Two dependent random walks

On the probability of two players hitting the same digit while walking over the same string of random digits.

T.A. de Graaf

Bachelor thesis, July 2, 2013

Supervisor: Prof. Dr. W.Th.F. den Hollander

Mathematical Institute, Leiden University

Contents

1 Introduction

Suppose that two players play the following game. From a string of 100 digits drawn randomly from 1 to 10, each player randomly chooses one of the first 10 digits. Afterwards they move forward by an amount equal to the value of the digit they hit, and so on, until they are about to pass the 100th digit. The game stops there. Now, what is the probability that they end up at the same digit? According to simulations this probability is approximately 0.972. But what is the exact probability? In this bachelor thesis we search for an answer to this question.

1.1 Problem definition

Let $\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, 2, ...\}$. Let $k, N \in \mathbb{N}$ be such that $k \leq N$. Define $(X_n)_{n\in\mathbb{N}}$ to be the i.i.d. sequence such that $X_n \in \{1, ..., k\}$ for all $n \in \mathbb{N}$, drawn according to the uniform distribution. The starting position is determined by the value of X_1, namely the starting position is the X_1-th digit in the string. Define $(S_n)_{n\in\mathbb{N}}$ as the nth partial sum, $S_n := X_1 + ... + X_n,\ n \in \mathbb{N}$. This sequence keeps track of the position in our string.

Figure 1 gives a sketch of the situation. Every horizontal bar is a digit in the string. The numbers under those bars represent the position of the digits and the $X_1, X_2, ...$ above the bars are the digits on that position. Note that on the blank spots there are supposed to be digits as well. But as long as you do not hit a certain digit, it does not matter what the value of that particular digit is.

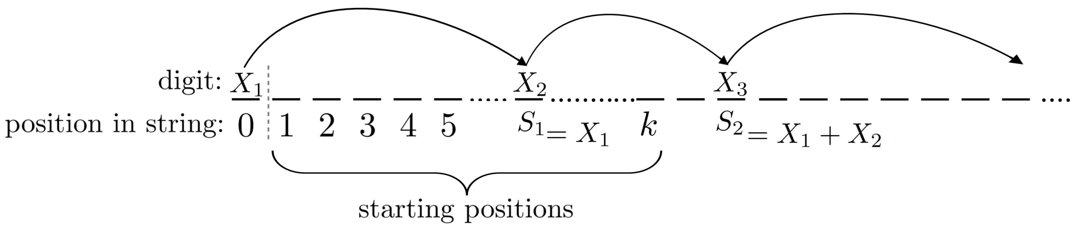

Figure 1: X_1 determines the starting position

1.2 Outline

In section 2 we review some of the basics of renewal theory and make the translation of renewal theory to our problem. In section 3 we differentiate between two models. The first model describes our initial problem, with two players walking on the same string. In the second model we consider our two players to be independent, i.e., they each walk on their own string. Then we will make the following important observation. The probabilities that they do not meet until a certain time n are the same in both models. So we can use the second model to give an answer to the question we raised for our initial problem, since in the second model the independence of the strings makes computations less complicated.

2 Renewal theory

2.1 Recursion relation

We call $(X_n)_{n\in\mathbb{N}}$ the *interarrival time sequence* and X_n the *nth interarrival time*. The sequence $(S_n)_{n\in\mathbb{N}}$, defined as above, is called the *arrival time sequence* and S_n the *nth arrival time*. Let $f = (f_n)_{n\in\mathbb{N}_0}$ be the common distribution of $(X_n)_{n\in\mathbb{N}}$:

$$f_n := \mathbb{P}(X_1 = n) = \begin{cases} \frac{1}{k} & \text{for } n = 1, ..., k, \\ 0 & \text{otherwise,} \end{cases}$$

with $f_0 := 0$. We call f the *waiting time distribution*. Let $(Z_n)_{n\in\mathbb{N}_0}$ be the sequence that denotes whether or not there occurs a renewal at time n:

$$Z_n := \begin{cases} 1 & \text{if } n = S_m \text{ for some } m \in \mathbb{N}_0, \\ 0 & \text{otherwise,} \end{cases}$$

with $Z_0 := 1$. Define u_n as the probability that a renewal occurs at time n:

$$u_n := \mathbb{P}(Z_n = 1),$$

with $u_0 := 1$. We can express $(u_n)_{n\in\mathbb{N}_0}$ in terms of $(f_n)_{n\in\mathbb{N}_0}$ via the recursion relation

$$u_n = \sum_{i=1}^{n} f_i u_{n-i}, \quad n \in \mathbb{N}.$$

For this observation we will use following definitions.

Definition 2.1 *Let $g, h : \mathbb{N}_0 \to \mathbb{R}$. The* discrete-time convolution product *of g and h is the function*

$$(g * h) : \mathbb{N}_0 \to \mathbb{R},\ j \mapsto \sum_{i=0}^{j} g(j-i)\, h(i).$$

Note that the discrete-time convolution product is associative, commutative and has the *identity element* $\delta : \mathbb{N}_0 \to \mathbb{R}$ with

$$\delta(j) := \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.2 *Let $g : \mathbb{N}_0 \to \mathbb{R}$. Define the n-*fold convolution product *of g, $g^{(n)} : \mathbb{N}_0 \to \mathbb{R}$, by*

$$\begin{aligned} g^{(0)}(j) &:= \delta(j), \\ g^{(1)}(j) &:= g(j), \\ g^{(n)}(j) &:= \underbrace{(g * g * ... * g)}_{n \text{ times}}(j), \quad n \geq 2. \end{aligned}$$

