# Lattice paths, sampling without replacement, and limiting distributions


M. Kuba, A. Panholzer and H. Prodinger*

Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria

kuba@dmg.tuwien.ac.at, Alois.Panholzer@tuwien.ac.at

and

Mathematics Department
Stellenbosch University
7602 Stellenbosch, South Africa

hproding@sun.ac.za



Submitted: Apr 22, 2008; Accepted: May 19, 2009; Published: May 29, 2009
Mathematics Subject Classification: 05A15, 60C05.



**Abstract**

In this work we consider weighted lattice paths in the quarter plane $\mathbb{N}_0 \times \mathbb{N}_0$. The steps are given by $(m,n) \to (m-1,n)$, $(m,n) \to (m,n-1)$ and are weighted as follows: $(m,n) \to (m-1,n)$ by $m/(m+n)$ and step $(m,n) \to (m,n-1)$ by $n/(m+n)$. The considered lattice paths are absorbed at lines $y = x/t - s/t$ with $t \in \mathbb{N}$ and $s \in \mathbb{N}_0$. We provide explicit formulæ for the sum of the weights of paths, starting at $(m,n)$, which are absorbed at a certain height $k$ at lines $y = x/t - s/t$ with $t \in \mathbb{N}$ and $s \in \mathbb{N}_0$, using a generating functions approach. Furthermore these weighted lattice paths can be interpreted as probability distributions arising in the context of Pólya-Eggenberger urn models, more precisely, the lattice paths are sample paths of the well known sampling without replacement urn. We provide limiting distribution results for the underlying random variable, obtaining a total of five phase changes.



*Keywords:* Lattice paths, Sampling without replacement, urn models, Lévy distribution



*This work was supported by the Austrian Science Foundation FWF, grant S9608-N23 and by the South African Science Foundation NRF, grant 2053748. The second author wants to thank the Department of Mathematical Sciences, University of Stellenbosch, for its support and hospitality during a research visit, where a part of this work has been carried out.

# 1 Introduction

## 1.1 Lattice paths

Let $\mathcal{S} \subseteq \mathbb{N}_0 \times \mathbb{N}_0$ denote a set of lattice points in the quarter plane[1]. We consider lattice paths with steps $(m, n) \to (m-1, n)$ and $(m, n) \to (m, n-1)$, starting at $(m, n) \in \mathcal{S}$. The steps are weighted as follows: step $(m, n) \to (m-1, n)$ is weighted by $m/(m+n)$ and step $(m, n) \to (m, n-1)$ is weighted by $n/(m+n)$. We are interested in weighted lattice paths starting at $(m, n) \in \mathcal{S}$, which touch or cross a certain line $y = f(x)$ at height $k$, with $0 \leq k \leq n$ and $\mathcal{S} = \mathcal{S}_y$ suitably defined. We consider the following types of lines: $y = x/t - s/t$, with $t \in \mathbb{N}$ and $s \in \mathbb{N}_0$. In this case we have $\mathcal{S} = \{(m, n) \mid m \geq tn + s, n \in \mathbb{N}_0\}$. We are interested in the sum of the weights of all paths starting at $(m, n) \in \mathcal{S}$ and touching $y = x/t - s/t$ at height $k$, with $0 \leq k \leq n$, not touching the line $y$ before, which we call *absorption at line* $y$.

From a probabilistic point of view we can interpret the desired numbers as probability distributions of a random variable $Y_{m,n}$, which can be described as follows. A particle is located at a certain point $(m, n) \in \mathcal{S}$, and moves randomly to the left or downwards with probabilities depending on the actual position:

$$\mathbb{P}\{(m, n) \to (m-1, n)\} = \frac{m}{m+n}, \quad \mathbb{P}\{(m, n) \to (m, n-1)\} = \frac{n}{m+n}.$$

The random variable $Y_{m,n}$ describes the height $k$ at which a particle starting at $(m, n)$ is absorbed, i.e., where it is touching or crossing a line $y = f(x)$ for the first time. The searched probability $\mathbb{P}\{Y_{m,n} = k\}$ is then equal to the sum of the weights of all lattice paths, starting at $(m, n)$, which touch or cross the line $y = f(x)$ at height $k$. We can also formulate this problem in the context of certain urn models.

