

Knuth's big-chooser matchbox process: the case of many matchboxes

Mark Dukes and Andrew Mullins

ABSTRACT. Banach's matchbox problem considers the setting of two matchboxes that each initially contain the same number of matches. Boxes are chosen with equal probability and a match removed each time. The problem concerns the law of the number of matches remaining in one box once the other box empties. Knuth considered a generalization of this problem whereby *big-choosers* arrive with probability p and remove a match from the box with the most number remaining, and *little-choosers* arrive with probability $1 - p$ and remove a match from the box with the least number remaining.

In this paper we consider Knuth's generalization for the case of k matchboxes in which there are *big-choosers* and *little-choosers*. We determine the generating function for the expected number of matches remaining in $k - 1$ matchboxes once a box first empties, a quantity we refer to as the 'residue'. Interestingly, this generating function is a quotient whose denominator contains a generating function for a special case of the Raney numbers. The form for this generating function allows us to give an expression for the expected residue in terms of a sum that involves diagonal state return probabilities, where a diagonal state is a configuration in which all matchboxes each contain the same number of matches. We use analytic techniques to determine the asymptotic behaviour of this expected value for all values of p , which involves the study of an asymmetric random walk.

In addition to this we consider the expected value of the order of the first return to a diagonal state and determine the asymptotic behaviour of this quantity. The coefficients of the diagonal state probability generating function are shown to be related to 'manila folder configurations in a filing cabinet', and we make this connection precise. This allows us to use known results for the enumeration of such manila folder configurations to give a closed form expression for the diagonal state return probabilities.

CONTENTS

1. Introduction	1
2. A recursion and diagonal-state numbers	3
2.1. A recursion for the expected residue	3
2.2. Diagonal state returns and the Raney numbers	3
3. The expected residue	6
4. Asymptotic analysis of the expected residue	11
5. Diagonal state matters	17
5.1. Expected order of first diagonal return	17
5.2. Diagonal state probabilities and manila folders	20
Acknowledgements	23
References	23

1. Introduction

Banach's matchbox problem [4, Sec. 8] considered two matchboxes each initially containing n matches. At every time-step a box is selected at random and a match removed. The problem is to determine the probability that one box contains a prescribed number of matches at the moment the other box empties. In his playfully entitled 'The Toilet Paper Problem' paper,

2020 *Mathematics Subject Classification.* 05A15; 05A19; 60C05.

Key words and phrases. Banach's matchbox problem; Bernoulli process; big-chooser; Raney number; asymptotics; mean first passage time; manila folder.

Knuth [15] generalized this problem by introducing a choice-rule relating to the box¹ containing the most number of matches. In his generalization, at every time-step the box with the most number of matches remaining is selected with probability p and a match removed, while the other box is selected with probability $q = 1 - p$. He showed that the expected number of matches remaining in one box once the other box empties (a quantity he refers to as the *residue*) is:

$$n - \sum_{i=1}^{n-1} (n-i) c_i p^i q^{i-1},$$

where $c_i = \frac{1}{i} \binom{2(i-1)}{i-1}$ is the i^{th} Catalan number.² This setting was further considered by Stirzaker [20] wherein he elegantly derived Knuth's results by framing the problem as a simple random walk on the integers. In addition to the expected value, Stirzaker was able to give a closed form for the probability generating function and extend the results to the case where the two matchboxes can initially contain a different number of matches. Separately, Duffy and Dukes [2] studied the large deviations of the residue. The limiting behaviour of the processes is discussed in Wallner [21, Sec. 3.5]. Related work includes the papers of Gutiérrez and Subercaseaux [10], Cacoullos [1], Lengyel [13], Stadje [19], and Keane et al. [12]. In addition to this, we find it interesting to note that Prodinger [16] used the kernel method to derive the generating function for the expected residue that was originally derived in [15].

In this paper we consider Knuth's generalization of Banach's matchbox problem for the case of k matchboxes. The k matchboxes each initially contain n matches. People arrive in succession and choose a match from one of the boxes. There are two types of people: *big-choosers* and *little-choosers*. A big-chooser takes a match from a box with the most number of matches remaining, while a little-chooser takes a match from a box with the least number remaining. In this context, it is not significant how one deals with the event that there are multiple boxes from which a big- or little-chooser may choose. We suppose that the probability of a person being a big-chooser is $p \in (0, 1)$ and the probability of a person being a little-chooser is $q = 1 - p$. Given this setup, we will consider the expected (aggregate) number of matches in the $k - 1$ non-empty boxes at the moment that the remaining box first becomes empty, and we will call this aggregate quantity the *residue*, in keeping with Knuth's terminology.

The state of the k matchboxes any time is represented by a k -tuple $(a_1, a_2, \dots, a_k)$ with $a_1 \geq a_2 \geq \dots \geq a_k$. The approach we take is more closely aligned with Knuth's than with Stirzaker's. While Stirzaker's solution was arguably more succinct, unfortunately for us one of the key features of a simple random walk which Stirzaker was able to exploit, namely that a walk from points a to b must visit every intervening point en route, is not true of more general walks where the steps to the left and right are not of equal length. Indeed, if a walk starts at point a and takes steps of $+2$ (with probability q) and -1 (with probability $1 - q$), then it is possible (for non-trivial q) for walks to avoid any specified intervening point en route to b .

In Theorem 3.3 we determine the generating function for the expected residue. Interestingly, this generating function is a quotient whose denominator contains a generating function for a special case of the Raney numbers. The form for this generating function allows us to give, in Proposition 3.4, an expression for the expected residue in terms of a sum that involves *diagonal state* return probabilities, where a diagonal state is a matchbox configuration in which all matchboxes each contain the same number of matches. We use analytic techniques to determine, in Theorem 4.1, the asymptotic behaviour of this expected residue for all values of p , and this involves the study of an asymmetric random walk.

In addition to this, in Theorem 5.1, we present both the generating function and an expression for the expected order of the first return to a diagonal state. The asymptotic behaviour of this quantity is determined in Proposition 5.2. The coefficients of the diagonal state probability

¹The setting of Knuth's generalization considered two rolls of toilet paper, each initially containing n sheets, instead of matchboxes. We choose to use the matchbox terminology.

²We use the Catalan numbers as defined in [15], whereas the modern convention is to refer to c_{n+1} as the n th Catalan number.

generating function are shown to be related to manila folder configurations in a filing cabinet, and we make this connection precise in Section 5.2. This allows us to use the known enumeration of such manila folder configurations to give a closed form expression for the diagonal state probabilities.

2. A recursion and diagonal-state numbers

2.1. A recursion for the expected residue. Let $X_n^{(k)}(p)$ be the random variable that is the sum of the number of matches in the $k - 1$ larger boxes at the moment the k th box empties, as defined in Section 1. We refer to this quantity as the *residue*. Let $M_n^{(k)}(p)$ be the expected value of $X_n^{(k)}(p)$, i.e. the expected residue. Notice that $X_n^{(k)}(p)$ takes values in the interval $[k - 1, (k - 1)n] := \{k - 1, \dots, (k - 1)n\}$, and so $M_n^{(k)}(p)$ also takes values in this range. For the case $k = 3$, the first few values are easy to calculate. The value $M_1^{(3)}(p) = 2$ since the first person empties the third box regardless of what type of chooser they are.

In considering $X_2^{(3)}(p)$, the state of the system goes from $(2, 2, 2) \rightarrow (2, 2, 1)$ with probability 1, and we represent this by writing $(2, 2, 2) \xrightarrow{1} (2, 2, 1)$. From there, $(2, 2, 1) \xrightarrow{q} (2, 2, 0)$ so that $\mathbb{P}(X_2^{(3)}(p) = 4) = q$, and alternatively $(2, 2, 1) \xrightarrow{p} (2, 1, 1)$. Subsequently, there are two possibilities for what happens until we have an empty box: $(2, 1, 1) \xrightarrow{p} (1, 1, 1) \xrightarrow{1} (1, 1, 0)$, which results in $X_2^{(3)}(p) = 2$, and $(2, 1, 1) \xrightarrow{q} (2, 1, 0)$, which results in $X_2^{(3)}(p) = 3$. It follows that $M_2^{(3)}(p) = 4q + p(2p + 3q) = 4 - p - p^2$. Figure 1 illustrates the experimental value of $M_{100}^{(3)}(p)$ as a function of p .

During the time up until a box first empties, a time that we will refer to as the *first phase*, the expected value can be calculated by recursions very similar to those stated by Knuth for the two-box problem. We have $M_n^{(k)}(p) = \text{First}_{(k-1)n, n}^{(k)}(p)$ where $\text{First}_{a, b}^{(k)}(p)$ is the expected value of the residue, given that the $k - 1$ larger boxes collectively contain a matches and the smallest box contains b matches. We may calculate $\text{First}_{a, b}^{(k)}(p)$ recursively using:

$$\text{First}_{a, 0}^{(k)}(p) = a; \tag{1}$$

$$\text{First}_{(k-1)a, a}^{(k)}(p) = \text{First}_{(k-1)a, a-1}^{(k)}(p) \quad \text{if } a > 0; \tag{2}$$

$$\text{First}_{a, b}^{(k)}(p) = p \cdot \text{First}_{a-1, b}^{(k)}(p) + q \cdot \text{First}_{a, b-1}^{(k)}(p) \quad \text{if } a > (k-1)b > 0. \tag{3}$$

The above recursions have straightforward explanations. The first is trivial. The second holds since $(a, a, \dots, a, a) \xrightarrow{1} (a, a, \dots, a, a - 1)$. The third holds by conditional expectation.

The solution to the recursions above depends on some numbers and power series that feature throughout this paper, so we will first present those.

2.2. Diagonal state returns and the Raney numbers. We call a state in which all k matchboxes have the same number of matches a *diagonal state*. Consider the event that there are no diagonal states encountered in between the diagonal states $(n, n, \dots, n)$ and $(n - i, n - i, \dots, n - i)$. We will refer to such an event as a *first diagonal return of order i* , or simply a *first diagonal return*. Let us denote by $s_i^{(k)}$ the number of ways in which this can happen.

The probability weight associated with a first diagonal return of order i is $p^{(k-1)i}q^{i-1} = \frac{1}{q}(p^{k-1}q)^i$. Notice that up until the first box empties, the difference in size between the largest and second smallest boxes can be no greater than one. The reason for this behaviour is that big-choosers have a leveling effect on the $(k - 1)$ non-smallest boxes. By considering the aggregate of the $k - 1$ larger boxes, $s_i^{(k)}$ is equal to the number of lattice paths from $((k - 1)n, n)$ to $((k - 1)(n - i), n - i)$ that take unit south and west steps, and lie strictly below the

$$\text{boundary line: } (k - 1)y - x = 0,$$

except at the two endpoints.