Our $X_1, ..., X_n$ are i.i.d. random variables with common distribution $f^{(1)}$. Now, $f^{(n)}$ is the distribution of $S_n = X_1 + ... + X_n$, i.e., $f_j^{(n)}$ is the probability that the $(n+1)$th renewal takes place at time j: $f_j^{(n)} = \mathbb{P}(S_n = j)$. Note that the first renewal takes place at time 0 since we have $Z_0 := 1$. We get

$$u_n = \mathbb{P}(Z_n = 1) = \mathbb{E}(Z_n) = \sum_{i=0}^{n} \mathbb{P}(S_i = n) = \sum_{i=0}^{n} f_n^{(i)}.$$

Note that this sum is finite due to the fact that

$$f_n^{(i)} = \mathbb{P}(S_i = n) = 0 \text{ for } i > n.$$

Now we are able to obtain the recursion relation connecting $(f_n)_{n\in\mathbb{N}_0}$ and $(u_n)_{n\in\mathbb{N}_0}$. For $n \in \mathbb{N}$,

$$\begin{aligned} u_n = \mathbb{P}(Z_n = 1) &= \mathbb{P}(S_1 = n) + \sum_{i=1}^{n-1} \mathbb{P}(S_1 = i)\,\mathbb{P}(Z_n = 1 | S_1 = i) \\ &= \mathbb{P}(X_1 = n) + \sum_{i=1}^{n-1} \mathbb{P}(X_1 = i)\,\mathbb{P}(Z_{n-i} = 1) \\ &= f_n + \sum_{i=1}^{n-1} f_i u_{n-i} \\ &= \sum_{i=1}^{n} f_i u_{n-i}. \end{aligned}$$

2.2 Translation to our problem

In this section we make the translation of renewal theory to our problem. For now, we consider only one player in our game.

In renewal theory X_1 denotes the length of the first interarrival time. In our case it determines the starting position of our player. Not by coincidence we choose, for all $i \in \mathbb{N}$, X_i randomly between 1 and k and also have exactly k starting positions. Now X_1 has the same distribution as $X_2, X_3, ...$, which makes computations easier.

In renewal theory we say that a renewal occurs at n when $S_n = m$ for some $m \in \mathbb{N}_0$. This corresponds in our problem to the event that the player hits digit n. Now, our interarrival time sequence $(X_n)_{n\in\mathbb{N}}$ denotes the successive lengths of our jumps and therefore determines the digits that are being hit in our string. In this way we can consider our problem as a renewal process.

Note that S_n, $n \in \mathbb{N}$, denotes the position after the nth jump. Suppose that $S_{n-1} + X_n \leq N$, but $S_n + X_{n+1} > N$. Then S_n is the end position in the string.

Definition 2.3 *Define the* generating functions *of* $(f_n)_{n\in\mathbb{N}_0}$ *and* $(u_n)_{n\in\mathbb{N}_0}$*, respectively, by*

$$F(z) := \sum_{n=0}^{\infty} f_n z^n, \quad U(z) := \sum_{n=0}^{\infty} u_n z^n, \quad z \in \mathbb{R}.$$

Lemma 2.4 *The generating functions of* $(f_n)_{n\in\mathbb{N}_0}$ *and* $(u_n)_{n\in\mathbb{N}_0}$ *are related by*

$$F(z) = \frac{U(z) - 1}{U(z)}, \quad U(z) = \frac{1}{1 - F(z)}.$$

Proof. Note that since $f_0 = 0$ and $u_n = \sum_{i=1}^{n} f_i u_{n-i}$, we have

$$\begin{aligned} U(z) &= \sum_{n=0}^{\infty} u_n z^n \\ &= u_0 + \sum_{n=1}^{\infty} \left[\sum_{i=1}^{n} f_i u_{n-i} \right] z^n \\ &= 1 + U(z)\,F(z). \end{aligned}$$

□

Lemma 2.5 *For $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} u_n = \frac{2}{k+1}.$$

Proof. Let $k \in \mathbb{N}$. By the Renewal Theorem[1] we know that

$$\lim_{n\to\infty} u_n = \frac{1}{\mathbb{E}(X_1)} = \frac{1}{\frac{1}{2}(k+1)} = \frac{2}{k+1}.$$

□

3 Two renewal processes

3.1 Link between two models

In this section we look at the case of two players. We differentiate between two models.

Definition 3.1 *We call the case of two players walking on the same string* model 1: M_1.
Definition 3.2 *We call the case of two players walking on two different strings* model 2: M_2.

Now we are at a point of formulating a theorem that is essential to us in solving our problem. This theorem makes lemma 3.4 below possible and therefore we are able to solve our problem via the path sketched in section 3.2 below.

Theorem 3.3 [The key theorem] *For $n \in \mathbb{N}$,*

$$\begin{aligned}&\mathbb{P}\left(\text{player 1 and 2 do not meet up to time } n|M_1\right)\\ &\quad= \mathbb{P}\left(\text{player 1 and 2 do not meet up to time } n|M_2\right).\end{aligned}$$

Proof. Define, for player $p = 1, 2$, $(S_n^{(p)})_{n\in\mathbb{N}}$ as before. Let $n \in \mathbb{N}$. Let $l \leq n$ be the largest digit that is being hit by one of the players, i.e., $l := \max\{S_i^{(p)} : n - S_i^{(p)} \geq 0,\ i \in \mathbb{N},\ p = 1, 2\}$. Let $r_p \in \mathbb{N}$ be such that $l = S_{r_p}^{(p)}$, $p \in \{1, 2\}$, i.e., the number of jumps of player p until it hits l. Let $r := \max\{r_1, r_2\}$.