## 1.2 Pólya-Eggenberger urn models and sampling without replacement

Pólya-Eggenberger urn models are defined as follows. We start with an urn containing $n$ white balls and $m$ black balls. The evolution of the urn occurs in discrete time steps. At every step a ball is chosen at random from the urn. The color of the ball is inspected and then the ball is reinserted into the urn. According to the observed color of the ball, balls are added/removed due to the following rules. If we have chosen a white ball, we put into the urn $a$ white balls and $b$ black balls, but if we have chosen a black ball, we put into the urn $c$ white balls and $d$ black balls. The values $a, b, c, d \in \mathbb{Z}$ are fixed integer values and the urn model is specified by the $2 \times 2$ ball replacement matrix $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

One of the most fundamental urn models is the so-called *sampling without replacement urn*, associated with the ball replacement matrix $M = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. In this urn model a parameter of interest is the number $Y_{m,n}$ of remaining white balls, after all black balls have been removed, starting with $n$ white and $m$ black balls. The formal setting is as follows. We have a state space $\mathcal{S}$, which is given by $\mathcal{S} := \{(m, n) \mid m, n \in \mathbb{N}_0\}$. Further we have a set of absorbing states $\mathcal{A} := \{(0, n) \mid n \in \mathbb{N}_0\}$, where the evolution of the urn stops.

[1] Throughout this work we use the notations $\mathbb{N} := \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 := \{0, 1, 2, \dots\}$.

The problem of absorption at line $y$ can also be formulated in the context of Pólya-Eggenberger urn models, where the state space $\mathcal{S}$ and the set of absorbing states $\mathcal{A}$ are suitably modified. E.g., for $y = x/t - s/t$ we consider a sampling urn with ball replacement matrix $M = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, where the state space is given by $\mathcal{S} = \{(m,n) \mid m \geq tn + s, n \in \mathbb{N}_0\}$ and the set of absorbing states by $\mathcal{A} = \{(tm+s, m) \mid m \in \mathbb{N}_0\}$.

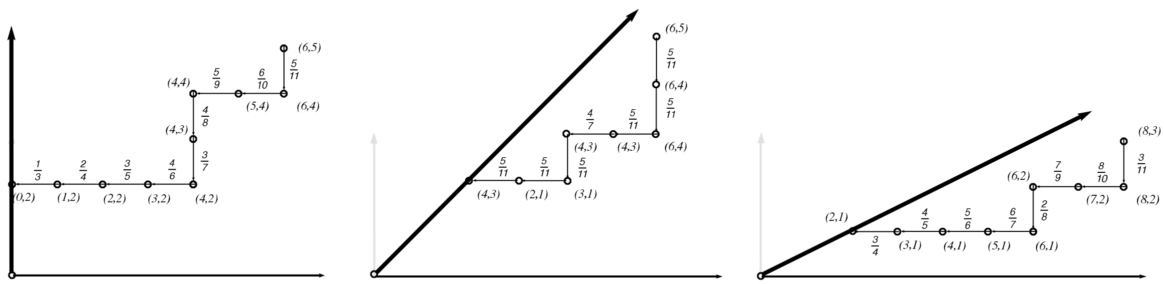


Figure 1: Sample paths for absorption at $x = 0$, $y = x$ and $y = x/2$.

## 1.3 Motivation and related work

The main motivation for this study is to combine the different areas of lattice path enumeration, see, e.g., Mohanty [11], Banderier and Flajolet [2], and Pólya-Eggenberger urn models, see Flajolet et al. [3, 4], and Hwang et al. [6]. For the weighted lattice paths and absorbing lines studied, we obtain closed formulæ for the probability of absorption at height $k$. These explicit results also allow a detailed study of the limiting behaviour of the random variable $Y_{m,n}$. We can completely characterize the limit laws of $Y_{m,n}$ and phase changes appearing depending on the growth behaviour of $m$ and $n$ of the starting point $(m, n)$.

The problem studied was also motivated by a combinatorial game involving card guessing, which has been analyzed by Levasseur [10], Zagier [14], Sulanke [13] and Knopfmacher and Prodinger [9]. One starts with a deck consisting of $m$ red and $n$ black cards. A guess is made as to the color of the top card, after which it is revealed and discarded. To maximize the number of correct guesses one chooses the color corresponding to the majority of cards remaining in the deck. We will revisit this problem and provide limiting distribution results.

Our analysis is based on a generating function approach: we will derive functional equations for suitably defined generating functions and use arguments of [1] in order to obtain explicit formulæ for the probability functions of the considered random variables.

## 1.4 Results

Here we collect the exact and asymptotic results obtained for the weighted lattice paths studied.

**Theorem 1.** *The probability $\mathbb{P}\{Y_{m,n} = k\}$ that a particle starting at $(m, n)$ is absorbed on the* *line $y = x/t - s/t$ at height $k$ is given by the following explicit formula:*

$$\mathbb{P}\{Y_{m,n} = k\} = \frac{\binom{(t+1)k+s}{k}\Big(\binom{m+n-1-s-(t+1)k}{n-k} - t\binom{m+n-1-s-(t+1)k}{n-k-1}\Big)}{\binom{m+n}{n}},$$

*for $m \geq tn + s$ and $0 \leq k \leq n$, with $t \in \mathbb{N}$, $s \in \mathbb{N}_0$.*

In the next theorem we state the obtained limiting distribution results of $Y_{m,n}$, depending on the growth of $m$ and $n$, for absorbing lines $y = x/t - s/t$, with $t \in \mathbb{N}$, $s \in \mathbb{N}_0$ fixed. We use here the notation $Y_n \xrightarrow{\mathcal{L}} Y$ for the weak convergence, i.e., the convergence in distribution, of a sequence of random variables $Y_n$ to a random variable $Y$. Furthermore we use $Y \stackrel{\mathcal{L}}{=} Z$ for the equality in distribution of two random variables $Y$ and $Z$.