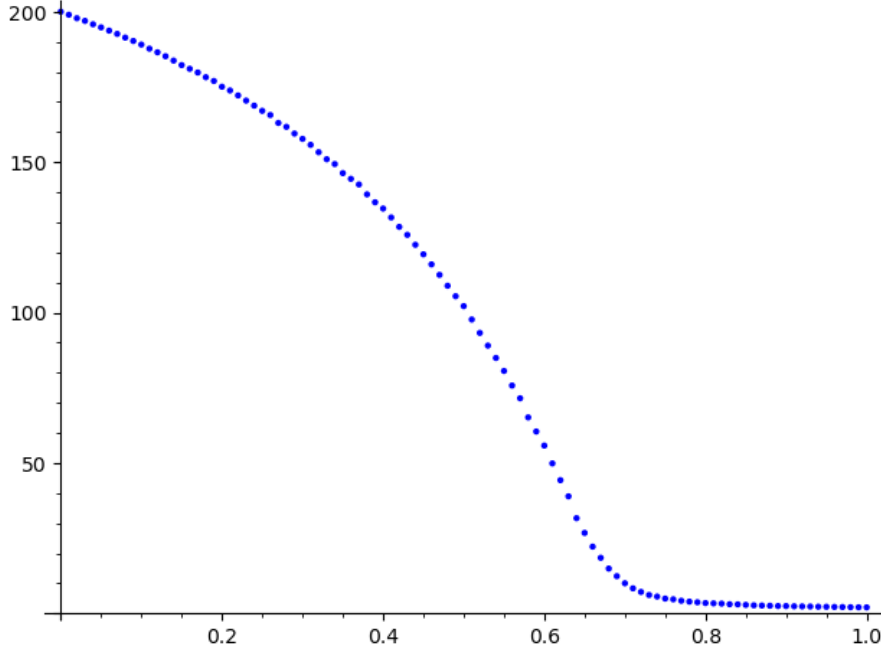


FIGURE 1. The experimental value of $M_{100}^{(3)}(p)$. The horizontal axis represents the value of p while the vertical axis represents $M_{100}^{(3)}(p)$.

Paths through the first quadrant of the lattice that are weakly bounded above by the boundary line are equivalent to ballot problems. Dvoretzky and Motzkin [3] established the following result about a ballot in which a votes are cast for candidate A and b votes are cast for candidate B :

If α is an integer satisfying $0 < \alpha < \frac{a}{b}$, then the probability that throughout the counting the number of votes registered for A is always greater than α times the number of votes registered for candidate B equals $\frac{a-\alpha b}{a+b}$.

In our case, any path from $((k-1)n, n)$ to $((k-1)(n-i), n-i)$ lying below the given boundary line must start with a step from $((k-1)n, n)$ to $((k-1)n, n-1)$. There are $\binom{(k-1)i+i-1}{i-1} = \binom{ki-1}{i-1}$ unbounded paths through the lattice from $((k-1)n, n-1)$ to $((k-1)(n-i), n-i)$. Dvoretzky and Motzkin's theorem tells us, by setting $a = (k-1)i$, $b = i-1$ and $\alpha = k-1$, that the probability of choosing one that lies strictly below $(k-1)y - x = 0$ is $k-1/(ki-1)$. Thus the number of such paths is

$$s_i^{(k)} = \frac{k-1}{ki-1} \binom{ki-1}{i-1} = \frac{k-1}{k(i-1)+k-1} \binom{k(i-1)+k-1}{i-1}.$$

This form of a binomial expression is well-known in enumerative combinatorics and represents a special case of the Raney numbers [17]:

$$\text{Raney}_{c,r}(\ell) = \frac{r}{\ell c + r} \binom{\ell c + r}{\ell}.$$

These numbers have several combinatorial interpretations and are also sometimes referred to as the two-parameter Fuss-Catalan numbers. For example, $\text{Raney}_{c,r}(\ell)$ is the number of sequences $(a_1, \dots, a_{\ell c + r})$ where $a_i \in \{1, 1-c\}$ with all partial sums $a_1 + \dots + a_j$ positive and $a_1 + \dots + a_{\ell c + r} = r$. The numbers in which we are interested are the special case:

$$s_i^{(k)} = \text{Raney}_{k,k-1}(i-1).$$

That our numbers are instances of the Raney numbers allows us to make use of identities that they satisfy. Interesting aspects of these numbers can be found in Graham et al. [8, Sec. 7.5, Example 5], Hilton and Pedersen [11], and Zhou and Yan [23].

Define the generating function

$$S^{(k)}(z) := \sum_{n \geq 1} s_n^{(k)} z^n.$$

It is known, see for example Liszewska and Młotkowski [14, Sec. 9.2], that $S^{(k)}(z)$ satisfies

$$S^{(k)}(z) = \frac{z}{(1 - S^{(k)}(z))^{k-1}}. \quad (4)$$

The first few such power series are:

$$S^{(2)}(z) = \sum_{n \geq 1} \frac{1}{n} \binom{2(n-1)}{n-1} z^n = z + z^2 + 2z^3 + 5z^4 + 14z^5 + 42z^6 + \dots$$

$$S^{(3)}(z) = \sum_{n \geq 1} \frac{1}{n} \binom{3(n-1)+1}{n-1} z^n = z + 2z^2 + 7z^3 + 30z^4 + 143z^5 + 728z^6 + \dots$$

$$S^{(4)}(z) = \sum_{n \geq 1} \frac{1}{n} \binom{4(n-1)+2}{n-1} z^n = z + 3z^2 + 15z^3 + 91z^4 + 612z^5 + 4389z^6 + \dots$$

The coefficients of $S^{(2)}(z)$ are the Catalan numbers and $S^{(2)}(z) = (1 - \sqrt{1-4z})/2$. An exact expression for $S^{(3)}(z)$ is also known

$$S^{(3)}(z) = \frac{4}{3} \sin^2 \left(\frac{1}{3} \arcsin \sqrt{\frac{27z}{4}} \right),$$

see e.g. Gessel [7, Lemma 3.2]. Both are easily seen to satisfy equation (4).

As we mentioned above, the probability associated with one of the paths counted by $s_i^{(k)}$ is $p^{(k-1)i} q^{i-1}$, and so we will be interested in the probability generating function

$$\sum_{i \geq 1} s_i^{(k)} p^{(k-1)i} q^{i-1} z^i = \frac{1}{q} S^{(k)}(p^{k-1} q z).$$

The quotient $1/(1 - \frac{1}{q} S^{(k)}(p^{k-1} q z))$ will also appear frequently in this paper so we discuss it briefly here. Let us write

$$f^{(k)}(z) = f_0^{(k)} + f_1^{(k)} z + f_2^{(k)} z^2 + \dots := \frac{1}{1 - \frac{1}{q} S^{(k)}(p^{k-1} q z)}. \quad (5)$$

The function $\frac{1}{q} S^{(k)}(p^{k-1} q z)$ is the probability generating function for the order of the first return to a diagonal state, i.e. the coefficient of z^i is the probability that the first return to a diagonal state occurs after ki steps. Consequently the quotient $1/(1 - \frac{1}{q} S^{(k)}(p^{k-1} q z))$ is the probability generating function for the event of being in a diagonal state after some multiple of k steps, but not necessarily for the first time. See e.g. Grimmett and Stirzaker's book [9, Sec. 5.3, Thm. 1(a)].

Given the discussion so far about first-quadrant lattice paths from $((k-1)n, n)$ to the origin (and other points) that lie weakly below the boundary line, it would be sensible to introduce some terminology related to this. Let $\text{Paths}_k(n)$ be the set of all those paths from $((k-1)n, n) \mapsto (0, 0)$ that take unit south and west steps and lie in the first quadrant weakly below the boundary line $(k-1)y - x = 0$. Let $\text{Paths}_k(n, i)$ be the subset of those paths in $\text{Paths}_k(n)$ that return to the boundary line precisely i times between the two endpoints. Let $\text{Paths}_k := \{\epsilon\} \cup_{n \geq 1} \text{Paths}_k(n)$, where ϵ is the empty path. Notice that members of all of the above sets can be represented by words in $\{D, L\}^*$ where D corresponds to a unit south step and L corresponds to a unit west step. Finally, let $\text{paths}_k(n) = |\text{Paths}_k(n)|$ and $\text{paths}_k(n, i) = |\text{Paths}_k(n, i)|$.

3. The expected residue

We are now in a position to solve the recurrence stated in equations (1)–(3). First we will prove a recursive expression for the expected residue $M_n^{(k)}(p)$. We will then use that recursion to give an expression for the generating function $M^{(k)}(z)$ for the numbers $\left(M_n^{(k)}(p)\right)_{n \geq 1}$. The generating function has a pleasing form and it is possible to extract the coefficient of z^n in order to give a non-recursive expression for $M_n^{(k)}(p)$ that involves only numbers that appear in Subsection 2.2.

Lemma 3.1. $M_1^{(k)}(p) = k - 1$. For all $n \geq 2$ we have

$$M_n^{(k)}(p) = \frac{1}{q} \sum_{0 < i < n} s_i^{(k)} (p^{k-1}q)^i M_{n-i}^{(k)}(p) + L_n^{(k)}(p), \quad (6)$$

where

$$L_n^{(k)}(p) = (p^{k-1}q)^n \frac{1}{q} \sum_{j=k}^{(k-1)n} j d_{n,j}^{(k)} (1/p)^j$$

and

$$d_{n,j}^{(k)} = \begin{cases} \binom{kn-2-j}{n-2} - \sum_{\ell=1}^{n-\lceil \frac{j}{k-1} \rceil} s_\ell^{(k)} \cdot \binom{k(n-\ell)-1-j}{n-\ell-1} & \text{if } k \leq j \leq (k-1)n \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The recurrences in equations (1)–(3) provide a means to calculate $M_n^{(k)}(p)$ for general n . The expression for $M_1^{(k)}(p)$ follows directly from these. One systematic way to evaluate $M_n^{(k)}(p) = M_{(k-1)n,n}^{(k)}(p)$ for $n \geq 2$ is to apply the recursion to all terms except those terms of the form $\text{First}_{(k-1)a,a}^{(k)}(p)$ for $a < n$. For example we can write:

$$\begin{aligned} M_2^{(3)}(p) &= p^2 M_1^{(3)}(p) + (3pq + 4q) \\ M_3^{(3)}(p) &= p^2 M_2^{(3)}(p) + 2p^4 q M_1^{(3)}(p) + (6p^3 q^2 + 8p^2 q^2 + 10pq^2 + 6q^2). \end{aligned}$$

Doing this for general n will result in an expression for $M_n^{(k)}(p)$ of the following type:

$$\begin{aligned} M_n^{(k)}(p) &= \text{First}_{(k-1)n,n}^{(k)}(p) = \sum_{i=1}^{n-1} \alpha_{n,i}^{(k)}(p) M_{(k-1)(n-i),n-i}^{(k)}(p) + \beta_n^{(k)}(p) \\ &= \sum_{i=1}^{n-1} \alpha_{n,i}^{(k)}(p) M_{n-i}^{(k)}(p) + \beta_n^{(k)}(p), \end{aligned}$$

where $\alpha_{n,i}^{(k)}(p)$ and $\beta_n^{(k)}(p)$ are yet-to-be determined functions. One may easily observe that, due to the form of the recursions, $\alpha_{n,i}(p)$ is a function of i , k , and p , but not of n . The quantity $\alpha_{n,i}(p)$ is the probability of a chooser process that returns to a ‘diagonal’ state $(x, x, \dots, x)$, for the first time, at $(n-i, n-i, \dots, n-i)$. From Subsection 2.2 this probability is $\alpha_i^{(k)}(p) = s_i^{(k)} (p^{k-1}q)^i / q$. Thus we have

$$M_n^{(k)}(p) = \frac{1}{q} \sum_{i=1}^{n-1} s_i^{(k)} (p^{k-1}q)^i M_{n-i}^{(k)}(p) + \beta_n^{(k)}(p). \quad (7)$$

The first term on the right hand side of equation (7) represents the expected residue amongst those chooser processes that cause a return to some diagonal state $(n-i, n-i, \dots, n-i)$ before

the first matchbox empties. The second term represents those that do not encounter a diagonal state. Therefore we can write

$$\beta_n^{(k)}(p) = \sum_j j d_{n,j}^{(k)} p^{(k-1)n-j} q^{n-1}$$

where $d_{n,j}^{(k)}$ is the number of lattice paths from $((k-1)n, n-1) \rightarrow \dots \rightarrow (j, 1) \rightarrow (j, 0)$ that do not touch the boundary line $(k-1)y - x = 0$.