Note that only the digits that are being hit by at least one of the players are relevant. For example, if player 1 does not hit digit $m \in \mathbb{N}$, i.e.,

$$\nexists\ m_1 \in \mathbb{N} : S_{m_1}^{(1)} = m,$$

then we are able to consider the value of digit m undetermined for player 2. This means that the value of digit m for player 2 does not necessarily have to be equal to the value of digit m for player 1. Hence

$$\begin{aligned}&\mathbb{P}(\text{player 1 and 2 do not meet up to time } n|M_1)\\ &\qquad= \mathbb{P}(\text{player 1 and 2 do not meet on } i = 1, ..., n|M_1)\\ &\qquad= \mathbb{P}(\forall\ i \in \{1, ..., r\} : S_i^{(1)} \neq S_i^{(2)}|M_1)\\ &\qquad= \mathbb{P}(\nexists\ i \in \{1, ..., r\} : S_i^{(1)} = S_i^{(2)}|M_1)\\ &\qquad= \mathbb{P}(\nexists\ i \in \{1, ..., r\} : S_i^{(1)} = S_i^{(2)}|M_2)\\ &\qquad= \mathbb{P}(\text{player 1 and 2 do not meet up to time } n|M_2).\end{aligned}$$

□

[1] See *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications*, Vlad Stefan Barbu and Nikolaos Limnios.

3.2 Two players

Since we are investigating the probability that the two players do not meet up to time n, we are able to use model 2. Define, for player $p = 1, 2$, $(Z_n^{(p)})_{n\in\mathbb{N}_0}$ as before. We define a "joint renewal" as the event that the two players hit the same digit:

$$\begin{aligned}\overline{u}_n &:= \mathbb{P}\,(\text{both players hit digit } n)\\ &= \mathbb{P}(Z_n^{(1)} = 1, Z_n^{(2)} = 1).\end{aligned}$$

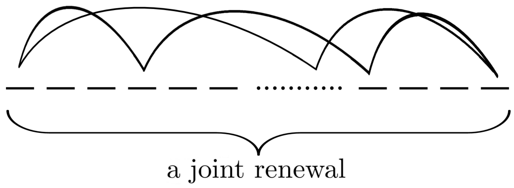

Figure 2: Two players hit the same digit

For this joint renewal process we also have

$$\overline{u}_n = \sum_{i=1}^{n} \overline{f}_i \overline{u}_{n-i}, \quad n \in \mathbb{N},$$

with $\overline{u}_0 := 0$, for an unknown distribution $\overline{f} : \mathbb{N}_0 \to \mathbb{R}$. Here $\overline{f}_i$, $i \in \mathbb{N}$, gives the probability that the two players meet for the first time on digit i. So the probability that two players do not meet up to time N is given by

$$\sum_{i>N} \overline{f}_i.$$

And therefore is the probability that two players end up at the same digit given by

$$\sum_{i=1}^{N} \overline{f}_i = 1 - \sum_{i>N} \overline{f}_i.$$

Lemma 3.4 *For $n \in \mathbb{N}$, $\overline{u}_n = u_n^2$.*

Proof. The probability to hit digit n is for both players u_n. So, under the assumption of independent strings, the probability that both players hit digit n is u_n^2. □

Define the generating functions of $(\overline{f}_n)_{n\in\mathbb{N}_0}$ and $(\overline{u}_n)_{n\in\mathbb{N}_0}$ respectively, by

$$\overline{F}\,(z) := \sum_{n=0}^{\infty} \overline{f}_n z^n, \quad \overline{U}\,(z) := \sum_{n=0}^{\infty} \overline{u}_n z^n, \quad z \in \mathbb{R}.$$

Then, analogously to the relation between $F(z)$ and $U(z)$ in lemma 2.4, we get

$$\overline{F}\,(z) = \frac{\overline{U}\,(z) - 1}{\overline{U}\,(z)}, \quad \overline{U}\,(z) = \frac{1}{1 - \overline{F}\,(z)}.$$

3.3 Structure of solution

Recall that our probability of interest is

$$\sum_{i=1}^{N} \overline{f}_i.$$

We would like to know the unknown distribution $\overline{f}$. In section 5.1 below we determine the distribution of $\overline{f}$ in the case of $k = 2$ and $N = 4, 5, 6$. Unfortunately we have not managed to do this analytically for general k and N. However we have managed to do this numerically with the help of `Maple`. We did this according to the following path:

$$f \xrightarrow{(1)} F \xrightarrow{(2)} U \xrightarrow{(3)} u \xrightarrow{(4)} \overline{u} \xrightarrow{(5)} \overline{U} \xrightarrow{(6)} \overline{F} \xrightarrow{(7)} \overline{f}.$$

The first two steps $(1), (2)$ are easy: the choice of $f = (f_n)_{n\in\mathbb{N}_0}$ determines $F(z)$, and the relation

$$U(z) = \frac{1}{1 - F(z)}$$

determines $U(z)$. Since know that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad |x| < 1,$$

we have for $|F(z)| < 1$ that

$$U(z) = \sum_{n=0}^{\infty} F(z)^n.$$

As the coefficients of the generating function $U(z)$ determine $u = (u_n)_{n\in\mathbb{N}_0}$, step (3) is now possible. But note that it might be rather difficult to compute these coefficients in closed form. How difficult this is depends on $F(z)$.