**Theorem 2.** *The limiting distribution behaviour of $Y_{m,n}$ is, for $m \to \infty$ and depending on the growth of $n = n(m)$, described as follows.*

1. *$n = o(m)$: The random variable $Y_{m,n}$ is asymptotically zero, as $m$ tends to infinity:*

$$\mathbb{P}\{Y_{m,n} = 0\} = \frac{(m-s-tn)\binom{m}{s}}{(m+n-s)\binom{m+n}{s}} \to 1.$$

2. *$n \sim \rho m$, such that $0 < \rho < \frac{1}{t}$: The random variable $Y_{m,n}$ weakly converges to a discrete random variable $X_\rho$,*

$$\mathbb{P}\{Y_{n,m} = k\} \sim \frac{1-t\rho}{1+\rho} \cdot \binom{(t+1)k+s}{k} \frac{\rho^k}{(1+\rho)^{(t+1)k+s}}, \quad k \in \mathbb{N}_0.$$

3. *$n \sim \frac{m}{t}$, such that $n = \frac{m}{t} - \frac{s}{t} - \ell$, with $m = o(\ell^2)$: The scaled random variable $\frac{\ell^2 t^2}{m^2} Y_{m,n}$ is asymptotically Gamma distributed with shape parameter $\frac{1}{2}$ and scale parameter $\frac{2(t+1)}{t}$,*

$$\frac{m^2}{t^2 l^2}\mathbb{P}\Big\{\frac{\ell^2 t^2 Y_{n,m}}{m^2} = x\Big\} \sim \frac{\sqrt{t}}{\sqrt{2(t+1)}\,\sqrt{\pi}\sqrt{x}} e^{-\frac{xt}{2(t+1)}}, \quad x > 0.$$

4. *$n \sim \frac{m}{t}$, such that $n = \frac{m}{t} - \frac{s}{t} - \ell$, with $\ell \sim \rho\sqrt{m}$ and $\rho > 0$: The scaled random variable $\frac{t}{m} Y_{m,n}$ weakly converges to a random variable $Y_\rho$ with density $f_\rho(x)$,*

$$\frac{m}{t}\mathbb{P}\Big\{\frac{t}{m} Y_{m,n} = x\Big\} \sim f_\rho(x) = \frac{\rho t}{\sqrt{2\pi x(1+t)}\,(1-x)^{\frac{3}{2}}} e^{-\frac{\rho^2 t^2 x}{2(1+t)(1-x)}}, \quad 0 < x < 1.$$

5. *$n \sim \frac{m}{t}$, such that $n = \frac{m}{t} - \frac{s}{t} - \ell$, with $\ell \to \infty$ and $\ell = o(\sqrt{m})$: The shifted and scaled random variable $\frac{1}{\ell^2}(\frac{m}{t} - Y_{m,n})$ is asymptotically Lévy distributed with parameter $\frac{t}{t+1}$,*

$$\ell^2 \mathbb{P}\Big\{\frac{1}{\ell^2}\Big(\frac{m}{t} - Y_{m,n}\Big) = x\Big\} \sim \frac{\sqrt{t}}{\sqrt{2\pi(t+1)}\,x^{\frac{3}{2}}} e^{-\frac{t}{2x(t+1)}}, \quad x > 0.$$

6. *$n \sim \frac{m}{t}$, such that $n \sim \frac{m}{t} - \frac{s}{t} - \ell$, with $\ell \in \mathbb{N}$ fixed: The shifted random variable $\frac{m}{t} - \frac{s}{t} - Y_{m,n}$ weakly converges to a discrete random variable $Y_\ell$,*

$$\mathbb{P}\Big\{\frac{m}{t} - \frac{s}{t} - Y_{m,n} = k\Big\} \sim \frac{t^{tk}}{(1+t)^{tk-\ell}}\binom{(t+1)k-\ell-1}{k-\ell}\frac{\ell}{k}, \quad k \geq \ell.$$

**Remark 1.** The Lévy distribution is a stable distribution. It is a special case of the Lévy skew alpha-stable distribution, which in its general form does not have an analytically expressible probability density. Furthermore the moments of the Lévy distribution do not exist. Hence, for $n \sim m/t$, such that $n = m/t - s/t - \ell$, with $\ell \to \infty$ and $\ell = o(\sqrt{m})$, the random variable $(m/t - Y_{m,n})/\ell^2$ converges in distribution, but without convergence of any integer moment. The occurrence of the Lévy distribution was some kind of surprise for the authors. Note that in the case of absorption at line $x = 0$ one can always prove moment convergence [3, 6, 7].