Consider the set of all paths from $((k-1)n, n-1) \rightarrow \dots \rightarrow (j, 1) \rightarrow (j, 0)$. The number of such paths is $\binom{((k-1)n-j)+(n-2)}{n-2} = \binom{kn-2-j}{n-2}$. Let us call a path in this set *good* if it does not touch the boundary line $(k-1)y - x = 0$, and *bad* otherwise. Every bad path is characterized by the path returning to the boundary line before reaching $(j, 1)$. Suppose this happens for a bad path for the first time at $((k-1)(n-\ell), n-\ell)$ for some $\ell \geq 1$. The number of ways a path can go from $((k-1)n, n-1) \rightarrow \dots \rightarrow ((k-1)(n-\ell), n-\ell)$ is s_ℓ and, since it is a bad path, there are no further restrictions on the steps from $((k-1)(n-\ell), n-\ell)$ to $(j, 1)$. That latter number is $\binom{(k-1)(n-\ell)-j+n-\ell-1}{n-\ell-1} = \binom{(k-\ell)-j-1}{n-\ell-1}$. Therefore the number of good paths is

$$d_{n,j}^{(k)} = \binom{kn-2-j}{n-2} - \sum_{\ell \geq 1} s_\ell^{(k)} \cdot \binom{(k-1)(n-\ell)-j-1}{n-\ell-1}. \quad \square$$

Let

$$M^{(k)}(z) = \sum_{n \geq 1} M_n^{(k)}(p) z^n \quad \text{and} \quad L^{(k)}(z) = \sum_{n \geq 2} L_n^{(k)}(p) z^n.$$

Note the starting index of $L^{(k)}(z)$ is 2. This is because $L_n^{(k)}(p)$ denotes the contribution to $M_n^{(k)}(p)$ of paths which have no diagonal return, and there can be no such paths in the case $n = 1$.

Lemma 3.2. *Let $k \geq 2$. Then*

$$L^{(k)}(z) = \frac{z(kp - (k-1))S^{(k)}(p^{k-1}qz) - qz^2((2k-1)p - (2k-2) + (k-1)(qz))}{q^2(1-z)^2}.$$

Proof. By definition,

$$L_n^{(k)}(p) = \sum_{j=k}^{(k-1)n} j d_{n,j}^{(k)} p^{(k-1)n-j} q^{n-1} = \frac{(p^{k-1}q)^n}{pq} \sum_{j=k}^{(k-1)n} j d_{n,j}^{(k)} (1/p)^{j-1}.$$

The generating function for these is then

$$\begin{aligned} L^{(k)}(z) &= \sum_{n \geq 2} L_n^{(k)}(p) z^n \\ &= \sum_{n \geq 2} z^n \left(\frac{(p^{k-1}q)^n}{pq} \sum_{j=k}^{(k-1)n} j d_{n,j}^{(k)} (p^{-1})^{j-1} \right) \\ &= \frac{1}{pq} \sum_{n \geq 2} \left(\sum_{j=k}^{(k-1)n} j d_{n,j}^{(k)} (p^{-1})^{j-1} \right) (p^{k-1}qz)^n \\ &= \frac{1}{pq} \frac{\partial D^{(k)}(x, y)}{\partial x} \Bigg|_{\substack{x \leftarrow 1/p \\ y \leftarrow p^{k-1}qz}} \end{aligned}$$

where $D^{(k)}(x, y) := \sum_{n \geq 2} D_n^{(k)}(x) y^n$ and $D_n^{(k)}(x) := \sum_{j=k}^{(k-1)n} d_{n,j}^{(k)} x^j$.

Any path to $(j, 1)$ either passes through $(j, 2)$ or through $(j+1, 1)$, but not both. It follows that $d_{n,j}^{(k)} = d_{n,j+1}^{(k)} + d_{n-1,j-(k-1)}^{(k)}$ holds for those values of n and j for which all quantities in the recursion are defined. It also turns out easier, from a generating function perspective, to

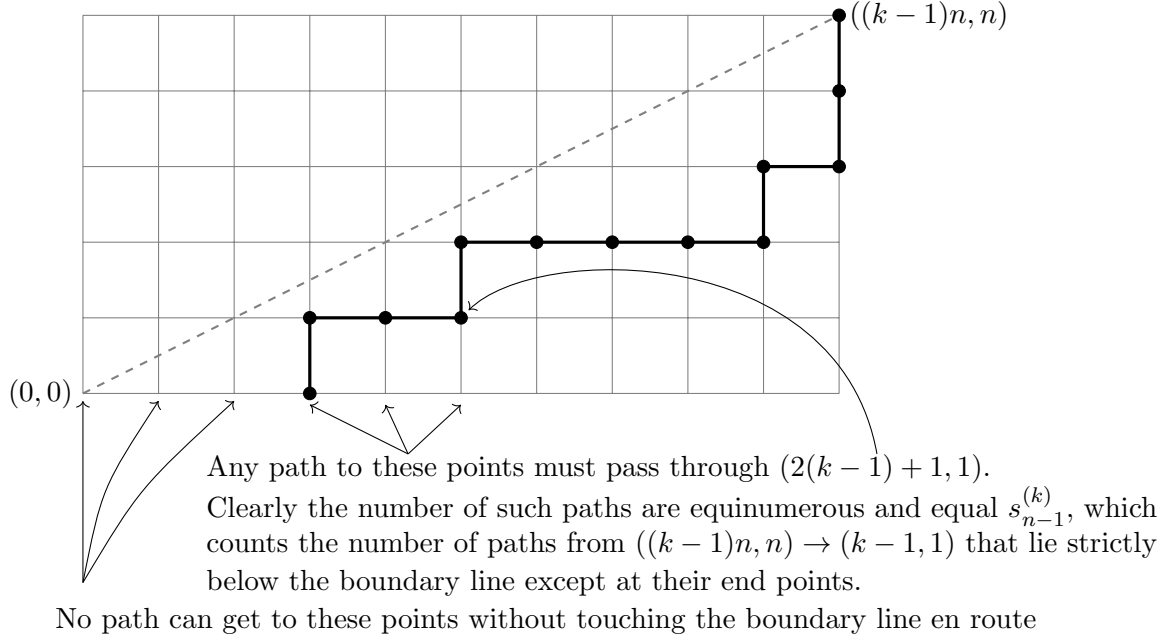


FIGURE 2. Paths counted by the numbers $d_{n,j}^{(k)}$: those paths from $((k-1)n, n)$ to $(j, 0)$ that do not touch the boundary line or the x -axis en-route. This diagram relates to the proof of Lemma 3.2. Here $k = 3$.

reverse the sequence $(d_{n,j}^{(k)})_j$. Set $t_{n,j}^{(k)} := d_{n,(k-1)n-j}^{(k)}$ for all $0 \leq j \leq (k-1)n - k$ and define $T_n^{(k)}(x) := \sum_{j=0}^{(k-1)n-k} t_{n,j}^{(k)} x^j$. Note that $D_n^{(k)}(x) = x^{(k-1)n} T_n(1/x)$. For example, we have

$$\begin{aligned} T_2^{(3)}(x) &= 1 + x \\ T_3^{(3)}(x) &= 1 + 2x + 2x^2 + 2x^3 \\ T_4^{(3)}(x) &= 1 + 3x + 5x^2 + 7x^3 + 7x^4 + 7x^5 \\ T_5^{(3)}(x) &= 1 + 4x + 9x^2 + 16x^3 + 23x^4 + 30x^5 + 30x^6 + 30x^7. \end{aligned}$$

Notice that the expression for $T_2^{(k)}(x)$ for a given value of k is

$$T_2^{(k)}(x) = 1 + x + x^2 + \dots + x^{k-2}.$$

The recurrence for $d_{n,j}^{(k)}$ above can be transformed into one for $t_{n,j}^{(k)}$:

$$t_{n,j}^{(k)} = t_{n,j-1}^{(k)} + t_{n-1,j}^{(k)}.$$

From this it is clear that the sequence $(t_{n,j}^{(k)})$, as a sequence over j , is the sequence of all partial sums of $(t_{n-1,j}^{(k)})$, with the largest term appended to the end of the new sequence $k-1$ times. This is because $t_{n,j}^{(k)}$ is only defined for $j \leq (k-1)n - k$. Notice that the largest value in the sequence $t_{n+1,j}^{(k)}$ for n fixed is $s_n^{(k)}$. See Figure 2 for an illustration of why this is the case. In terms of $T_{n+1}^{(k)}(x)$ and $T_n^{(k)}(x)$ this is

$$T_{n+1}^{(k)}(x) = \frac{T_n^{(k)}(x)}{1-x} - \frac{x^{(k-1)n} s_n^{(k)}}{1-x}.$$

8

This recursion now allows us to give an expression for

$$\begin{aligned}
T^{(k)}(x, y) &:= \sum_{n \geq 2} T_n^{(k)}(x) y^n = T_2^{(k)}(x) y^2 + \sum_{n \geq 2} T_{n+1}^{(k)}(x) y^{n+1} \\
&= T_2^{(k)}(x) y^2 + \sum_{n \geq 2} y^{n+1} \left(\frac{T_n^{(k)}(x)}{1-x} - \frac{x^{(k-1)n} s_n^{(k)}}{1-x} \right) \\
&= T_2^{(k)}(x) y^2 + \frac{x^{k-1} y^2}{1-x} - \frac{y S^{(k)}(x^{k-1} y)}{1-x} + \frac{y}{1-x} T^{(k)}(x, y).
\end{aligned}$$

Rearranging gives

$$\begin{aligned}
T^{(k)}(x, y) \left(1 - \frac{y}{1-x} \right) &= \frac{(1-x) T_2^{(k)}(x) y^2 + x^{k-1} y^2 - y S^{(k)}(x^{k-1} y)}{1-x} \\
&= \frac{(1-x^{k-1}) y^2 + x^{k-1} y^2 - y S^{(k)}(x^{k-1} y)}{1-x},
\end{aligned}$$

and so

$$T^{(k)}(x, y) = \frac{y^2 - y S^{(k)}(x^{k-1} y)}{1-x-y}.$$

Now since $D_n^{(k)}(x) = x^{(k-1)n} T_n(1/x)$ we have

$$\begin{aligned}
D^{(k)}(x, y) &= \sum_{n \geq 2} D_n^{(k)}(x) y^n \\
&= \sum_{n \geq 2} x^{(k-1)n} T_n(1/x) y^n \\
&= T(1/x, x^{k-1} y) \\
&= \frac{(x^{k-1} y)^2 - (x^{k-1} y) S^{(k)}((x^{-1})^{k-1} (x^{k-1} y))}{1 - x^{-1} - x^{k-1} y} \\
&= \frac{x^k y S^{(k)}(y) - x^{2k-1} y^2}{1 + x^k y - x}.
\end{aligned} \tag{8}$$