Step (4) is done by lemma 3.4. Steps (5) through (7) are possible by the generating functions $\overline{U}(z)$, $\overline{F}(z)$ and their relation. Steps (5) and (6) are easy. Step (7) is hard, again due to the fact that from $\overline{F}(z)$ it is in general not easy to deduce its coefficients $\overline{f}_i$, $i \in \mathbb{N}_0$, in closed form.

3.4 Singularities

Recall our distribution

$$f_n : \mathbb{N}_0 \to [0,1], \ n \mapsto \begin{cases} \frac{1}{k} & \text{for } n = 1, ..., k, \\ 0 & \text{otherwise.} \end{cases}$$

We get

$$F(z) = \sum_{i=1}^{k} \frac{1}{k} z^i = \frac{1}{k} \frac{z(1-z^k)}{1-z}.$$

Write $z = 1 - \varepsilon$. Then

$$\begin{aligned} F(z) &= \sum_{i=1}^{k} \frac{1}{k}(1-\varepsilon)^i \\ &= \frac{1}{k}(1-\varepsilon) + \frac{1}{k}(1-\varepsilon)^2 + \frac{1}{k}(1-\varepsilon)^2 + ... + \frac{1}{k}(1-\varepsilon)^k \\ &= \frac{1}{k}(k - \varepsilon - 2\varepsilon - 3\varepsilon - ... - k\varepsilon) + \mathcal{O}\left(\varepsilon^2\right) \\ &= \frac{1}{k}\left(k - \frac{k}{2}(k+1)\varepsilon\right) + \mathcal{O}\left(\varepsilon^2\right) \\ &= 1 - \frac{k+1}{2}\varepsilon + \mathcal{O}\left(\varepsilon^2\right), \end{aligned}$$

and hence

$$\begin{aligned} U(z) &= \frac{1}{1-F(z)} \\ &= \frac{2}{k+1}\frac{1}{\varepsilon} + \mathcal{O}\left(\frac{1}{\varepsilon^2}\right). \end{aligned}$$

Let $h : \mathbb{N}_0 \to \mathbb{R}$ be such that $u_n = \frac{2}{k+1} + h_n$ for $n \in \mathbb{N}_0$. Then

$$\begin{aligned} H(z) &:= \sum_{i=0}^{\infty} z^i h_i \\ &= U(z) - \frac{2}{k+1}\frac{1}{\varepsilon}. \end{aligned}$$

We expect h_n to converge exponentially fast to zero, since for R_k the radius of convergence of $H(z)$ we have

$$\forall\ z \in \mathbb{R} \text{ with } |z| < R_k :\ H(z) = \sum_{i=0}^{\infty} z^i h_i < \infty.$$

So we expect $h_n = (\frac{1}{R_k})^{n+o(n)}$. Let us first investigate the singularities of $H(z)$.

Lemma 3.5 *$H(z)$ has no singularity at $z = 1$.*

Proof. Let us express the Taylor polynomial of $F(z)$ around $z = 1$ as

$$\begin{aligned} F(z) &= \sum_{i=0}^{\infty} \frac{F^{(i)}(1)}{i!}(z-1)^i \\ &= 1 + F'(1)(z-1) + Q(z), \end{aligned}$$

with

$$Q(z) = \frac{F''(1)}{2}(z-1)^2 + \sum_{i=3}^{\infty} \frac{F^{(i)}(1)}{i!}(z-1)^i.$$

Then

$$\begin{aligned} H(z) &= U(z) - \frac{2}{k+1}\frac{1}{1-z} \\ &= \frac{1}{1-F(z)} - \frac{2}{k+1}\frac{1}{1-z} \\ &= \frac{F'(1)(1-z) - 1 + F(z)}{(1-F(z))F'(1)(1-z)} \\ &= \frac{Q(z)}{(1-F(z))F'(1)(1-z)} \xrightarrow{z\to 1} \frac{F''(1)}{2F'(1)^2}. \end{aligned}$$

So $H(z)$ indeed has no singularity at $z = 1$. □

Lemma 3.6 *For k odd, there is no $z \in \mathbb{R}$ for which $H(z)$ has a singularity.*

Proof. A singularity of $H(z)$ different from $z = 1$ must be a solution of $1 - F(z) = 0$, $z \neq 1$, i.e., a solution of

$$k(1-z) = z(1-z^k),$$

for $z \in \mathbb{R}$, since we are only interested in real solutions at the moment. Put $g(z) := z(1-z^k)$. Then $g(1) = 0$, $g(0) = 0$, $g'(z) = 1-(k+1)z^k$. Note that $[k(1-z)]' = -k$ and $g'(1) = -k$. There is no intersection on $[0, 1)$, since $g(z) < k(1-z)$ for $0 \leq z < 1$. For $z > 1$ we have

$$1-(k+1)z^k < -k \iff (k+1)z^k > k+1,$$

and so no intersection occurs on $(1,\infty)$ either. For k odd and $z<0$, we have $z(1-z^k)<0$, and so also no intersection occurs on $(-\infty,0)$. $\square$

Lemma 3.7 *For $k \geq 2$ even, $H(z)$ has a singularity at $z_k^* \in \mathbb{R}$, with $-2 \leq z_k^* < -1$.*

Proof. From the same reasoning as in the case of k odd, we can conclude that there is no intersection on $[0,1)$ nor on $(1,\infty)$. But $g(-1)=0$ and

$$\begin{aligned} &k(1--2) = 3k, \\ &-2(1-(-2)^k) = -2+2^{k+1}. \end{aligned}$$

We can prove by induction that, for $k \geq 2$ even, $3k \leq -2+2^{k+1}$. And since $g'(z) < -k$ for $-2 \leq z < 1$, we do have an intersection in $[-2,-1)$.