# 2 Lines $y = x/t - s/t$, with $t \in \mathbb{N}$, $s \in \mathbb{N}_0$

## 2.1 Recurrences

Let $\varphi_{m,n}(v) = \sum_{k\geq 0} \mathbb{P}\{Y_{m,n} = k\}v^k$ denote the probability generating function of $Y_{m,n}$, where $Y_{m,n} = Y_{m,n}(s,t)$. We usually drop the dependence of $Y_{m,n}$ on $s$ and $t$ for the sake of simplicity. By using the basic decomposition of the paths according to the first step and taking into account the absorbing lines, the problem can be translated into the following recurrence:

$$\varphi_{m,n}(v) = \frac{m}{m+n}\varphi_{m-1,n}(v) + \frac{n}{m+n}\varphi_{m,n-1}(v), \quad \text{for } m > tn+s, \quad n \geq 1,$$
$$\varphi_{tm+s,m}(v) = v^m, \quad \text{for } m \geq 0, \qquad \varphi_{m,0}(v) = 1, \quad \text{for } m \geq s.$$

This recurrence will be treated by introducing the normalized functions

$$\Phi_{m,n}(v) = \binom{m+n}{m}\varphi_{m,n}(v).$$

We obtain

$$\Phi_{m,n}(v) = \Phi_{m-1,n}(v) + \Phi_{m,n-1}(v), \quad \text{for } m > tn+s \geq 1, \tag{1a}$$
$$\Phi_{tm+s,m}(v) = \binom{(t+1)m+s}{m}v^m, \quad \text{for } m \geq 0, \qquad \Phi_{m,0}(v) = 1, \quad \text{for } m \geq s. \tag{1b}$$

## 2.2 Generating functions

We introduce the trivariate generating function

$$F(z,u,v) = \sum_{n\geq 0}\sum_{m\geq tn+s} \Phi_{m,n}(v)z^m u^{m-tn-s},$$

and also $t+1$ auxiliary functions $F_k(z,v)$ defined by

$$F_k(z,v) = \sum_{n\geq 0} \Phi_{tn+s+k,n}(v) z^{tn+s+k}, \quad \text{for } 0 \leq k \leq t.$$

Due to (1b) the generating function $F_0(z,v)$ is already known:

$$F_0(z,v) = z^s \sum_{n\geq 0} \Phi_{tn+s,n}(v) z^{tn} = z^s \sum_{n\geq 0} \binom{(t+1)n+s}{n} (vz^t)^n. \tag{2}$$

Using (1a) we obtain the following functional equation for $F(z,u,v)$:

$$(1 - zu - \frac{1}{u^t})F(z,u,v) = (1 - \frac{1}{u^t})F_0(z,v) - \frac{1}{u^t}\sum_{k=1}^{t} u^k F_k(z,v). \tag{3}$$

It is advantageous to write equation (3) in the following form:

$$(zu^{t+1} - u^t + 1)F(z,u,v) = (1 - u^t)F_0(z,v) + \sum_{k=1}^{t} u^k F_k(z,v). \tag{4}$$

**Remark 2.** The standard approach for solving equation (4) is the kernel method[2], we refer to Prodinger [12] for a survey about this method, and the works of Banderier et al. [1, 2] for applications. We will proceed in a slightly different way using a variation of the kernel method based on arguments of Mireille Bousquet-Mélou [1].

## 2.3 Solving the functional equation

Equation (4) gives a simple relation between the unknown functions $F(z,u,v)$ and $F_k(z,v)$, $1 \leq k \leq t$. In order to solve (4) we consider the so-called characteristic equation

$$P(z,u) = zu^{t+1} - u^t + 1 = 0.$$

By general considerations on the roots of the characteristic polynomial $P(z,u)$, as figured out in [2], it follows that $P(z,u)$ can be written in the following form:

$$P(z,u) = (zu - \lambda_{t+1}(z))(u - \lambda_1(z))(u - \lambda_2(z)) \cdots (u - \lambda_t(z)),$$

with functions $\lambda_1(z), \dots, \lambda_{t+1}(z)$ analytic around $z = 0$. In the following we use the abbreviation $\lambda_i := \lambda_i(z)$, $1 \leq i \leq t+1$, where we do not express explicitly the dependence of $\lambda_i$ on $z$. Now we use the fact that $F(z,u,v)$ is an analytic function in a neighbourhood of $z = 0$. Thus we can evaluate $F(z,u,v)$ and therefore equation (4) at $u = \lambda_1(z), \dots, u = \lambda_t(z)$ for $z$ in a

[2]A preliminary version of this work, where the kernel method has been used, can be found on the authors websites and the arXiv.