This gives

$$\begin{aligned}
\frac{\partial D^{(k)}(x, y)}{\partial x} &= \frac{(1 + x^k y - x)(k x^{k-1} y S^{(k)}(y) - (2k-1) x^{2k-2} y^2)}{(1 + x^k y - x)^2} \\
&\quad - \frac{(k x^{k-1} y - 1)(x^k y S^{(k)}(y) - x^{2k-1} y^2)}{(1 + x^k y - x)^2} \\
&= \frac{x^{k-1}(k - (k-1)x) y S^{(k)}(y)}{(1 + x^k y - x)^2} \\
&\quad - \frac{x^{2k-2} y^2((2k-1) - (2k-2)x + (k-1)x^k y)}{(1 + x^k y - x)^2}.
\end{aligned}$$

Using this expression it follows that

$$\begin{aligned}
L(z) &= \frac{1}{pq} \frac{\partial D^{(k)}(x, y)}{\partial x} \bigg|_{\substack{x \leftarrow 1/p \\ y \leftarrow p^{k-1}qz}} = \frac{1}{pq} \frac{\partial D(x, p^{k-1}qz)}{\partial x} \bigg|_{x \leftarrow 1/p} \\
&= \frac{(p^{-1})^{k-1}(k - (k-1)p^{-1})p^{k-1}qzS^{(k)}(p^{k-1}qz)}{pq(1 + (p^{-1})^k(p^{k-1}qz) - p^{-1})^2} \\
&\quad - \frac{(p^{-1})^{2k-2}(p^{k-1}qz)^2((2k-1) - (2k-2)p^{-1} + (k-1)(p^{-1})^k(p^{k-1}qz))}{pq(1 + (p^{-1})^k(p^{k-1}qz) - p^{-1})^2} \\
&= \frac{z(kp - (k-1))S^{(k)}(p^{k-1}qz) - qz^2((2k-1)p - (2k-2) + (k-1)(qz))}{q^2(1-z)^2}. \quad \square
\end{aligned}$$

Having worked through these intermediary lemmas, we are now ready to state the main theorem concerning the generating function for the sequence of expected residues.

Theorem 3.3. $M^{(k)}(z) = \frac{z}{(1-z)^2} \cdot \left(k - \frac{1}{q}\right) + \frac{z}{1-z} \cdot \frac{p}{q} \cdot \frac{1}{\left(1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)\right)}.$

Proof. Multiply both sides of the equation in Lemma 3.1 by z^n and sum over all $n \geq 2$ to get

$$M^{(k)}(z) - M_1^{(k)}(p)z = \frac{1}{q}S^{(k)}(p^{k-1}qz)M^{(k)}(z) + L^{(k)}(z).$$

Rearrange this, while also using the fact that $M_1^{(k)}(p) = k-1$, to get

$$M^{(k)}(z) = \frac{(k-1)z + L^{(k)}(z)}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)}.$$

Lemma 3.2 provides an expression for $L^{(k)}(z)$ and so the numerator above equals

$$\begin{aligned}
(k-1)z + L^{(k)}(z) &= (k-1)z + \frac{z(kp - (k-1))S^{(k)}(p^{k-1}qz)}{q^2(1-z)^2} \\
&\quad - \frac{qz^2((2k-1)p - (2k-2) + (k-1)(pqz))}{q^2(1-z)^2} \\
&= \frac{z}{q^2(1-z)^2} \left((pk - (k-1))S_k(p^{k-1}qz) - pqz + (k-1)q^2 \right).
\end{aligned}$$

Substituting this back into the expression for $M^{(k)}(z)$ gives:

$$\begin{aligned}
M^{(k)}(z) &= \frac{z \left((pk - (k-1))S_k(p^{k-1}qz) - pqz + (k-1)q^2 \right)}{q^2(1-z)^2 \left(1 - \frac{1}{q}S^{(k)}(p^{k-1}qz) \right)} \\
&= \frac{z(kq-1)}{q(1-z)^2} + \frac{pz}{q(1-z) \left(1 - \frac{1}{q}S^{(k)}(p^{k-1}qz) \right)}. \quad \square
\end{aligned}$$

From Theorem 3.3 we now deduce a succinct expression for the expected residue $M_n^{(k)}(p)$.

Proposition 3.4. *For all $n \geq 1$,*

$$M_n^{(k)}(p) = (k-1)n - \frac{p}{q} \left((1 - f_0^{(k)}) + (1 - f_1^{(k)}) + \dots + (1 - f_{n-1}^{(k)}) \right), \quad (9)$$

where $f_i^{(k)}$, defined in equation (5), is the probability of being in a diagonal state after ki steps.

Proof. Extract the coefficient of z^n in $M^{(k)}(z)$ from Theorem 3.3 to find

$$\begin{aligned}
M_n(p) &= [z^n] \frac{kq-1}{q} \cdot \frac{z}{(1-z)^2} + [z^n] \frac{p}{q} \cdot \frac{z}{(1-z) \left(1 - \frac{1}{q} S^{(k)}(p^{k-1}qz)\right)} \\
&= \frac{(k-1)q-p}{q} \cdot [z^n] \frac{z}{(1-z)^2} + \frac{p}{q} \cdot [z^n] \frac{z}{(1-z) \left(1 - \frac{1}{q} S^{(k)}(p^{k-1}qz)\right)} \\
&= (k-1)n - n \frac{p}{q} + \frac{p}{q} \sum_{i=0}^{n-1} [z^i] \frac{1}{1 - \frac{1}{q} S^{(k)}(p^{k-1}qz)} \\
&= (k-1)n - \frac{p}{q} \left(n - \sum_{i=0}^{n-1} [z^i] \frac{1}{1 - \frac{1}{q} S^{(k)}(p^{k-1}qz)} \right).
\end{aligned}$$

From Subsection 2.2 we have

$$\begin{aligned}
M_n^{(k)}(p) &= (k-1)n - \frac{p}{q} \left(n - (f_0^{(k)} + f_1^{(k)} + \dots + f_{n-1}^{(k)}) \right) \\
&= (k-1)n - \frac{p}{q} \left((1 - f_0^{(k)}) + (1 - f_1^{(k)}) + \dots + (1 - f_{n-1}^{(k)}) \right),
\end{aligned}$$

where $f_i^{(k)}$ is the probability that a path starting from the diagonal state $(n, n, \dots, n)$ reaches the diagonal state $(n-i, n-i, \dots, n-i)$. Note that it may well have gone through diagonal states in between. $\square$

4. Asymptotic analysis of the expected residue

Now we turn our attention to the asymptotic behaviour of $M_n^{(k)}(p)$ for fixed p as $n \rightarrow \infty$.

Theorem 4.1. Let $R^* = \frac{(k-1)^{k-1}}{k^k} \cdot \frac{1}{p^{k-1}q}$ and let $\epsilon > 0$. Then

$$M_n^{(k)}(p) = \begin{cases} \left(\left(k - \frac{1}{q}\right)n + \frac{p}{q} \cdot \frac{1}{1 - \lambda_{k,p}} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) \right) & \text{if } q > \frac{1}{k} \\ \sqrt{\frac{2k(k-1)n}{\pi}} + \mathcal{O}\left(\frac{1}{n^{1/2}}\right) & \text{if } q = \frac{1}{k} \\ \frac{(k-1)(2-kq)}{2(1-kq)} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) & \text{if } q < \frac{1}{k}, \end{cases}$$

where $\lambda_{k,p} := \frac{1}{q} S^{(k)}(p^{k-1}q) \in (0, 1)$ when $q > 1/k$.

The values of $\lambda_{k,p}$ for $k = 2, 3, 4$, and 5 are shown in Figure 3. We require several technical lemmas in order to prove Theorem 4.1.

Lemma 4.2. $\frac{1}{q} S_k(p^{k-1}qz)$ converges for $|z| < \frac{(k-1)^{k-1}}{k^k} \frac{1}{p^{k-1}q}$.

Proof. Let $(\alpha_n)_{n \geq 1}$ be the sequence generated by $\frac{1}{q} S^{(k)}(p^{k-1}qz)$ as a power series in z , so that

$$\alpha_n = \frac{1}{q} \cdot \frac{k-1}{kn-1} \binom{kn-1}{n-1} (p^{k-1}qz)^n.$$

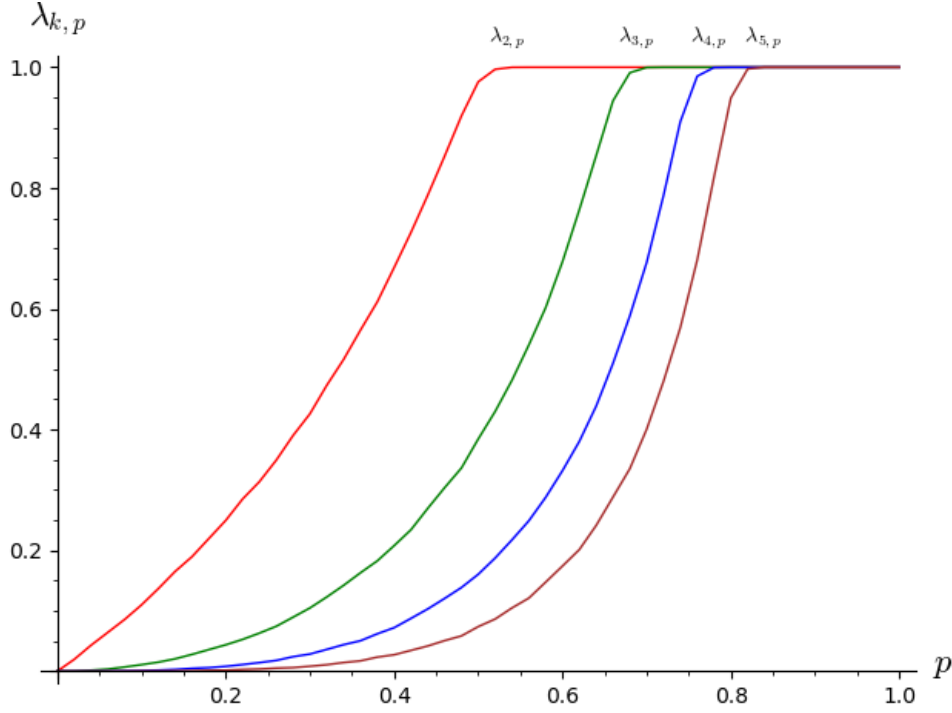


FIGURE 3. The value of $\lambda_{k,p}$ for $k \in \{2, 3, 4, 5\}$. The lines for $\lambda_{2,p}$, $\lambda_{3,p}$, $\lambda_{4,p}$, and $\lambda_{5,p}$ are coloured red, green, blue, and brown, respectively.

Apply the ratio test to $(\alpha_n)_{n \geq 1}$ to find

$$\begin{aligned}
\frac{|\alpha_{n+1}|}{|\alpha_n|} &= \frac{\frac{1}{q} \frac{k-1}{k(n+1)-1} \binom{k(n+1)-1}{n} (p^{k-1}q|z|)^{n+1}}{\frac{1}{q} \frac{k-1}{kn-1} \binom{kn-1}{n-1} (p^{k-1}q|z|)^n} \\
&= \left(\frac{kn-1}{kn+k-1} \right) p^{k-1}q|z| \cdot \frac{(kn+k-1)!}{((k-1)n+k-1)!n!} \cdot \frac{((k-1)n!(n-1)!)}{(kn-1)!} \\
&= \left(\frac{kn-1}{kn+k-1} \right) p^{k-1}q|z| \cdot \frac{1}{n} \cdot \frac{(kn+k-1)(kn+k-2) \cdots (kn)}{((k-1)n+k-1)((k-1)n+k-2) \cdots ((k-1)n+1)} \\
&\rightarrow \frac{k^k}{(k-1)^{k-1}} p^{k-1}q|z| \text{ as } n \rightarrow \infty.
\end{aligned}$$

Therefore $\frac{1}{q} S^{(k)}(p^{k-1}qz)$ converges for $|z| < \frac{(k-1)^{k-1}}{k^k} \frac{1}{p^{k-1}q}$ and diverges for $|z| > \frac{(k-1)^{k-1}}{k^k} \frac{1}{p^{k-1}q}$. $\square$

Notice that, on the interval $p \in (0, 1)$, the function $p \mapsto p^{k-1}q$ has a maximum value of $\frac{(k-1)^{k-1}}{k^k}$ that occurs uniquely at $p = \frac{k-1}{k}$. Therefore $\frac{1}{q} S^{(k)}(p^{k-1}qz)$ is holomorphic on a disc centered at the origin of radius $R^* := \frac{(k-1)^{k-1}}{k^k} \frac{1}{p^{k-1}q}$ and, moreover, $R^* > 1$ for $q \neq \frac{1}{k}$.

