oh transfer theorem (see Theorem VI.3 and IX.16 of [17]; note that there exists a suitable Δ -domain

in which the domination property of the roots holds, as we deal with an algebraic function, which has therefore isolated singularities). Using the change of variable $z = 1 - \frac{t}{n}$, we obtain

$$b_n^{>x\sigma\sqrt{n}} = \frac{1}{2i\pi} \oint_{\Gamma'} \frac{1}{\sigma\sqrt{n}} \frac{e^t e^{-2x\sqrt{2t}}}{\sqrt{t}} \times \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) dt, \quad (19)$$

where Γ' is a Hankel contour winding clockwise from $-\infty$ around the origin. Expanding $e^{-2x\sqrt{2t}}$, making the substitution $t \rightsquigarrow -t$, integrating term-wise, and using the Hankel contour representation⁽ⁱⁱⁱ⁾ for the Gamma function,

$$\frac{1}{\pi} \sin(\pi s) \Gamma(1-s) = -\frac{1}{2i\pi} \int_{+\infty}^{(0)} (-t)^s e^{-t} dt, \quad \text{for all } s \in \mathbb{C}, \quad (20)$$

we obtain after simplification as n tends to infinity (since an unbounded weighted walk of length n is a bridge with probability $b_n^{<\infty} = \frac{1}{\sigma\sqrt{n}}(1 + O(\frac{1}{\sqrt{n}}))$, see Theorem 3 from [4]):

$$\beta_n^{>x\sigma\sqrt{n}} = b_n^{>x\sigma\sqrt{n}} / b_n^{<\infty} \quad (21)$$

$$= -\frac{1}{2i\pi} \sum_{j=0}^{\infty} (-1)^j \frac{2^j (\sqrt{2}x)^j}{j!} \int_{+\infty}^{(0)} e^{(-t)} (-t)^{(j-1)/2} dt \times \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \quad (22)$$

$$= \sum_{k=0}^{\infty} (-1)^k \frac{(\sqrt{2}x)^{2k}}{k!} \times \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) = e^{-2x^2} \times \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \quad (23)$$

where $\beta_n^{>x\sigma\sqrt{n}}$ (i.e. the probability that a bridge of length n goes upon the barrier $y = x\sigma\sqrt{n}$) thus follows a Rayleigh limit law for $x \in]0, +\infty[$. This concludes the first part of our proof.

The proof of the concentration property follows the same lines as the proof for $x \in]0, \infty[$ with different error terms.

This concludes our proof in the case of non-periodic walks.

In the case of walks with period d , there are d singularities on the circle of convergence at positions ω^j where ω is the primitive d -th root of 1. For unconditioned walks, this corresponds to d saddle-points on the circle $|z| = 1$. In the case of upper-bounded walks, we use a star-shape Cauchy contour that is deformed up to d Hankel contours, the j th contour coming from $\omega^j \times (+\infty)$ winding around the point ω^j and coming back to $\omega^j \times (+\infty)$. In both cases, there is a multiplicative factor d occurring in $b_n^{>x\sigma\sqrt{n}}$ and in $b_n^{<\infty}$; this factor cancels when taking the ratio of the two quantities. \square

4.2 Łukasiewicz bridges

When considering the case of Łukasiewicz walks (i.e. $c = 1$, see the Section 1), we obtain more precise asymptotics for the convergence to the Rayleigh law. Indeed, starting from Equation 5, one has:

$$Q(u_1(z)) = \frac{1}{p_d z} \frac{\partial}{\partial u} \frac{u(1-zP(u))}{u-v_1(z)} \Big|_{u=u_1(z)} = \frac{1}{p_d z^2} \frac{u_1(z)}{u_1'(z)(u_1(z)-v_1(z))}. \quad (24)$$

⁽ⁱⁱⁱ⁾ See a proof of this representation in Flajolet-Sedgewick [17], p. 745.