neighbourhood of $0$. Since $P(z,\lambda_i(z))=0$, for $1\le i\le t$, we obtain from (4) after plugging in $u=\lambda_i$, $1\le i\le t$, a system of $t$ linear equations for the unknown functions $F_k(z,v)$, $1\le k\le t$:

$$\begin{aligned}(1-\lambda_1^t)F_0(z,v)+\sum_{k=1}^{t}\lambda_1^k F_k(z,v)&=0,\\ \vdots\qquad\qquad\qquad &\ \ \vdots\\ (1-\lambda_t^t)F_0(z,v)+\sum_{k=1}^{t}\lambda_t^k F_k(z,v)&=0.\end{aligned}$$

Applying Cramer's rule we can write the solution of this linear system of equations as a quotient of determinants, with $1\le k\le t$:

$$F_k(z,v)=-\begin{vmatrix}\lambda_1 & \cdots & \lambda_1^{k-1} & 1-\lambda_1^t & \lambda_1^{k+1} & \cdots & \lambda_1^t\\ \lambda_2 & \cdots & \lambda_2^{k-1} & 1-\lambda_2^t & \lambda_2^{k+1} & \cdots & \lambda_2^t\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \lambda_t & \cdots & \lambda_t^{k-1} & 1-\lambda_t^t & \lambda_t^{k+1} & \cdots & \lambda_t^t\end{vmatrix}\cdot\begin{vmatrix}\lambda_1 & \lambda_1^2 & \cdots & \lambda_1^t\\ \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^t\\ \vdots & \vdots & \vdots & \vdots\\ \lambda_t & \lambda_t^2 & \cdots & \lambda_t^t\end{vmatrix}^{-1}F_0(z,v). \tag{5}$$

By Bousquet-Mélou's [1] observation we only need to derive $F_t(z,v)$. In the case $k=t$ we can split the $t$-th row in the determinant appearing in the numerator of (5) and obtain easily:

$$F_t(z,v)=\Big(\frac{(-1)^t}{\lambda_1\lambda_2\cdots\lambda_t}+1\Big)F_0(z,v). \tag{6}$$

Now let $N(z,u)$ denote the right-hand-side of the functional equation (4) for $F(z,u,v)$:

$$N(z,u):=(1-u^t)F_0(z,v)+\sum_{k=1}^{t}u^kF_k(z,v).$$

The quantity $N(z,u)$ is a polynomial in $u$ with leading coefficient $F_t(z,v)-F_0(z,v)$, whose zeros are exactly $\lambda_1,\ldots,\lambda_t$. Hence, after normalization, we have a leading monomial $u^t$, and the normalized polynomial factors nicely into the following expression:

$$\frac{N(z,u)}{F_t(z,v)-F_0(z,v)}=\prod_{k=1}^{t}(u-\lambda_k).$$

Since

$$(-1)^{t+1}\lambda_1\cdots\lambda_t\lambda_{t+1}=1,$$

which is a direct consequence of the factorization of $P(z,u)$, we get

$$\begin{aligned}F_t(z,v)-F_0(z,v)&=\Big(\frac{(-1)^t}{\lambda_1\cdots\lambda_t}+1\Big)F_0(z,v)-F_0(z,v)=\frac{(-1)^t}{\lambda_1\ldots\lambda_t}F_0(z,v)\\ &=-\lambda_{t+1}\,F_0(z,v).\end{aligned}$$

Hence we finally obtain:

$$F(z,u,v)=\frac{N(z,u)}{P(z,u)}=\frac{-\lambda_{t+1}F_0(z,v)\prod_{k=1}^t(u-\lambda_k)}{(zu-\lambda_{t+1})\prod_{k=1}^t(u-\lambda_k)}=\frac{\lambda_{t+1}}{\lambda_{t+1}-zu}F_0(z,v). \tag{7}$$

### 2.4 Extracting coefficients

To obtain the required probabilities we only have to extract coefficients of (7). By using the definition of $F(z,u,v)$ we get

$$\begin{aligned}
\mathbb{P}\{Y_{m,n}=k\} &= \frac{1}{\binom{m+n}{m}}[z^m u^{m-tn-s}v^k]F(z,u,v) = \frac{1}{\binom{m+n}{m}}[z^{tn+s}(uz)^{m-tn-s}v^k]\frac{F_0(z,v)}{1-\frac{zu}{\lambda_{t+1}}} \\
&= \frac{1}{\binom{m+n}{m}}[z^{tn+s}v^k]\frac{F_0(z,v)}{\lambda_{t+1}^{m-tn-s}} = \frac{1}{\binom{m+n}{m}}[z^{tn}v^k]\frac{\sum_{l\ge 0}\binom{(t+1)l+s}{l}(vz^t)^l}{\lambda_{t+1}^{m-tn-s}} \\
&= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}[z^{t(n-k)}]\frac{1}{\lambda_{t+1}(z)^{m-tn-s}}.
\end{aligned}$$