Lemma 4.3. *If $\frac{1}{q} S^{(k)}(p^{k-1}qz) = 1$, then $z = 1$.*

Proof. By equation (4), $S^{(k)}(p^{k-1}qz) = \frac{p^{k-1}qz}{(1 - S^{(k)}(p^{k-1}qz))^{k-1}}$. Replacing $S^{(k)}(p^{k-1}qz)$ with q , we find:

$$q = \frac{p^{k-1}qz}{(1-q)^{k-1}}.$$

Rearranging gives $z = 1$. $\square$

Lemma 4.4. *We have*

$$\lambda_{k,p} := \frac{1}{q} S^{(k)}(p^{k-1}q) \begin{cases} = 1 & \text{if } q < \frac{1}{k} \\ \in (0, 1) & \text{if } q > \frac{1}{k}. \end{cases}$$

Proof. Recall that the function $\frac{1}{q} S^{(k)}(p^{k-1}qz)$ is the probability generating function for the order of the first return to a diagonal state. So $\frac{1}{q} S^{(k)}(p^{k-1}q)$ is the probability that an infinite walk will ever return to the boundary line. The initial step in our paths process is a south step and any return to the diagonal requires $k-1$ as many west steps as south steps. We will consider a related random walk in order to determine whether or not this will occur with probability 1.

Consider a random walk $S_n = S_0 + (X_1 + \dots + X_n)$ on $\mathbb{Z}$ that starts at $S_0 = -(k-1)$ and takes steps $X_i \in \{-(k-1), 1\}$ with

$$\begin{aligned} \mathbb{P}(X_i = +1) &= p \\ \mathbb{P}(X_i = -(k-1)) &= q. \end{aligned}$$

The event $S_i = 0$ for this random walk is equivalent to the event of a diagonal return in our paths process.

We are interested in the probability that $S_i = 0$ for some $i > 0$. As the X_i are independent, the expectation

$$\begin{aligned} \mathbb{E}(S_i) &= S_0 + \sum_{j=1}^i \mathbb{E}(X_j) \\ &= -(k-1) + i\mathbb{E}(X_1) \\ &= -(k-1) + i(1 - qk), \end{aligned}$$

since $\mathbb{E}(X_1) = 1 - qk$. Notice that $\mathbb{E}(X_1) > 0$ for $q < 1/k$ and $\mathbb{E}(X_1) < 0$ for $q > 1/k$.

Suppose that $q < 1/k$, and let $1 - qk = \mu$ ($\mu > 0$). Let $\text{var}(X_1) = \sigma^2$ where $0 < \sigma < \infty$. Define $S'_n := X_1 + X_2 + \dots + X_n$. We are interested in $\mathbb{P}(S_n < 0)$ which is equal to $\mathbb{P}(S'_n < k-1)$. The event $S'_n < (k-1)$ is equivalent to $S'_n - n\mu < (k-1) - n\mu$, which in turn is equivalent to

$$\frac{S'_n - n\mu}{\sqrt{n}\sigma} < \frac{(k-1) - n\mu}{\sqrt{n}\sigma} = \frac{\mu}{\sigma} \sqrt{n}.$$

By the central limit theorem (see e.g. Grimmett and Stirzaker [9, §5.10]), we have

$$\frac{S'_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0, 1) \text{ as } n \rightarrow \infty.$$

Therefore, for n sufficiently large,

$$\mathbb{P}(S'_n < k-1) \rightarrow \mathbb{P}\left(Y < \frac{(k-1) - n\mu}{\sqrt{n}\sigma} = \frac{\mu}{\sigma} \sqrt{n}\right), \text{ where } Y \sim N(0, 1).$$

But $\frac{(k-1) - n\mu}{\sqrt{n}\sigma} = \frac{\mu}{\sigma} \sqrt{n} \rightarrow -\infty$ as $n \rightarrow \infty$, so $\mathbb{P}(S'_n < (k-1)) \rightarrow 0$. This implies $\mathbb{P}(S_n \geq 0) \rightarrow 1$ as $n \rightarrow \infty$. As the random walk moves in the positive direction by only taking unit steps, the event $S_n \geq 0$ guarantees that $S_i = 0$ for some $i \leq n$. Therefore

$$\sum_{i \geq 0} s_i^{(k)} p^{(k-1)i} q^{i-1} = \frac{1}{q} S^{(k)}(p^{k-1}q) = 1.$$

Now suppose that $q > 1/k$. Notice that on the closed interval $[0, 1]$ the continuous function $f(q) := p^{k-1}q$ takes the minimum value 0 at $q = 0$ and $q = 1$. It takes a local maximum uniquely at $q = 1/k$. Therefore, by the intermediate value theorem, for every $q_1 > 1/k$ there exists a $q_0 < 1/k$ such that

$$p_0^{k-1}q_0 = p_1^{k-1}q_1,$$

where $p_0 := 1 - q_0$ and $p_1 := 1 - q_1$. We have already seen that $\frac{1}{q_0} S^{(k)}(p_0^{k-1} q_0) = 1$ for any $q_0 < 1/k$. Therefore, for $q_1 > \frac{1}{k}$, we have

$$\begin{aligned} \frac{1}{q_1} S^{(k)}(p_1^{k-1} q_1) &= \frac{q_0}{q_1} \frac{1}{q_0} S^{(k)}(p_1^{k-1} q_1) \\ &= \frac{q_0}{q_1} \frac{1}{q_0} S^{(k)}(p_0^{k-1} q_0) \\ &= \frac{q_0}{q_1} \cdot 1 < 1, \end{aligned}$$

since $q_0 < q_1$. □

A more thorough treatment of this random walk can be found in Sen and Saran [18].

Lemma 4.5. *The derivative*

$$\left. \frac{d}{dz} S^{(k)}(p^{k-1} qz) \right|_{z \leftarrow 1} = \begin{cases} \frac{pq}{1-kq} & \text{if } q < \frac{1}{k} \\ \frac{(1 - q\lambda_{k,p})p^{k-1}q}{(1 - q\lambda_{k,p})^k - (k-1)p^{k-1}q} & \text{if } q > \frac{1}{k}. \end{cases}$$

Proof. By Lemma 4.2, $S^{(k)}(p^{k-1} qz)$ is holomorphic in a region around $z = 1$, for $q \neq \frac{1}{k}$, and so we can evaluate $\left. \frac{d}{dz} S^{(k)}(p^{k-1} qz) \right|_{z \leftarrow 1}$. Starting with $S^{(k)}(p^{k-1} qz) = \frac{p^{k-1} qz}{(1 - S^{(k)}(p^{k-1} qz))^{k-1}}$, we have:

$$\begin{aligned} \frac{d}{dz} S^{(k)}(p^{k-1} qz) &= p^{k-1} qz(1-k)(1 - S^{(k)}(p^{k-1} qz))^{-k} \left(-\frac{d}{dz} S^{(k)}(p^{k-1} qz) \right) + p^{k-1} q(1 - S^{(k)}(p^{k-1} qz))^{1-k}. \end{aligned}$$

By rearranging, one now has an expression for the derivative of $S^{(k)}(p^{k-1} qz)$:

$$\frac{d}{dz} S^{(k)}(p^{k-1} qz) = \frac{p^{k-1} q(1 - S^{(k)}(p^{k-1} qz))}{(1 - S^{(k)}(p^{k-1} qz))^k - (k-1)p^{k-1} qz}. \quad (10)$$

We wish to evaluate this at $z = 1$. Lemma 4.4 gives us two cases:

$$S^{(k)}(p^{k-1} q) = \begin{cases} q & \text{if } q < \frac{1}{k} \\ q\lambda_{k,p} & \text{if } q > \frac{1}{k}. \end{cases}$$

The result follows. □

Lemma 4.6. *For $q < \frac{1}{k}$,*

$$\left. \frac{d^2}{dz^2} S^{(k)}(p^{k-1} qz) \right|_{z \leftarrow 1} = \frac{pq^2(k-1)(2-qk)}{(1-qk)^3}.$$

Proof. Differentiating equation (10), we have:

$$\begin{aligned} \frac{d^2}{dz^2} S^{(k)}(p^{k-1} qz) &= \frac{-p^{k-1} q \frac{d}{dz} S^{(k)}(p^{k-1} qz)}{(1 - S^{(k)}(p^{k-1} qz))^k - (k-1)p^{k-1} qz} \\ &\quad + \frac{p^{k-1} q(1 - S^{(k)}(p^{k-1} qz))}{((1 - S^{(k)}(p^{k-1} qz))^k - (k-1)p^{k-1} qz)^2} (-1) \\ &\quad \times \left[k(1 - S^{(k)}(p^{k-1} qz))^{k-1} \left(-\frac{d}{dz} S^{(k)}(p^{k-1} qz) \right) - (k-1)p^{k-1} q \right] \end{aligned}$$

Evaluating this at $z = 1$, and using Lemmas 4.4 and 4.5, yields the result. □

We are now in a position to prove Theorem 4.1.

Proof. (Proof of Theorem 4.1) In order to determine the asymptotics of $M_n^{(k)}(p) = [z^n]M^{(k)}(z)$, we need to determine the nature of the poles of the function

$$M^{(k)}(z) = \left(k - \frac{1}{q}\right) \frac{z}{(1-z)^2} + \frac{p}{q} \left(\frac{z}{1-z}\right) \left(\frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)}\right).$$

If $\frac{1}{q}S^{(k)}(p^{k-1}qz) = 1$ then $z = 1$, by Lemma 4.3, so the only pole is at $z = 1$. We split our proof into three cases.