$\square$

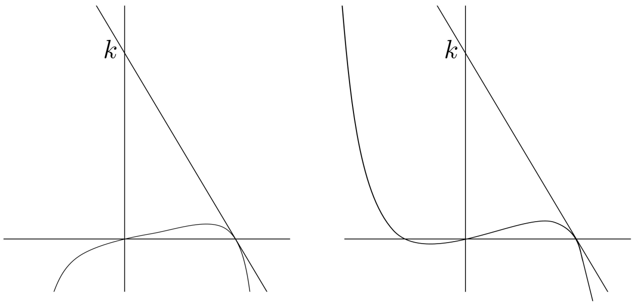

Figure 3: On the left: k is odd, on the right: k is even

It is remarkable that for k odd $H(z)$ has no singularity at any $z \in \mathbb{R}$. At first glance this could mean that for k odd h_n converges faster than any exponent to 0, while for k even h_n converges as $(\frac{1}{-z_k^*})^{n+o(n)}$. But recall that we are looking for the radius of convergence of $H(z)$. Therefore, we must also investigate the complex solutions of $F(z)=1$. Let $k \in \mathbb{N}$. Define $V_k := \{z \in \mathbb{C} : \; F(z) = 1\}$ and

$$R_k := \min_{z \in V_k \setminus \{1\}} ||z||,$$

as the radius of convergence of $H(z)$.

Lemma 3.8 *For $k \in \mathbb{N}$, $R_k > 1$.*

Proof. Let $k \in \mathbb{N}$. We have to prove that for $z \in V_k$, $z \neq 1$: $||z|| > 1$. We know that if $z \in V_k$, then $||F(z)|| = 1$. By the triangle inequality we get

$$\left|\left|\frac{1}{k}(z+...+z^k)\right|\right| \leq \frac{1}{k}\left(||z||+...+||z^k||\right) = \frac{1}{k}\left(||z||+...+||z||^k\right).$$

Put $z = re^{i\phi}$. Then $1 \leq \frac{1}{k}(r+...+r^k)$ and hence $r \geq 1$.

We know that the triangle inequality becomes an equality if and only if the angles of all $z, z^2, ..., z^k$ are equal. We know that $z = e^0$ is a solution. Put $z^j = r^j e^{j(i\phi)}$. Now

$$\left|\left|\frac{1}{k}(z+...+z^k)\right|\right| = \frac{1}{k}\left(||z||+...+||z^k||\right)$$

if and only if $j(i\phi) = 0$ for all $j \in \{1,...,k\}$, i.e., if and only if $\phi = 0$. This brings us back to the solution $z = 1$. So for all $z \in V_k$ with $||z|| = 1$, we have $z = 1$. $\square$

Table: A list of radii of convergence and real solutions of $F(z) = 1$ for different k, rounded off to 6 decimal places.

k	R_k	$-z_k^*$
2	2	2
3	$\sqrt{3}$	
4	1.556701	1.650629
5	1.445045	
10	1.218111	1.338591
11	1.197691	
20	1.107025	1.198964
21	1.101823	
100	1.021000	1.054191
101	1.020791	
200	1.010472	1.030347
201	1.010420	

Lemma 3.9 *For $k \in \mathbb{N}$, $k > 2$ even, $R_k < 2$.*

Proof. Let $k \in \mathbb{N}_{>2}$, k even. By lemma 3.7 we know that $z_k^* \in (-2, -1)$. And since by definition $R_k \leq |-z_k^*|$, we have $R_k \in (1, 2)$. □

We would have wanted lemma 3.9 for the case of k odd as well, but unfortunately we have not managed to proof this yet.

4 Asymptotic behaviour

4.1 The main theorem

Lemma 4.1 *For $k \in \mathbb{N}$,*

$$\limsup_{n\to\infty} |h_n|^{\frac{1}{n}} = \frac{1}{R_k}.$$

Proof. This is the well known Cauchy-Hadamard theorem[2] for power series. □

The main result for our initial problem is the following.

Theorem 4.2 *For $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \left(\sum_{i>n} \overline{f}_i \right)^{\frac{1}{n}} = \frac{1}{R_k}.$$

Proof. Let $k \in \mathbb{N}$. Recall that for $n \in \mathbb{N}$,

$$\overline{u}_n = \sum_{i=1}^{n} \overline{f}_i \overline{u}_{n-i}.$$

Substitute $\overline{u}_n = u_n^2 = (\frac{2}{k+1} + h_n)^2$. Then

$$\left(\frac{2}{k+1} + h_n \right)^2 = \sum_{i=1}^{n} \overline{f}_i \left(\frac{2}{k+1} + h_{n-i} \right)^2,$$

[2] See *Real Analysis*, Dipak Chatterjee.

or

$$\frac{4}{(k+1)^2}+\frac{4}{k+1}h_n+h_n^2=\sum_{i=1}^{n}\overline{f}_i\left(\frac{4}{(k+1)^2}+\frac{4}{k+1}h_{n-i}+h_{n-i}^2\right).$$