To extract coefficients from this expression we consider the characteristic equation $zu^{t+1}-u^t+1=0$. Multiplying with $z^t$ and using the substitution $\lambda := zu$ leads then to the equation

$$z^t = \lambda^t(1-\lambda).$$

Of course, $\lambda=\lambda_{t+1}(z)$ is exactly the function implicitly defined by this equation, which satisfies $\lambda_{t+1}(0)=1$. To apply the Lagrange inversion formula we introduce the substitutions $\tilde{z} := z^t$ and $w := \lambda - 1$ leading to the following equation, which is suitable for that:

$$\tilde{z} = -(1+w)^t w.$$

Thus we obtain further:

$$\begin{aligned}
\mathbb{P}\{Y_{m,n}=k\} &= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}[\tilde{z}^{n-k}]\frac{1}{(1+w)^{m-tn-s}} \\
&= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}\frac{1}{n-k}[w^{n-k-1}]\Big(\frac{1}{(1+w)^{m-tn-s}}\Big)'\Big(-\frac{1}{(1+w)^t}\Big)^{n-k} \\
&= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}\frac{(-1)^{n-k-1}(m-s-tn)}{n-k}[w^{n-k-1}]\frac{1}{(1+w)^{m+1-tk-s}} \\
&= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}\frac{m-s-tn}{n-k}\binom{m-s+n-1-(t+1)k}{n-k-1} \\
&= \frac{\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}\frac{m-s-tn}{m-s+n-(t+1)k}\binom{m-s+n-(t+1)k}{n-k},
\end{aligned}$$

for $m \ge tn+s$ and $0 \le k \le n$. Now we rewrite the probabilities obtained in the following form, which is given in Theorem 1:

$$\mathbb{P}\{Y_{m,n}=k\} = \frac{\binom{(t+1)k+s}{k}\Big(\binom{m+n-1-s-(t+1)k}{n-k} - t\binom{m+n-1-s-(t+1)k}{n-k-1}\Big)}{\binom{m+n}{n}},$$

for $m \ge tn+s$, with $0 \le k \le n$.

**Remark 3.** Due to the simplicity of the result concerning the lines $y = x/t - s/t$ it is natural to ask for a more direct derivation of the probabilities. In the following we will sketch such an alternative combinatorial derivation. It is well known that the number of (unweighted) lattice paths from $(m, n)$ to the origin $(0, 0)$ with unit steps to the left or downwards is given by $\binom{m+n}{m}$. We need the following result, which can be found in Mohanty [11].

**Lemma 1.** *The number of (unweighted) lattice paths from the origin to* $(m, n)$*, which never pass above the line* $y = x/t$*, is given by*

$$\frac{m - tn + 1}{m + n + 1}\binom{m+n+1}{n} = \binom{m+n}{n} - t\binom{m+n}{n-1}.$$

Now we obtain the probabilities $\mathbb{P}\{Y_{m,n} = k\}$ by fixing the last step and using Lemma 1:

$$\mathbb{P}\{Y_{m,n} = k\} = \mathbb{P}\{(m, n) \to (tk + s, k) \mid \ \ y = (x - s)/t \text{ is not touched}\}.$$

By fixing the last step we get further

$$\begin{aligned}\mathbb{P}\{Y_{m,n} = k\} = \mathbb{P}\{&(m, n) \to (tk + s + 1, k) \mid \ \ y = (x - s)/t \text{ is not touched}\} \\ &\times \mathbb{P}\{(tk + s + 1, k) \to (tk + s, k)\}.\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{P}\{Y_{m,n} = k\} &= \frac{\#_{\text{Paths}}\big((m - tk - s - 1, n - k) \to (0, 0) \mid y = x/t \text{ is not passed}\big)}{\#_{\text{Paths}}\big((m, n) \to (0, 0)\big)} \\ &\qquad \times \#_{\text{Paths}}\big((tk + s + 1, k) \to (0, 0)\big) \cdot \frac{tk + s + 1}{(t + 1)k + s + 1} \\ &= \frac{\Big(\binom{m+n-1-s-(t+1)k}{n-k} - t\binom{m+n-1-s-(t+1)k}{n-k-1}\Big)\binom{(t+1)k+s+1}{k}}{\binom{m+n}{m}} \cdot \frac{tk + s + 1}{(t + 1)k + s + 1} \\ &= \frac{\Big(\binom{m+n-1-s-(t+1)k}{n-k} - t\binom{m+n-1-s-(t+1)k}{n-k-1}\Big)\binom{(t+1)k+s}{k}}{\binom{m+n}{m}}.\end{aligned}$$