Case $q > \frac{1}{k}$: By Lemma 4.4 we have $\frac{1}{q}S^{(k)}(p^{k-1}q) =: \lambda_{k,p} \in (0, 1)$. Now

$$\begin{aligned} [z^n]M^{(k)}(z) &= [z^n] \left(\left(k - \frac{1}{q}\right) \cdot \frac{z}{(1-z)^2} + \frac{p}{q} \cdot \frac{z}{1-z} \cdot \frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)} \right) \\ &= \left(k - \frac{1}{q}\right) n + [z^n] \frac{p}{q} \cdot \frac{z}{1-z} \cdot \frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)}. \end{aligned}$$

As $\frac{1}{q}S^{(k)}(p^{k-1}qz) \neq 1$ at $z = 1$, we have that

$$\frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)} = \sum_{n \geq 0} c_n (1-z)^n,$$

for some constants c_n . We can set $z = 1$ to determine $c_0 = \frac{1}{1 - \lambda_{k,p}}$. Therefore

$$\begin{aligned} \frac{p}{q} \cdot \frac{z}{1-z} \cdot \frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)} &= \frac{p}{q} \cdot \left(\frac{1}{1-z} - 1\right) \cdot \sum_{n \geq 0} c_n (1-z)^n \\ &= \frac{p}{q} \cdot \frac{1}{1-z} \cdot \frac{1}{1 - \lambda_{k,p}} + \sum_{n \geq 0} \alpha_n (1-z)^n, \end{aligned}$$

for some constants α_n , $n \geq 0$. It follows that

$$[z^n] \frac{p}{q} \cdot \frac{z}{1-z} \cdot \frac{1}{1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)} = \frac{p}{q} \cdot \frac{1}{1 - \lambda_{k,p}} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right),$$

for any given $\epsilon > 0$. (See e.g. Wilf [22, Theorem 5.5]) Here R^* is the radius of convergence of $\frac{1}{q}S^{(k)}(p^{k-1}qz)$, as derived in Lemma 4.2. Bringing these results together we have:

$$[z^n]M^{(k)}(z) = \left(k - \frac{1}{q}\right) n + \frac{p}{q} \cdot \frac{1}{1 - \lambda_{k,p}} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right),$$

as required.

Case $q < \frac{1}{k}$: Let us write

$$M^{(k)}(z) = \left(k - \frac{1}{q}\right) \frac{z}{(1-z)^2} + \frac{1}{z-1} H^{(k)}(z),$$

where $H^{(k)}(z) := \frac{pz}{S^{(k)}(p^{k-1}qz) - q}$. Consider first the function $H^{(k)}(z)$. We know from Lemma 4.4 that $H^{(k)}(z)$ has a pole at $z = 1$. Noting that the derivative $\frac{d}{dz}(S^{(k)}(p^{k-1}qz) - q)|_{z \leftarrow 1} \neq 0$, by Lemma 4.5, shows that this is a simple pole. Therefore we can write

$$H^{(k)}(z) = \sum_{n \geq -1} c_n (z-1)^n,$$

and now seek to determine both c_{-1} and c_0 . First notice that

$$c_{-1} = \text{Res}(H^{(k)}(z) : z = 1) = \frac{pz|_{z \leftarrow 1}}{\frac{d}{dz}(S^{(k)}(p^{k-1}qz) - q)|_{z \leftarrow 1}} = \frac{p}{\frac{pq}{1-kq}} = \frac{1}{q} - k.$$

We can now calculate

$$c_0 = \left(H^{(k)}(z) - \frac{c_{-1}}{z-1} \right) \Big|_{z \leftarrow 1}.$$

This value is

$$\begin{aligned} c_0 &= \lim_{z \rightarrow 1} \frac{pz(z-1) - \left(\frac{1}{q} - k\right) (S^{(k)}(p^{k-1}qz) - q)}{(z-1) (S^{(k)}(p^{k-1}qz) - q)} \\ &= \lim_{z \rightarrow 1} \frac{p(z-1) + pz - \left(\frac{1}{q} - k\right) \frac{d}{dz} S^{(k)}(p^{k-1}qz)}{(S^{(k)}(p^{k-1}qz) - q) + (z-1) \frac{d}{dz} S^{(k)}(p^{k-1}qz)}, \end{aligned}$$

by applying l'Hôpital's rule. As both the numerator and denominator of the last expression are zero when $z = 1$, we apply l'Hôpital's rule once again to find

$$\begin{aligned} c_0 &= \lim_{z \rightarrow 1} \frac{2p - \left(\frac{1}{q} - k\right) \frac{d^2}{dz^2} S^{(k)}(p^{k-1}qz)}{\frac{d}{dz} S^{(k)}(p^{k-1}qz) + \frac{d}{dz} S^{(k)}(p^{k-1}qz) + (z-1) \frac{d^2}{dz^2} S^{(k)}(p^{k-1}qz)} \\ &= \frac{2p - (1 - kq) \frac{pq(k-1)(2-qk)}{(1-qk)^3}}{\frac{2pq}{1-kq}} \\ &= \frac{2p(1-qk)^2 - pq(k-1)(2-qk)}{2pq(1-kq)} \\ &= \frac{2(1+q) - kq(6+q-3kq)}{2q(1-kq)}, \end{aligned}$$

wherein the second equality is the result of applying Lemma 4.6. These two values for c_0 and c_{-1} allow us to complete the asymptotic analysis for $M_n^{(k)}(p) = [z^n]M^{(k)}(z)$:

$$\begin{aligned} M_n^{(k)}(p) &= \left(k - \frac{1}{q}\right) n + [z^n] \frac{c_{-1}}{(z-1)^2} + [z^n] \frac{c_0}{z-1} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) \\ &= \left(k - \frac{1}{q}\right) n + \left(\frac{1}{q} - k\right) (n+1) - \frac{2(1+q) - kq(6+q-3kq)}{2q(1-kq)} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) \\ &= \frac{(1-k)(kq-2)}{2(1-kq)} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right), \end{aligned}$$

for a given $\epsilon > 0$, as required. Once again R^* is the radius of convergence of $\frac{1}{q}S^{(k)}(p^{k-1}qz)$, as derived in Lemma 4.2.

Case $q = \frac{1}{k}$: Rewrite equation (4) as $S^{(k)}(z) = z\phi(S^{(k)}(z))$ where $\phi(x) := (1-x)^{-(k-1)}$. Functional equations of this form have been considered by Flajolet and Sedgewick [6, Section VI.7]. The functions $\phi(x)$ and $S^{(k)}(z)$ satisfy the conditions of [6, Theorem VI.6] and, moreover, the unique positive solution to the characteristic equation $\phi(\tau) - \tau\phi'(\tau) = 0$ is $\tau = \frac{1}{k}$. Writing

$$\rho = \frac{\tau}{\phi(\tau)} = \frac{(k-1)^{k-1}}{k^k}$$

for the radius of convergence of $S^{(k)}(z)$ at 0, we therefore have that, near ρ ,

$$S^{(k)}(z) = \frac{1}{k} - \sqrt{\frac{2(k-1)}{k^3}} \sqrt{1 - \frac{z}{\rho}} + \sum_{j \geq 2} (-1)^j d_j \left(1 - \frac{z}{\rho}\right)^{j/2},$$

for some computable constants d_j . Since ρ is $p^{k-1}q$ when $q = 1/k$ we have, as $z \rightarrow 1$,

$$S^{(k)}(p^{k-1}qz) \sim \frac{1}{k} - \sqrt{\frac{2(k-1)}{k^3}} \sqrt{1-z}.$$

Replacing this into the expression for $M^{(k)}(z)$ in Theorem 3.3 we find

$$M^{(k)}(z) \sim \sqrt{\frac{k(k-1)}{2}} (1-z)^{-\frac{3}{2}}.$$

Therefore, by [6, Theorem VI.1], we have:

$$[z^n]M^{(k)}(z) \sim \sqrt{\frac{k(k-1)}{2}} \frac{\sqrt{n}}{\Gamma(3/2)},$$

where $\Gamma(3/2) = \int_0^\infty e^{-t} t^{1/2} dt = \sqrt{\pi}/2$. □

5. Diagonal state matters

A natural aspect of the big-chooser process to consider is that of returning to a diagonal state, a state for which all matchboxes have an equal number of matches. Knuth [15] made the striking observation that the expected residue in the $k = 2$ case was equal to the expected value of i , where $(n-i, n-i)$ is the first time the two matchboxes are again equal in size after the process begins. Stirzaker gave further context for this observation in his paper [20]. In this section we will consider this expected value for general k .

5.1. Expected order of first diagonal return. Let $Y := Y(k, n, p)$ be the order of the first diagonal return for the general k -matchbox process that begins in state $(n, n, \dots, n)$. That is, the first return to a diagonal state after leaving the initial configuration $(n, n, \dots, n)$ is at $(n-Y, n-Y, \dots, n-Y)$. Define the generating function

$$R^{(k)}(z) := \sum_{n \geq 1} R_n^{(k)}(p) z^n,$$

where $R_n^{(k)}(p) := \mathbb{E}(Y(k, n, p))$. The first value is clearly $R_1^{(k)}(p) = 1$.

Theorem 5.1. *For $n \geq 2$, the expected order of the first diagonal return is*

$$R_n^{(k)}(p) = n - \frac{1}{q} \sum_{i=1}^{n-1} (n-i) s_i^{(k)} (p^{k-1}q)^i.$$

Moreover, its generating function is

$$R^{(k)}(z) = \frac{z}{(1-z)^2} \left(1 - \frac{1}{q} S^{(k)}(p^{k-1}qz) \right).$$

Proof. The probability weights and numbers introduced in Section 2.2 allow us to write

$$\mathbb{P}(Y = i) = s_i^{(k)} (p^{k-1}q)^i / q$$

for all $1 \leq i \leq n-1$, while the remaining probability

$$\mathbb{P}(Y = n) = 1 - \frac{1}{q} \left(s_1^{(k)} (p^{k-1}q)^1 + s_2^{(k)} (p^{k-1}q)^2 + \dots + s_{n-1}^{(k)} (p^{k-1}q)^{n-1} \right).$$

From this it follows that

$$\begin{aligned} R_n^{(k)}(p) &= \left(\sum_{i=1}^{n-1} i \cdot s_i^{(k)} (p^{k-1}q)^i / q \right) \\ &\quad + n \left(1 - \frac{1}{q} \left(s_1^{(k)} (p^{k-1}q)^1 + s_2^{(k)} (p^{k-1}q)^2 + \dots + s_{n-1}^{(k)} (p^{k-1}q)^{n-1} \right) \right) \\ &= n - (n-1) s_1^{(k)} (p^{k-1}q) \frac{1}{q} - (n-2) s_2^{(k)} (p^{k-1}q)^2 \frac{1}{q} - \dots - s_{n-1}^{(k)} (p^{k-1}q)^{n-1} \frac{1}{q}. \end{aligned}$$

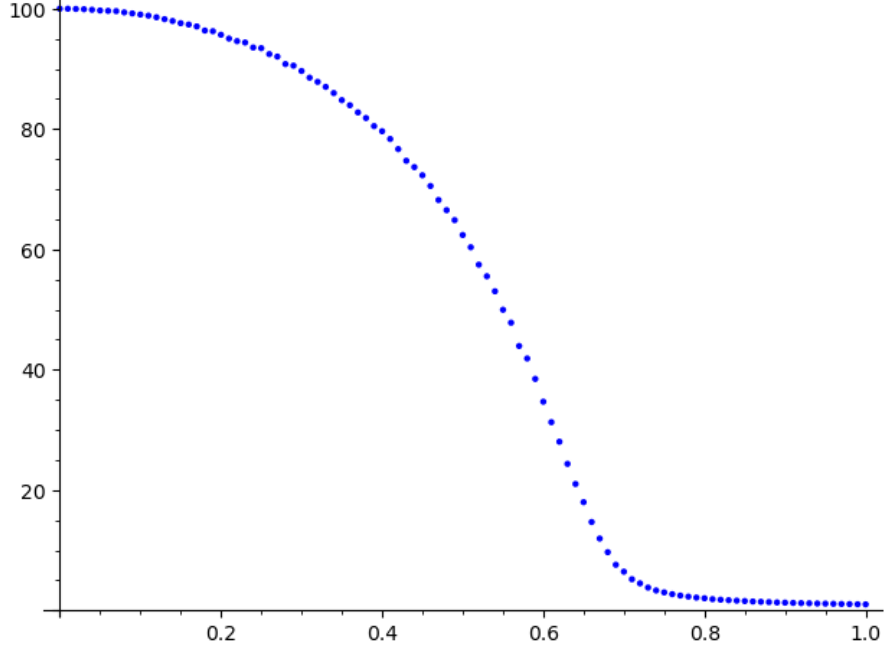


FIGURE 4. The experimental value of the random variable Y where the first return to the diagonal state is at $(n - Y, n - Y, n - Y)$ for $n = 100$.