The value of $Q(v_1)$ follows by interchanging the rôles of u_1 and v_1 . Equation (15) thus becomes

$$[u^0]F^{[>h]}(z, u) = z \left(\frac{u_1(z)}{v_1(z)} \right)^h \times \frac{-v_1'(z)u_1(z)}{v_1(z)^2} \times (1 + O(A^h)). \quad (25)$$

By following the same steps as in Section 4.1 but at a higher asymptotic order, we find, with $a_1 = \sqrt{2/P''(1)}$, that

$$\beta_n^{>x\sqrt{n}} = \exp\left(-\frac{\rho x^2 a_1^2}{\tau^2}\right) + O\left(\frac{1}{\sqrt{n}}\right) \quad (26)$$

in which we recognise the Rayleigh distribution of parameter $\sigma = \frac{\tau}{a_1\sqrt{2\rho}}$. In the probabilistic setting (with zero drift), we give in the next formula more precise error terms:

$$\begin{aligned} \frac{\beta_n^{>x\sigma\sqrt{n}}}{\exp(-2x^2)} &= 1 + \frac{(-2/3)x\xi/\zeta^{3/2} - 6x/\sqrt{\zeta}}{\sqrt{n}} + \frac{1}{n} \left((-2 - \frac{10}{9}\frac{\xi^2}{\zeta^3} + \frac{2}{3}\frac{\theta}{\zeta^2} - \frac{16}{3\zeta} - \frac{8}{3}\frac{\xi}{\zeta^2})x^4 \right. \\ &\quad \left. + (\frac{24}{\zeta} + \frac{5}{3}\frac{\xi^2}{\zeta^3} + 3 - \frac{\theta}{\zeta^2} + \frac{20}{3}\frac{\xi}{\zeta^2})x^2 - \frac{5}{\zeta} - \frac{3}{8} - \frac{7}{6}\frac{\xi}{\zeta^2} - \frac{5}{24}\frac{\xi^2}{\zeta^3} + \frac{1}{8}\frac{\theta}{\zeta^2} + \frac{5}{24}\frac{\xi^3}{\zeta^3} - \frac{1}{8}\frac{\theta^2 - 3\zeta^2}{\zeta^2} \right) \\ &\quad + O\left(\frac{1}{n^{3/2}}\right), \end{aligned} \quad (27)$$

where $\zeta = \sigma^2 = P''(1)$, $\xi = P'''(1)$ and $\theta = P''''(1)$. The *algolib* Maple package (more precisely, the *gdev* and *equivalent* functions developed by Bruno Salvy, see algo.inria.fr/librairies) can naturally push the expansion to higher orders.

5 Bioinformatics application

A set of G genes is expressed in a given tissue; this provides a ranking of level of expression of these genes. Considering now the same ranking and a subset of specific interest of g genes, the question is how to characterise if the ranks of these g genes form an unexpected pattern. If these g genes have a high level of expression, they will mostly appear at the top of the ranking; on the contrary, if they have a low level of expression, they will mostly appear at the bottom of the ranking. Both cases are assumed to correspond to biological disorders. The aim is to provide a statistical estimator for exceptional behaviours. Keller *et al.* [21] proposed the following non-parametric approach. While scanning from left to right the ranking of the G genes, build a random walk $(B_i)_{0 \leq i \leq G}$ such that its jump at time i is

$$\begin{cases} G - g & \text{if the gene at rank } i \text{ belongs to } g, \\ -g & \text{if the gene at rank } i \text{ belongs to } G - g. \end{cases}$$

By construction, we have $B_0 = B_G = 0$ and these walks are therefore bridges. The tail probability (commonly referred to as p -value in applied mathematics) that Keller *et al.* choose as statistical indicator is $p\text{-value} = \Pr(\max_{1 \leq i \leq G} |B_i| > h)$, for any chosen h . They provide a dynamic programming algorithm computing this indicator in complexity $O(G \times g)$. We consider in this extended abstract the case of over-expressed genes which corresponds to the maximum of the bridges. If $(G - g)/g$ is entire, we are in a particular case of Łukasiewicz walk where $P(u) = \frac{u^d}{d+1} + \frac{d}{d+1} \frac{1}{u}$ and $d = \frac{G-g}{g}$. This gives

$P''(1) = d, P'''(1) = d(d-4), \dots$; substituting these values in Equation (27) leads to an efficient formula (which has a constant time complexity in $n, h,$ and d).

Heuristics. We expect that using Equation (27) with a fractional $d = (G - g)/g$ will give a good approximation of the probability that the height of a walk $(+(G - g), -g)$ goes upon a given value, although we cannot prove this yet.