## 2.5 Deriving the limiting distributions

The main results of this paper, which describe the limiting distribution of $Y_{m,n}$ depending on the growth behaviour of $m$ and $n$, are obtained from the probability mass function given in Theorem 1 after a careful application of Stirling's formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\Big(1 + \mathcal{O}\big(\frac{1}{n}\big)\Big). \tag{8}$$

Note that one main difficulty is to "guess" the right normalizations required. Once the right guess was made, we still had to carry out quite lengthy and tedious calculations in order to obtain the stated results. Since these calculations are very lengthy, we will only present as an example the derivation of the local limit for the case $s = 0$ and $n \sim \frac{m}{t}$, such that $n = \frac{m}{t} - \ell$, with $\ell \sim \rho\sqrt{m}$ and $\rho > 0$. Other computations are carried out in a similar manner.

$$
\begin{aligned}
\frac{m}{t}\mathbb{P}\{\frac{tY_{m,n}}{m} = x\} &\sim \frac{\rho t}{\sqrt{2\pi(t+1)x}\,(1-x)^{\frac{3}{2}}} \frac{(1 - \frac{\rho}{\sqrt{m}(1+\frac{1}{t})(1-x)})^{(1+\frac{1}{t})(1-x)m - \rho\sqrt{m}}}{(1 - \frac{\rho t}{\sqrt{m}(1-x)})^{(1-x)\frac{m}{t} - \rho\sqrt{m}}} \\
&\quad \times \frac{(1 - \frac{\rho t}{\sqrt{m}})^{\frac{m}{t} - \rho\sqrt{m}}}{(1 - \frac{\rho}{\sqrt{m}(1+\frac{1}{t})})^{(1+\frac{1}{t})m - \rho\sqrt{m}}} \\
&\sim \frac{\rho t}{\sqrt{2\pi(t+1)x}\,(1-x)^{\frac{3}{2}}} \Big(\frac{(1 - \frac{\rho t}{\sqrt{m}(1-x)})(1 - \frac{\rho t}{\sqrt{m}(1+t)})}{(1 - \frac{\rho t}{\sqrt{m}})(1 - \frac{\rho t}{\sqrt{m}(1-x)(1+t)})}\Big)^{\rho\sqrt{m}} \times \\
&\quad \times \Big(\frac{(1 - \frac{\rho t}{\sqrt{m}(1+t)(1-x)})^{t+1}}{(1 - \frac{\rho t}{\sqrt{m}(1-x)})}\Big)^{\frac{m}{t}(1-x)} \Big(\frac{(1 - \frac{\rho t}{\sqrt{m}})}{(1 - \frac{\rho t}{\sqrt{m}(1+t)})^{t+1}}\Big)^{\frac{m}{t}} \\
&\sim \frac{\rho t}{\sqrt{2\pi(t+1)x}\,(1-x)^{\frac{3}{2}}} \frac{e^{-\frac{\rho^2 t}{1-x}} e^{-\frac{\rho^2 t}{1+t}}}{e^{-\rho^2 t} e^{-\frac{\rho^2 t}{(1+t)(1-x)}}} e^{\frac{\rho^2 t\binom{t+1}{2}}{(1+t)^2(1-x)}} e^{\frac{-\rho^2 t\binom{t+1}{2}}{(1+t)^2}} \\
&\sim \frac{\rho t}{\sqrt{2\pi x(1+t)}\,(1-x)^{\frac{3}{2}}} e^{-\frac{\rho^2 t^2 x}{2(1+t)(1-x)}}.
\end{aligned}
$$

# 3 Limit laws for the card guessing game again revisited

One starts with a deck consisting of $m$ red and $n$ black cards. A guess is made as to the color of the top card, after which it is revealed and discarded. To maximize the number of correct guesses one chooses the color corresponding to the majority of cards remaining in the deck. Let $Z_{m,n}$ denote the random variable counting the number of correct guesses starting with $m$ red and $n$ black cards. The following result was obtained by Sulanke [13], and also by Knopfmacher and Prodinger [9].

**Theorem 3** (Sulanke; Knopfmacher and Prodinger)**.** *The exact distribution of the random variable $Z_{m,n}$ counting the number of correct guesses in the card guessing game starting with $m$ red and $n$ black cards is given as follows:*

$$
\mathbb{P}\{Z_{m,n} = k\} = \frac{\binom{m+n}{k} - \binom{m+n}{k+1}}{\binom{m+n}{m}}, \quad \textit{for } m \le k \le m+n.
$$

Since no limit laws were derived in [9, 13], we complete the analysis by stating the limit laws for $Z_{m,n}$.