Figure 4 illustrates the experimental value of $R_{100}^{(3)}(p)$.

The generating function $R^{(k)}(z) = \sum_{n \geq 1} R_n^{(k)}(p) z^n$ is therefore

$$\begin{aligned}
R^{(k)}(z) &= z + \sum_{n \geq 2} \left(n - (n-1) s_1^{(k)}(p^{k-1}q) \frac{1}{q} - \dots - s_{n-1}^{(k)}(p^{k-1}q)^{n-1} \frac{1}{q} \right) z^n \\
&= z + \sum_{n \geq 2} n z^n \\
&\quad - \frac{1}{q} \sum_{n \geq 2} \left((n-1) z^{n-1} \cdot s_1^{(k)}(p^{k-1}q) z + \dots + z \cdot s_{n-1}^{(k)}(p^{k-1}q)^{n-1} z^{n-1} \right) \\
&= \sum_{n \geq 1} n z^n - \frac{1}{q} \left(\sum_{n \geq 1} n z^n \right) S^{(k)}(p^{k-1}qz) \\
&= \frac{z}{(1-z)^2} - \frac{1}{q} \frac{z}{(1-z)^2} S^{(k)}(p^{k-1}qz) \\
&= \frac{z(q - S^{(k)}(p^{k-1}qz))}{q(1-z)^2}.
\end{aligned}$$

□

So although $M^{(2)}(z) = R^{(2)}(z)$, we can clearly see that this relationship does not hold for any $k > 2$. It is relatively straightforward to determine the asymptotic behaviour of $R_n(p)$.

Proposition 5.2. Let $R^* = \frac{(k-1)^{k-1}}{k^k} \cdot \frac{1}{p^{k-1}q}$ and let $\epsilon > 0$. Then

$$R_n(p) = \begin{cases} \frac{p}{1-kq} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) & \text{if } q < \frac{1}{k} \\ \sqrt{\frac{8(k-1)n}{k\pi}} + \mathcal{O}\left(\frac{1}{n^{1/2}}\right) & \text{if } q = \frac{1}{k} \\ n(1 - \lambda_{k,p}) + \frac{p^{k-1}(1 - q\lambda_{k,p})}{(1 - q\lambda_{k,p})^k - (k-1)qp^{k-1}} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right) & \text{if } q > \frac{1}{k}, \end{cases}$$

as $n \rightarrow \infty$.

Proof. As $R^{(k)}(z) = \frac{z}{(z-1)^2} \left(1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)\right)$ we see that, for $q \neq 1/k$, $R^{(k)}(z)$ only has a pole at $z = 1$, and that it is of order 2. For the case $q = 1/k$ this is not true and an algebraic singularity at $z = 1$ appears. In any case the (unique) dominant singularity is at $z = 1$.

Let $1 - \frac{1}{q}S^{(k)}(p^{k-1}qz) = \sum_{i \geq 0} c_i(z-1)^i$, and write

$$R^{(k)}(z) = \sum_{j \geq -2} \alpha_j(z-1)^j = \left(\frac{1}{(z-1)^2} + \frac{1}{z-1}\right) \sum_{i \geq 0} c_i(z-1)^i.$$

We wish to know α_{-2} and α_{-1} . We can see that $\alpha_{-2} = c_0$ and $\alpha_{-1} = c_1 + c_0$. Now,

$$c_0 = \left(1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)\right)\Big|_{z \leftarrow 1} = \begin{cases} 0 & \text{if } q < \frac{1}{k} \\ 1 - \lambda_{k,p} & \text{if } q > \frac{1}{k}, \end{cases}$$

where $\lambda_{k,p} := (1 - \frac{1}{q}S^{(k)}(p^{k-1}q))$ was defined in Lemma 4.4. Also,

$$c_1 = \frac{d}{dz} \left[1 - \frac{1}{q}S^{(k)}(p^{k-1}qz)\right]\Big|_{z \leftarrow 1} = \begin{cases} \frac{-p}{1-kq} & \text{if } q < \frac{1}{k} \\ \frac{-p^{k-1}(1 - q\lambda_{k,p})}{(1 - q\lambda_{k,p})^k - (k-1)qp^{k-1}} & \text{if } q > \frac{1}{k}, \end{cases}$$

by Lemma 4.5. Finally, let $R^* = \frac{(k-1)^{k-1}}{k^k} \cdot \frac{1}{p^{k-1}q}$ and let $\epsilon > 0$. R^* is the radius of convergence of $\frac{1}{q}S^{(k)}(p^{k-1}qz)$, as derived in Lemma 4.2. Now (e.g. using Wilf [22, Theorem 5.5]) we have

$$[z^n]R^{(k)}(z) \rightarrow [z^n]\frac{\alpha_{-2}}{(z-1)^2} + [z^n]\frac{\alpha_{-1}}{(z-1)} + \mathcal{O}\left(\left(\frac{1}{R^*} + \epsilon\right)^n\right),$$

as $n \rightarrow \infty$, from which the first and third results follow.

For the case $q = 1/k$, and using $\rho = (k-1)^{k-1}/k^k$ as we did in the proof of Theorem 4.1, as $z \rightarrow 1$ the generating function $R^{(k)}(z) = z(1 - kS^{(k)}(\rho z))/(1-z)^2$ is asymptotically

$$\begin{aligned} R^{(k)}(z) &\sim \frac{1}{(1-z)^2} \left(1 - k \left(\frac{1}{k} - \sqrt{\frac{2(k-1)}{k^3}} \sqrt{1-z}\right)\right) \\ &= \sqrt{\frac{2(k-1)}{k}} (1-z)^{-\frac{3}{2}}. \end{aligned}$$

Again, using [6, Theorem VI.1] as we did in the proof of Theorem 4.1, we have

$$[z^n]R^{(k)}(z) \sim \sqrt{\frac{2(k-1)}{k}} \frac{2\sqrt{n}}{\sqrt{\pi}},$$

as required. □

5.2. Diagonal state probabilities and manila folders. The quantity $f_n^{(k)}$ is the probability that the system will return to a diagonal state, not necessarily for the first time, after nk choosers have selected matches. We can give a closed form expression for $f_n^{(k)}$ in terms of a sum, but in order to do so we require some terminology and a lemma. We will present a closed form expression for $\text{paths}_k(n, i)$ (defined at the end of Section 2) by showing the paths that it counts are equinumerous to another seemingly forgotten combinatorial object that was introduced and enumerated by Finucan [5].

A manila folder is a filing cabinet folder that has a designated number of compartments. Each manila folder has $k - 1$ compartments so that there are k dividers including the front and back dividers. Suppose we have n such manila folders. The folders are placed in a filing cabinet, fronts facing forward (in our diagrams the front is the right-hand side). It is permitted to (repeatedly) place a folder inside a compartment of another folder.

Let $\text{Manila}_k(n, i)$ be the set of possible folder arrangements consisting of n manila folders each having $k - 1$ compartments, and with precisely $i + 1$ of the folder spines visible to a hypothetical observer looking up from below the cabinet. Let $\text{manila}_k(n, i) = |\text{Manila}_k(n, i)|$.

For example, suppose that we have $n = 2$ folders each with $k - 1 = 2$ compartments. There are 3 distinct ways of arranging the folders in a filing cabinet, see Figure 5. Two of these arrangements, (b) and (c) in Figure 5, have only one spine visible from below. Configuration (b) can be represented by the sequence $((\epsilon, \epsilon), \epsilon)$ and configuration (c) by the sequence $(\epsilon, (\epsilon, \epsilon))$, where ϵ represents an empty compartment. So $\text{manila}_3(2, 1 - 1) = \text{manila}_3(2, 0) = 2$.

The arrangement in (a) has two spines visible from below. This arrangement is two empty folders side-by-side and can be represented by the sequence $(\epsilon, \epsilon)(\epsilon, \epsilon)$. So $\text{manila}_3(2, 2 - 1) = \text{manila}_3(2, 1) = 1$.

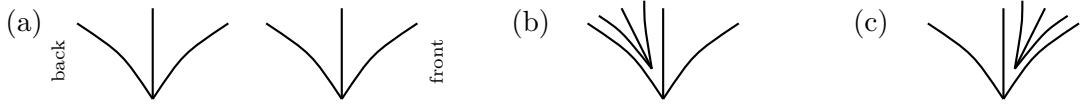


FIGURE 5. There are three ways to arrange two Manila folders, each with two compartments. $|\text{Manila}_3(2, 0)| = 2$ and $|\text{Manila}_3(2, 1)| = 1$.

Although it is intuitively obvious what $\text{Manila}_k(n, i)$ contains, we find it useful to make a more formal recursive definition.

Definition 5.3. Fix $k \in \mathbb{N}$. Then, for each $n \in \mathbb{N}$,

- M1. $\epsilon \in \text{Manila}_k(0, -1)$ (the ‘no manila folder’ element).
- M2. $(M_1, \dots, M_{k-1}) \in \text{Manila}_k(n, 0)$, where $M_j \in \text{Manila}_k(n_j, \cdot)$ for $1 \leq j \leq k - 1$, and $n_1 + \dots + n_{k-1} = n - 1$.
- M3. $M_1 M_2 \dots M_\ell \in \text{Manila}_k(n, \ell - 1)$, where $\ell > 1$, $M_j \in \text{Manila}_k(n_j, 0)$ for $1 \leq j \leq \ell$, and $n_1 + \dots + n_\ell = n$.

This recursive definition generates (in a unique way) any configuration of manila folders. The set of paths in $\text{Paths}_k(n, i)$ has a similar recursive description.

Definition 5.4. Fix $k \in \mathbb{N}$. Then, for each $n \in \mathbb{N}$,

- P1. $\epsilon \in \text{Paths}_k(0, -1)$ (the ‘no path’/‘empty path’ element).
- P2. $DP_1 LP_2 L \dots LP_{k-1} L \in \text{Paths}_k(n, 0)$, where $P_j \in \text{Paths}_k(n_j, \cdot)$ for $1 \leq j \leq k - 1$, and $n_1 + \dots + n_{k-1} = n - 1$.
- P3. $P_1 P_2 \dots P_\ell \in \text{Paths}_k(n, \ell - 1)$, where $\ell > 1$, $P_j \in \text{Paths}_k(n_j, 0)$ for $1 \leq j \leq \ell$, and $n_1 + \dots + n_\ell = n$.