Simulations. In order to get enough precision, we apply here Equation (27), with $d = 3$ and $d = 19$. We plot the obtained distributions in Figure 2. The simulations are done by drawing random non-conditional walks of length n and then by reversing jumps at uniformly chosen random positions until we obtain bridges. We deal here with periodic walks with 2 jumps $(+d$ and $-c)$; for such walks, bridges have length $n = 0 \pmod{c+d}$ and each reversing leads to a walk ending nearer and nearer from 0, while still ending in the same half-plane. This explains why our algorithm takes on average $O(\sqrt{n})$ iterations since a walk of length n ends typically at altitude $O(\sqrt{n})$.

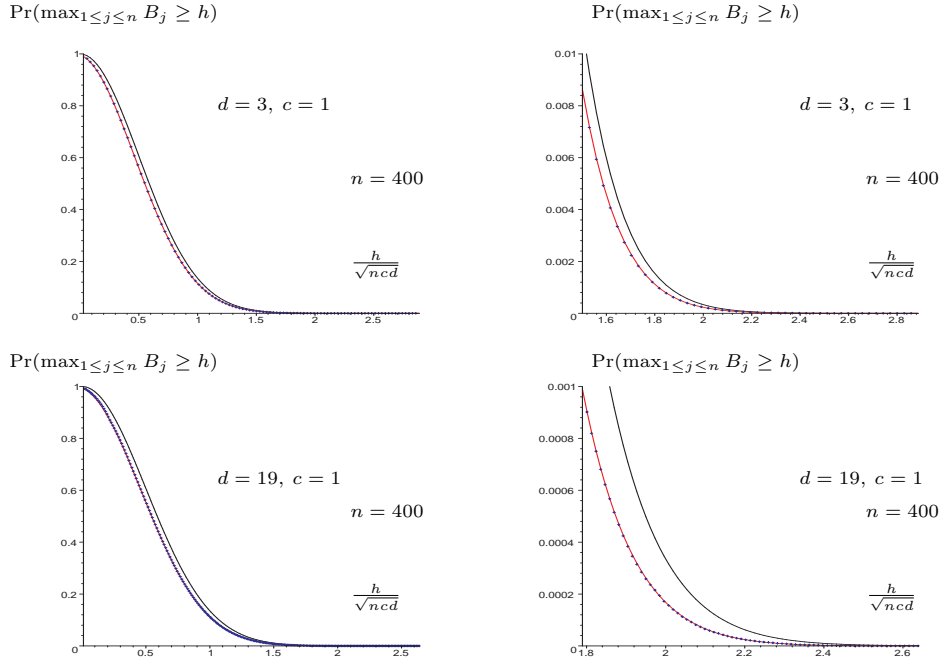


Fig. 2: Plots of the Rayleigh distribution (upper continuous curves), simulations (dots) and asymptotics expansions obtained by Equation (27) (lower continuous curve) for the distribution of the height h of the maximum value of the walk. In these examples, we performed 10^7 iterations for the simulations. The plots on the right are zooms for smaller p -values. These plots show a perfect concordance of the simulations with the discrete asymptotics we obtained, and a clear discrepancy with the Rayleigh distribution, perfectly coherent with our expansion terms.

6 Perspectives and conclusion

We provided in this article the generating functions for several cases of bounded lattice paths. We proved that the height of discrete bridges converges weakly to the Rayleigh distribution. In the case of Łukasiewicz bridges, we are able to compute at any order the asymptotic expansion of the discrete distribution of this height. Applying this method to a statistical estimator originating from bioinformatics provides excellent agreements with the corresponding simulations. The generating functions we gave for walks between one or two walls gives access to the asymptotics of the height of meanders (the drift then plays a rôle). It is also possible to apply the same proofs to directed walks in dimension 2. From a computer algebra point of view, it is also interesting to note that most of the formulas we got are involving symmetric functions, which can be handled in an efficient way with some variants of the platypus algorithm [4]. The full version of this article will also consider convergence of moments of the height in our different models. Note that asymptotics for bridges between two barriers also have applications in bioinformatics.

Acknowledgements. We thank the referees for their comments.