Use $\sum_{i=1}^{\infty}\overline{f}_i=1$ to write the latter equation as

$$\left(\frac{4}{(k+1)^2}+\frac{4}{k+1}h_n+h_n^2\right)\sum_{i>n}\overline{f}_i=\frac{4}{k+1}\sum_{i=1}^{n}\overline{f}_i\left(h_{n-i}-h_n\right)+\sum_{i=1}^{n}\overline{f}_i\left(h_{n-i}^2-h_n^2\right).$$

Since we are investigating the asymptotic behaviour as $n\to\infty$, we may neglect the non-leading terms. For instance, in the left-hand side $\frac{4}{k+1}h_n+h_n^2$ is negligible with respect to $\frac{4}{(k+1)^2}$ when $n\to\infty$. Similarly, in the right-hand side $h_{n-i}^2-h_n^2=(h_{n-i}-h_n)(h_{n-i}+h_n)$ is negligible with respect to $\frac{4}{k+1}(h_{n-i}+h_n)$ when $n\to\infty$. Thus, we get

$$\frac{1}{k+1}\sum_{i>n}\overline{f}_i\sim\sum_{i=1}^{n}\overline{f}_i\left(h_{n-i}-h_n\right),$$

with '$\sim$' being defined as: $a_n\sim b_n$ when $\lim_{n\to\infty}\frac{a_n}{b_n}=1$.

Lemma 4.1 says that $h_n=(\frac{1}{R_k})^{n+\varepsilon_n}$ for some $\varepsilon_n=o\left(n\right)$, i.e., $\lim_{n\to\infty}\frac{\varepsilon_n}{n}=0$. Hence

$$\begin{aligned}\sum_{i=1}^{n}\overline{f}_i\left(h_{n-i}-h_n\right)&=\sum_{i=1}^{n}\overline{f}_i\left(\left(\frac{1}{R_k}\right)^{n-i+\varepsilon_{n-i}}-\left(\frac{1}{R_k}\right)^{n+\varepsilon_n}\right)\\&=\left(\frac{1}{R_k}\right)^{n+\varepsilon_n}\sum_{i=1}^{n}\overline{f}_i\left(R_k^{i-\varepsilon_{n-i}+\varepsilon_n}-1\right).\end{aligned}$$

Claim: For some $\delta_n=o\left(n\right)$, i.e., $\lim_{n\to\infty}\frac{\delta_n}{n}=0$,

$$\sum_{i=1}^{n}\overline{f}_iR_k^{i-\varepsilon_{n-i}+\varepsilon_n}\sim\sum_{i=1}^{n}\overline{f}_iR_k^{i+\delta_n}.$$

Since $\lim_{n\to\infty}\frac{|\varepsilon_n|}{n}=0$ and $\sup_{1\le i\le n}|\varepsilon_{n-i}|=\sup_{1\le i\le n}|\varepsilon_i|<\infty$, we have

$$\lim_{n\to\infty}\frac{\sup\limits_{1\le i\le n}|\varepsilon_i|}{n}=0.$$

Put $\delta_n=2\sup_{1\le i\le n}|\varepsilon_i|$. Then $i-\varepsilon_{n-i}+\varepsilon_n\le i+\delta_n=i+2\sup_{1\le i\le n}|\varepsilon_i|$ for $i,n\in\mathbb{N}$. Hence

$$\lim_{n\to\infty}\left[\frac{1}{k+1}\sum_{i>n}\overline{f}_i\right]^{\frac{1}{n}}=\lim_{n\to\infty}\left[\left(\frac{1}{R_k}\right)^{n+\varepsilon_n}\sum_{i=1}^{n}\overline{f}_i\left(R_k^{i+\delta_n}-1\right)\right]^{\frac{1}{n}}$$

or

$$\begin{aligned}\lim_{n\to\infty}\left[\sum_{i>n}\overline{f}_i\right]^{\frac{1}{n}}&=\lim_{n\to\infty}\left(\frac{1}{R_k}\right)^{\frac{n+\varepsilon_n}{n}}\left[\sum_{i=1}^{n}\overline{f}_i\left(R_k^{i+\delta_n}-1\right)\right]^{\frac{1}{n}}\\&=\frac{1}{R_k}.\end{aligned}$$

In the last equality we use that

$$\begin{aligned}\lim_{n\to\infty}\left[\sum_{i=1}^{n}\overline{f}_i\left(R_k^{i+\delta_n}-1\right)\right]^{\frac{1}{n}} &= \lim_{n\to\infty}\left[R_k^{\delta_n}\sum_{i=1}^{n}\overline{f}_i\left(R_k^i-\frac{1}{R_k^{\delta_n}}\right)\right]^{\frac{1}{n}}\\ &= \lim_{n\to\infty}\left(R_k^{\delta_n}\right)^{\frac{1}{n}}\lim_{n\to\infty}\left[\sum_{i=1}^{n}\overline{f}_i\left(R_k^i-\frac{1}{R_k^{\delta_n}}\right)\right]^{\frac{1}{n}}\\ &= \lim_{n\to\infty}\left[\overline{F}\left(R_k\right)-\frac{1}{R_k^{\delta_n}}\overline{F}\left(1\right)\right]^{\frac{1}{n}}.\end{aligned}$$

Now $\overline{F}(1)=1$ and, to obtain the value of $\overline{F}(R_k)$, we note that by lemma 2.5, lemma 3.4 and lemma 3.8 we get

$$\lim_{n\to\infty}\overline{u}_n R_k^n=\infty.$$

So

$$\overline{F}(R_k)=1-\frac{1}{\overline{U}(R_k)}=1-\frac{1}{\sum\limits_{n=0}^{\infty}\overline{u}_n R_k^n}=1.$$