This recursive definition generates all of the paths in the unit lattice that start at the boundary line $(k - 1)y = x$ and take unit south and west steps, staying weakly below that boundary line, until ending at $(0, 0)$. To see why P2 works, consider Figure 6. Notice that, for any given path P , and any pair of dotted lines (B_i, B_{i+1}) ($0 \leq i \leq k - 2$), there will necessarily be a unique

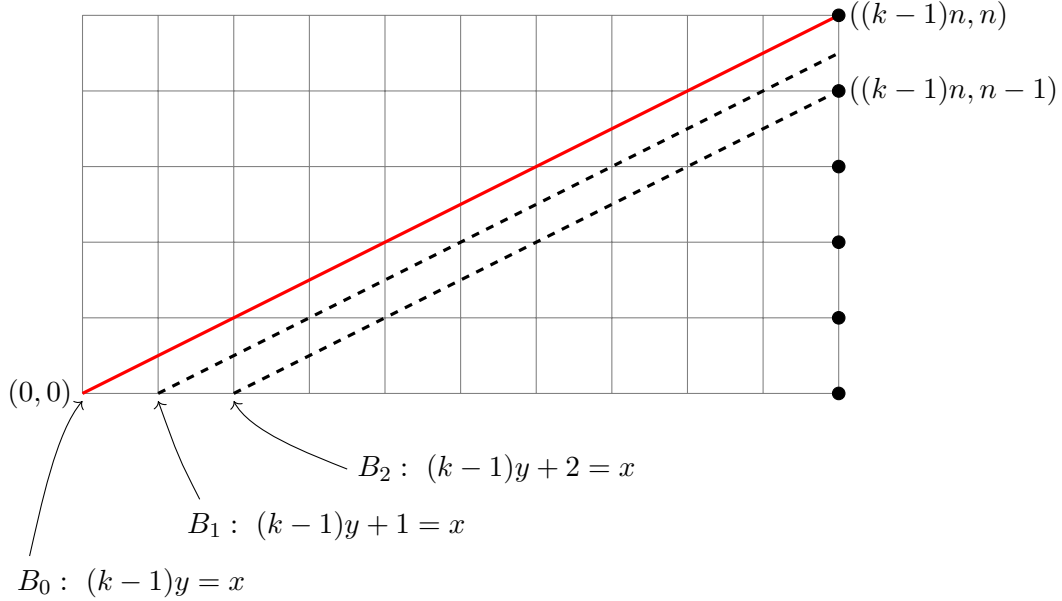


FIGURE 6. Illustration of the decomposition discussion about P2 in the paragraph following it

left step in P that takes P from B_{i+1} to B_i for the first time. This step corresponds to the $(k-1-i)^{th}$ L in the decomposition in P2.

Lemma 5.5. *The sets $\text{Manila}_k(n, i)$ and $\text{Paths}_k(n, i)$ are equinumerous.*

Proof. This follows by comparing Definitions 5.3 and 5.4. □

Equinumerosity of the sets is sufficient for our purposes but we find it interesting to take a moment and outline a bijective mapping between the two.

Consider the function $\mu : \text{Manila}_k(n, i) \rightarrow \text{Paths}_k(n, i)$ defined as follows. Given $A \in \text{Manila}_k(n, i)$, let $\mu(A)$ be the sequence derived in the following way. Label the front of each folder D (for *down*). Label each subsequent divider and the back of the folder L (for *left*). Let $\mu(A)$ be the reading of the labels on the folders read in the order that they appear from the front to the back of the cabinet. Equivalently, if using the bracketed ϵ representation of A , read it from right to left and

- for every right-parenthesis that appears append a D to $\mu(A)$,
- for every left-parenthesis and comma that appears append an L to $\mu(A)$.

See Figure 7 for an example of this labelling and the corresponding paths. For example, for the arrangement $A \in \text{Manila}_3(2, 0)$ shown in Figure 7(c), the label on the front facing folder is D . Reading the labels we encounter moving from right to left in the diagram we find $\mu(A) = DDLLLL$.

The mapping μ applied to each of the three types of manila folder configuration in Definition 5.3 results in the following paths:

- $\mu 1.$ $\mu(\epsilon) = \epsilon$, the empty path.
- $\mu 2.$ $\mu(M_1 M_2 \cdots M_\ell) = P_1 P_2 \cdots P_\ell$ where $\ell \geq 2$ and each $M_j \in \text{Manila}_k(\cdot, 0)$. The resulting path returns to the boundary line precisely $\ell - 1$ times between endpoints.
- $\mu 3.$ $\mu((M_1, \dots, M_{k-1})) = DP_1 LP_2 L \dots LP_{k-1} L$ where each $M_j \in \text{Manila}_k(\cdot, \cdot)$. The resulting path does not touch the boundary line between endpoints.

We now present a nice formula for $\text{paths}_k(n, i)$, the number of paths from $((k-1)n, n) \rightarrow (0, 0)$ that revisit the boundary line $(k-1)y = x$ precisely i times en route.

Proposition 5.6. $\text{paths}_k(n, i) = \frac{i+1}{n} \binom{kn-i-2}{n-i-1}$

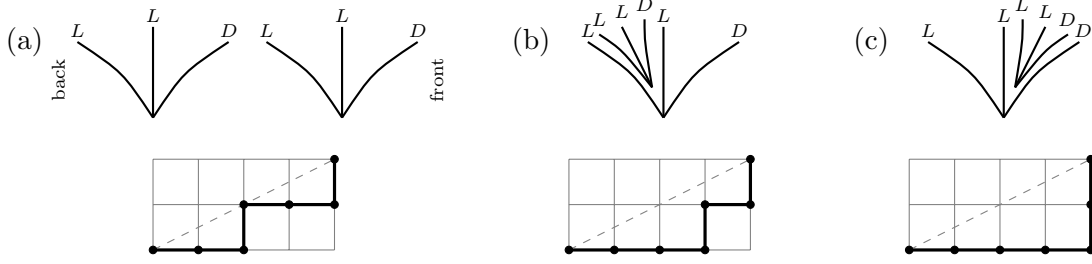


FIGURE 7. Illustration of the bijection between arrangements of Manila folders and paths through the lattice.

Proof. Finucan showed [5, §1.3] that the number of arrangements of n manila folders that each contain k compartments and for which h spines are hidden (equivalent to $n - h$ spines being visible) is $\frac{n-h}{n} \binom{nk+h-1}{h}$. He³ called this number $F(k, n, h)$. As our $\text{manila}_k(n, i)$ counts the number of arrangements of n folders, each with $(k - 1)$ compartments, and with $(i + 1)$ visible spines, we have:

$$\begin{aligned}
 \text{manila}_k(n, i) &= F(k - 1, n, n - i - 1) \\
 &= \frac{n - (n - i - 1)}{n} \binom{n(k - 1) + n - i - 2}{n - i - 1} \\
 &= \frac{i + 1}{n} \binom{kn - i - 2}{n - i - 1} \\
 &= \text{paths}_k(n, i),
 \end{aligned}$$

by Lemma 5.5. □

This allows us to give an exact expression for $f_n^{(k)}$.

Proposition 5.7. *For $n \geq 1$ we have:*

$$f_n^{(k)} = p^{(k-1)n} \sum_{i=0}^{n-1} \frac{n-i}{n} \binom{(k-1)n + i - 1}{i} q^i.$$

Proof. We can interpret $f_n^{(k)}$ using (probability) weighted lattice paths from $\text{Paths}_k(n)$. By conditioning on the number of returns to the boundary line it follows that

$$\begin{aligned}
 f_n^{(k)} &= p^{(k-1)n} \sum_{i=0}^{n-1} \text{paths}_k(n, i) q^{n-1-i} \\
 &= p^{(k-1)n} \sum_{i=0}^{n-1} \frac{i+1}{n} \binom{n(k-1) + n - i - 2}{n - (i+1)} q^{n-1-i} \\
 &= p^{(k-1)n} \sum_{j=0}^{n-1} \frac{n-1-j+1}{n} \binom{n(k-1) + n - (n-1-j) - 2}{n - (n-1-j+1)} q^j \\
 &= p^{(k-1)n} \sum_{j=0}^{n-1} \frac{n-j}{n} \binom{n(k-1) + j - 1}{j} q^j,
 \end{aligned}$$

wherein the second equality uses Proposition 5.6 □

³Finucan actually proved a more general result and this formula is a special case of his equation (4). See Finucan [5, Eqn. 4 and §4.2].

Acknowledgements

The authors thank David Stirzaker for sharing a copy of [20] with them. They also thank the anonymous referee for valuable suggestions that have improved the paper.

References

- [1] T. Cacoullos. Asymptotic Distribution for a Generalized Banach Match Box Problem. *Journal of the American Statistical Association* **62** (1967), no. 320, 1252–1257.
- [2] K. Duffy and W. M. B. Dukes. On Knuth’s generalization of Banach’s matchbox problem. *Mathematical Proceedings of the Royal Irish Academy* **104** (2004), 107–118.
- [3] A. Dvoretzky and Th. Motzkin. A Problem of Arrangements. *Duke Mathematical Journal* **14** (1947), no. 2, 305–313.
- [4] W. Feller. *An Introduction to Probability Theory and its Applications, Volume 1*. Wiley, 3rd Edition, 1968.
- [5] H. M. Finucan. *Some Decompositions of Generalised Catalan Numbers*. Lecture Notes in Mathematics, 952, from Combinatorial Mathematics IX. Proceedings, Brisbane, Australia, Springer-Verlag, 1981.
- [6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, USA, 2009.
- [7] I. M. Gessel and G. Xin. The generating function of ternary trees and continued fractions. *Electronic Journal of Combinatorics* **13** (2006), Paper R53.
- [8] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, Boston, 1994.
- [9] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford Science Publications, 2nd Edition, 1992.
- [10] F. Gutiérrez and B. Subercaseaux. Assortment Optimization For Conference Goodies With Indifferent Attendees. Preprint, arXiv:2403.03330.
- [11] P. Hilton and J. Pedersen. Catalan numbers, their generalization, and their uses. *Mathematical Intelligencer* **13** (1991), 64–75.
- [12] M. Keane, E. Kranakis, D. Krizanc, and L. Narayanan. Routing on Delay Tolerant Sensor Networks. In: Dolev, S. (eds) *Algorithmic Aspects of Wireless Sensor Networks. ALGOSENSORS 2009. Lecture Notes in Computer Science* **5804**, Springer, Berlin, Heidelberg, 2009.
- [13] T. Lengyel. Gambler’s ruin and winning a series by m games. *Annals of the Institute of Statistical Mathematics* **63** (2011), 181–195.
- [14] E. Liszewska and W. Młotkowski. Some relatives of the Catalan sequence. *Advances in Applied Mathematics* **121** (2020), 102105.
- [15] D. E. Knuth. The Toilet Paper Problem. *The American Mathematical Monthly* **91** (1984), no. 8, 465–470.
- [16] H. Prodinger. The kernel method: a collection of examples. *Séminaire Lotharingien de Combinatoire* **50** (2004), Article B50f.
- [17] G. N. Raney. Functional composition patterns and power series reversion. *Transactions of the American Mathematical Society* **94** (1960), 441–451.
- [18] K. Sen and J. Saran. On the distribution of crossings in a generalized random walk. *Journal of Statistical Planning and Inference* **7** (1983), 371–385.
- [19] W. Stadje. Asymptotic probabilities in a sequential urn scheme related to the matchbox problem. *Advances in Applied Probability* **30** (1998), no. 3, 831–849.
- [20] D. Stirzaker. A generalization of the matchbox problem. *The Mathematical Scientist* **13** (1988), 104–114.
- [21] M. Wallner. A half-normal distribution scheme for generating functions. *European Journal of Combinatorics* **87** (2020), 103138.
- [22] H. S. Wilf. *Generatingfunctionology*. CRC Press, 3rd Edition, 2005.
- [23] R. D. P. Zhou and S. H. F. Yan. The Raney numbers and $(s, s + 1)$ -core partitions. *European Journal of Combinatorics* **59** (2017), 114–121.

SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY COLLEGE DUBLIN, DUBLIN 4, IRELAND.

Email address: `mark.dukes@ucd.ie`

Email address: `andrew.mullins@ucdconnect.ie`