References

- [1] BANDERIER, C. *Combinatoire analytique des chemins et des cartes*. Ph.D. thesis, Université Paris VI, 2001.
- [2] BANDERIER, C. Limit laws for basic parameters of lattice paths with unbounded jumps. In *Chauvin, Brigitte (ed.) et al., Mathematics and computer science II*. Birkhäuser. Trends in Mathematics., 2002, pp. 33–47.
- [3] BANDERIER, C., BOUSQUET-MÉLOU, M., DENISE, A., FLAJOLET, P., AND GARDY, D. Generating functions for generating trees. *Discrete Math.* 246, 1-3 (2002), 29–55.
- [4] BANDERIER, C., AND FLAJOLET, P. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science* 281, Issue 1-2 (2002), 37–80. (Special volume dedicated to M. Nivat).
- [5] BANDERIER, C., FLAJOLET, P., SCHAEFFER, G., AND SORIA, M. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Struct. Algorithms* 19, 3-4 (2001), 194–246.
- [6] BANDERIER, C., AND GITTENBERGER, B. Analytic combinatorics of lattice paths: Enumeration and asymptotics for the average area. *Discrete Mathematics and Theoretical Computer Science AG* (2006), 345–355. (Proceedings of the 4th Colloquium of Mathematics and Computer Science).
- [7] BOSTAN, A., AND KAUERS, M. The complete generating function for Gessel walks is algebraic. *Proceedings of the American Mathematical Society* (2010). (With an appendix by M. van Hoeij).
- [8] BOUSQUET-MÉLOU, M. Discrete Excursions. *Sémin. Lothar. Comb.* 57 (2008).
- [9] BOUSQUET-MÉLOU, M., AND PETKOVŠEK, M. Linear recurrences with constant coefficients: The multivariate case. *Discrete Math.* 225, 1-3 (2000), 51–75.

- [10] CHATTERJEE, S. A new approach to strong embeddings. *Probab. Theory and Related Fields* (2010). To Appear.
- [11] CHUNG, F., GRAHAM, R., MORRISON, J., AND ODLYZKO, A. Pebbling a chessboard. *Am. Math. Mon.* 102, 2 (1995), 113–123.
- [12] DE BRUIJN, N. G., KNUTH, D. E., AND RICE, S. O. The average height of planted plane trees. In *Graph theory and computing*. Academic Press, New York, 1972, pp. 15–22.
- [13] DONSKER, M. D. An invariance principle for certain probability limit theorems. *Mem. Am. Math. Soc.* 6 (1951), 1–12. (Four papers on probability).
- [14] DURRETT, R. *Probability: Theory and Examples*. Statistics/Probability Series. Wadsworth & Brooks/Cole, 1991. 453 pages.
- [15] FAYOLLE, G., IASNOGORODSKI, R., AND MALYSHEV, V. *Random walks in the quarter-plane. Algebraic methods, boundary value problems and applications*. Applications of Mathematics. 40. Berlin: Springer. xvi, 160 p. , 1999.
- [16] FLAJOLET, P., AND ODLYZKO, A. M. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences* 25, 2 (1982), 171–213.
- [17] FLAJOLET, P., AND SEDGEWICK, R. *Analytic combinatorics*. Cambridge University Press, 2009. 810 pages. <http://algo.inria.fr/flajolet/Publications/books.html>.
- [18] IGLEHART, D. L. Random walks with negative drift conditioned to stay positive. *J. Appl. Probability* 11 (1974), 742–751.
- [19] KAIGH, W. D. An invariance principle for random walk conditioned by a late return to zero. *Ann. Probability* 4, 1 (1976), 115–121.
- [20] KAUERS, M., KOUTSCHAN, C., AND ZEILBERGER, D. Proof of Ira Gessel’s Lattice Path Conjecture. *Proceedings of the National Academy of Sciences* 106(28) (2009), 11502–11505.
- [21] KELLER, A., BACKES, C., AND LENHOF, H. P. Computation of significance scores of unweighted gene set enrichment analyses. *BMC Bioinformatics* 290, 8 (2007).
- [22] KNUTH, D. E. *The Art of Computer Programming*, third ed., vol. 1. Addison-Wesley, 1997.
- [23] KOMLÓS, J., MAJOR, P., AND TUSNÁDY, G. An approximation of partial sums of independent RV’s and the sample DF. I. *Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32 (1976), 111–131.
- [24] KURKOVA, I. A., AND SUHOV, Y. M. Malyshev’s theory and JS-queues. Asymptotics of stationary probabilities. *Ann. Appl. Probab.* 13, 4 (2003), 1313–1354.
- [25] PANHOLZER, A., AND PRODINGER, H. Towards a more precise analysis of an algorithm to generate binary trees: A tutorial. *The Computer Journal* 41 (1998), 201–204.
- [26] RUBEY, M. Transcendence of generating functions of walks on the slit plane. In *Drmotič (ed.) et al., Mathematics and computer science III*. Birkhäuser. Trends in Mathematics., 2004, pp. 49–58.