**Corollary 1.** *The random variable $Z_{m,n}$ counting the number of correct guesses in the card guessing game starting with $m$ red and $n$ black cards satisfies, for $m \to \infty$ and depending on the growth of $n = n(m)$, the following limit laws.*

- *The region $n = o(m)$: the centered random variable $Z^*_{m,n} = Z_{m,n} - m$ is asymptotically zero:*
$$\mathbb{P}\{Z^*_{m,n} = 0\} = 1 - \frac{n}{m+1} \sim 1.$$

- *The region $n \sim \rho m$, with $0 < \rho < 1$: the centered random variable $Z^*_{m,n} = Z_{m,n} - m$ is asymptotically geometrically distributed with parameter $\rho$:*
$$\mathbb{P}\{Z^*_{m,n} = k\} \sim (1-\rho)\rho^k, \quad k \in \mathbb{N}_0.$$

- *The region $n = m - \ell$, with $m = o(\ell^2)$: the centered and scaled random variable $Z^*_{m,n} = \frac{\ell}{m}(Z_{m,n} - m)$ is asymptotically exponential distributed:*
$$\frac{m}{\ell}\mathbb{P}\{Z^*_{m,n} = x\} \sim e^{-x}, \quad x \geq 0.$$

- *The region $n = m - \ell$, with $\ell \sim \rho\sqrt{m}$ and $\rho > 0$: the centered and scaled random variable $Z^*_{m,n} = (Z_{m,n} - m)/\sqrt{m}$ weakly converges to $Z_\rho$ with density $f_{Z_\rho}(x)$,*
$$\sqrt{m}\,\mathbb{P}\{Z^*_{m,n} = x\} \sim (\rho + 2x)e^{-x(\rho+x)}, \quad x \geq 0.$$

- *The region $n = m - \ell$, with $\ell = o(\sqrt{m})$: the centered and scaled random variable $Z^*_{m,n} = (Z_{m,n} - m)/\sqrt{m}$ is asymptotically Rayleigh distributed:*
$$\sqrt{m}\,\mathbb{P}\{Z^*_{m,n} = x\} \sim 2te^{-t^2}, \quad x \geq 0.$$

*Proof sketch.* The limiting distributions can be derived by an application of Stirling's formula. The calculations are again quite lengthy, so we leave the details to the interested reader. ☐

# Conclusion and Acknowledgement

We have analysed the distribution of weighted lattice paths absorbed at certain lines using a generating function approach. Moreover, we have derived limiting distributions for the underlying random variable. A wealth of different distributions arises in the limit depending on the starting position. In one case we obtained the Lévy distribution, which implies that for this case moment convergence does not hold. We address here the interesting open problem of extending the studies of the limiting distribution behaviour of the height of the absorption to other absorbing lines, and also to other classes of weighted lattice paths, or equivalently to other types of Pólya-Eggenberger urn models, such as the OK-Corral urn model [3, 8] with ball replacement matrix $M = \left(\begin{smallmatrix} 0 & -1 \\ -1 & 0 \end{smallmatrix}\right)$.

The authors thank the referee for valuable remarks improving the presentation of this work.

## References


[1] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, D. Gouyou-Beauchamps. Generating Functions of Generating Trees. *Discrete Mathematics* 246, 29–55, 2002.

[2] C. Banderier and P. Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science* 281, 37–80, 2002.

[3] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. *Discrete Mathematics and Computer Science, Proceedings of Fourth Colloquium on Mathematics and Computer Science*, Ph. Chassaing Editor., vol. AG, 59–118, 2006.

[4] P. Flajolet, J. Gabarró, H. Pekari. Analytic Urns, *Annals of Probability* 33, 1200–1233, 2005.

[5] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Reading, 1994.

[6] H.-K. Hwang, M. Kuba and A. Panholzer, Analysis of some exactly solvable diminishing urn models. *Proceedings of the 19th International Conference on Formal Power Series and Algebraic Combinatorics*, Nankai University, Tianjin, 2007.

Online available: `http://www.fpsac.cn/PDF-Proceedings/Posters/43.pdf`

[7] H.-K. Hwang, M. Kuba and A. Panholzer, Diminishing urn models: Analysis of exactly solvable models, manuscript, 2009.

[8] J. F. C. Kingman, Martingales in the OK Corral, *Bulletin of the London Mathematical Society* 31, 601–606, 1999.

[9] A. Knopfmacher and H. Prodinger, A simple card guessing game revisited, *Electronic journal of combinatorics* 8, paper R13, 9 pages, 2001.

[10] K. Levasseur, How to beat your kids at their own game, *Mathematical Magazine* 61, 301–305, 1988.

[11] S.G. Mohanty, *Lattice path counting and applications*, Academic Press, New York, 1979.

[12] H. Prodinger, The kernel method: a collection of examples, *Séminaire Lotharingien de Combinatoire*, B50f, 19 pp., 2004.

[13] R. A. Sulanke, Guessing, Ballot Numbers, and Refining Pascal's Triangle. Manuscript. Online available: `http://math.boisestate.edu/~sulanke/PAPERS/guessingweb.ps`

[14] D. Zagier, How often should you beat your kids? *Mathematical Magazine 63*, 89–92, 1990.