Hence

$$\lim_{n\to\infty}\left[\sum_{i=1}^{n}\overline{f}_i\left(R_k^{i+\delta_n}-1\right)\right]^{\frac{1}{n}}=\lim_{n\to\infty}\left[1-\frac{1}{R_k^{\delta_n}}\right]^{\frac{1}{n}}=1.$$

□

4.2 Examples

Let us work out the case we met in our introduction: $k=10$, $N=100$. The `Maple` code in the appendix gives us the probability

$$\sum_{i>100}\overline{f}_i\approx 0.028271.$$

To determine whether our asymptotic result is getting close, we need the radius of convergence of $H(z)$. From the table of section 3.4 we know that $R_{10}\approx 1.218111$. Now we compute

$$\left(\frac{1}{R_{10}}\right)^{100}\approx 2.7\cdot 10^{-9},$$

and so we get

$$\sum_{i>100}\overline{f}_i\approx 2.7\cdot 10^{-9}.$$

This is very far off. We would have wanted $(\frac{1}{R_{10}})^{100}\approx 0.028271$, i.e., $R_{10}\approx 1.036303$. So our asymptotic result is far from accurate.

Let us examine two other examples: $k=2$ for $N=4$ and $N=6$. We know from section 3.4 that $R_2=2$. From section 5.1 below we know that $\sum_{i>4}\overline{f}_i=0.0625$ and $\sum_{i>6}\overline{f}_i=0.015625$. We get

$$\left(\frac{1}{2}\right)^4=0.0625,\quad \left(\frac{1}{2}\right)^6=0.015625.$$

So that is exactly what we wanted.

Let us see whether this is still the case for $N=10$. Our `Maple` code gives: $\sum_{i>10}\overline{f}_i=0.0009765625$. We get

$$\left(\frac{1}{2}\right)^{10}=0.0009765625.$$

This is still the exact result that we were looking for.

5 Results

5.1 The first non-trivial case

Let us investigate the first non-trivial case, $k = 2$, for $N = 4, 5, 6$. First, it is easy to calculate the exact probability of two players meeting for the first time. Define, for player $i = 1, 2$, $(X_n^{(i)})_{n\in\mathbb{N}}$ as before. We have

$$\begin{aligned}
\overline{f}_1 &= \mathbb{P}(\text{players meet at first digit}) = \left(\frac{1}{2}\right)^2 = \frac{1}{4} \\
\overline{f}_2 &= \mathbb{P}(\text{players meet for the first time on digit 2}) \\
&= \mathbb{P}(\text{one player starts at digit 2, the other at digit 1, the first digit has value 1}) \\
&\quad + \mathbb{P}(\text{both players start at digit 2}) \\
&= \mathbb{P}(X_1^{(1)} = 1, X_1^{(2)} = 2, X_2^{(1)} = 1) + \mathbb{P}(X_1^{(1)} = 2, X_1^{(2)} = 1, X_2^{(2)} = 1) \\
&\quad + \mathbb{P}(X_1^{(1)} = 2, X_1^{(2)} = 2) \\
&= 2 \cdot \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} \\
\overline{f}_3 &= \mathbb{P}(\text{players meet for the first time on digit 3}) \\
&= \mathbb{P}(X_1^{(1)} = 1, X_2^{(1)} = 2, X_1^{(2)} = 2, X_2^{(2)} = 1) + \mathbb{P}(X_1^{(1)} = 2, X_2^{(1)} = 1, X_1^{(2)} = 1, X_2^{(2)} = 2) \\
&= 2 \cdot \left(\frac{1}{2}\right)^4 = \frac{1}{8} \\
\overline{f}_4 &= \mathbb{P}(\text{players meet for the first time on digit 4}) = ... = 2 \cdot \left(\frac{1}{2}\right)^5 = \frac{1}{16} \\
\overline{f}_5 &= \mathbb{P}(\text{players meet for the first time on digit 5}) = ... = 2 \cdot \left(\frac{1}{2}\right)^6 = \frac{1}{32}.
\end{aligned}$$

So in this particular case the probability that both players end up at the same digit, i.e., $1 - \sum_{i>N} \overline{f}_i = \sum_{i=1}^N \overline{f}_i$ is for each N easy to calculate.

Case	$\sum_{i=1}^N \overline{f}_i$
$k = 2,\ N = 4$	0.9375
$k = 2,\ N = 5$	0.96875
$k = 2,\ N = 6$	0.984375

Next, let us also analyze this case according to the path we sketched in section 3.2. We get

$$F(z) = \frac{1}{2}(z + z^2), \quad U(z) = \frac{1}{1 - \frac{1}{2}(z + z^2)}.$$

We know that, for $|x| < 1$, $\sum_{i=0}^\infty x^i = \frac{1}{1-x}$. So, for $\left|\frac{1}{2}(z + z^2)\right| < 1$, we have

$$U(z) = \sum_{i=0}^\infty \left(\frac{1}{2}(z + z^2)\right)^i.$$

This gives us $U(z) = 1 + \frac{1}{2}z + \frac{3}{4}z^2 + ...$ Define $r := \lfloor \frac{n}{2} \rfloor$ and $q := \lceil \frac{n}{2} \rceil$. With the binomium of Newton we can see that, for $n \in \mathbb{N}$,

$$u_n = \sum_{i=0}^r \binom{n - r + i}{n - q - i} \frac{1}{2^{n-r+i}}.$$

So $\overline{U}(z)=\sum_{i=0}^{\infty}\overline{u}_i z^i=\sum_{i=0}^{\infty}u_i^2 z^i$. Thus

$$\overline{F}(z)=1-\frac{1}{\overline{U}(z)}=1-\frac{1}{\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n}$$

for $|\frac{1}{2}(z+z^2)|<1$, i.e., $-2<z<1$.

We would like to have $\overline{F}(z)$ in closed form, so that we can determine the coefficients. We have not managed to do so, but what we can do is determine $\overline{f}_i$, for all $i\in\mathbb{N}_0$, by computing the Taylor series of $\overline{F}(z)$ around $z=0$:

$$\overline{F}(0)=1-\left.\frac{1}{\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n}\right|_{z=0}=0$$

$$\frac{d\overline{F}}{dz}(0)=\left.\frac{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]'}{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]^2}\right|_{z=0}=\frac{1}{4}$$

$$\begin{aligned}\frac{d^2\overline{F}}{dz^2}(0)=&\left.\frac{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]''\cdot\left[\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]^2\right]}{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]^4}\right|_{z=0}\\&-\left.\frac{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]'\cdot\left[\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]^2\right]'}{\left[\sum\limits_{n=0}^{\infty}\left[\sum\limits_{i=0}^{r}\binom{n-r+i}{n-q-i}\frac{1}{2^{n-r+i}}\right]^2 z^n\right]^4}\right|_{z=0}\\=&1\end{aligned}$$

$$\begin{aligned}\frac{d^3\overline{F}}{dz^3}(0)&=...=\frac{3}{4}\\\frac{d^4\overline{F}}{dz^4}(0)&=...=\frac{3}{2}\\\frac{d^5\overline{F}}{dz^5}(0)&=...=\frac{15}{4}.\end{aligned}$$

So we can write $\overline{F}(z)$ around $z=0$ as

$$\overline{F}(z)=\sum_{i=0}^{\infty}\frac{\overline{F}^{(i)}(0)}{i!}z^i=\frac{1}{4}z+\frac{1}{2}z^2+\frac{1}{8}z^3+\frac{1}{16}z^4+\frac{1}{32}z^5+\mathcal{O}(z^6).$$

Thus we recovered the distribution of $\overline{f}_i$, $i\in\mathbb{N}_0$, found above.

5.2 Table of results from `Maple`

With the help of the `Maple` code in the appendix, we can determine for every k and N what the probability is that both players end up at the same digit. These results are highly accurate. Unfortunately they are not exact, due to the fact that `Maple` can not work with the infinite summation of the generating function $U(z)$. However, if we add up a large number of summands, say 300, then we get a very accurate result for small N. As we see in our table, the results become less accurate when N becomes larger. We need to add up a larger number of summands to get a good result. This requires a powerful computer.

Input		Result `Maple`	Result `C++`	Absolute difference
$k = 2$,	$N = 4$	0.937500	0.937480	0.000020
$k = 2$,	$N = 5$	0.968750	0.968771	0.000021
$k = 2$,	$N = 6$	0.984375	0.984372	0.000003
$k = 10$,	$N = 100$	0.971729	0.971724	0.000005
$k = 15$,	$N = 50$	0.539373	0.539352	0.000021
$k = 20$,	$N = 100$	0.596406	0.596439	0.000033
$k = 40$,	$N = 50$	0.061067	0.061064	0.000003
$k = 100$,	$N = 500$	0.097700	0.171539	0.073839
$k = 100$,	$N = 1000$	0.097415	0.320431	0.223016

In the case of $k = 100$, $N = 1000$, we compare the results when we add up to even higher numbers. Let $M \in \mathbb{N}$ be the number of summands, i.e., $U(z) = \sum_{n=0}^{M} u_n z^n$.

M	Result `Maple`	Absolute difference with result `C++`
400	0.128270	0.192161
800	0.246537	0.073894
2000	0.320377	0.000054
4000	0.320377	0.000054

6 Appendix

6.1 C++ code

```
#include <iostream>
using namespace std;

//create a random number between 1..s
unsigned long int lehmer(long int s){
   static unsigned long long a = 2007, b = 4194301, c = 2147483647, z = b;
   if (s < 0){ s = -s; a = s; }
   z = (a + b * z) % c;
   return (z % s) + 1;
}

int main(int argc, char * argv[]){
   int N = 100;
   int k = 10;
   unsigned long int cycles = 1000000;
   unsigned long int i, j, l, equal = 0;
   unsigned long int list[N], number;

  for (i = 1; i <= cycles; i++){
     //create clean random-integer-filled list
     for (int i = 1; i <= N; i++){ list[i] = lehmer(k); }

    //pick starting places and run until we would pass the 100th digit
    j = lehmer(k);
    while (j < N){
       number = list[j];
       if (j + number <= N){ j += number; }
       else { break; }
    }
    l = lehmer(k);
    while (l < N){
       number = list[l];
       if (l + number <= N){ l += number; }
       else { break; }
    }

    //if both endpoints are the same, increment number of same endpoints
    if (l == j){ equal++; }
  }
  printf("%f\n", (double)equal/(double)cycles);
}
```

6.2 Maple code

```
> with(Statistics):
  k:= 10:
  N:= 100:
  X:= RandomVariable(DiscreteUniform(1,k)):
  f:= x->ProbabilityFunction(X,x):
  u:= proc(n) option remember;
  if n=0 then 1 else add(f(i)*u(n-i),i=1..n) fi; end:
  ubar:= n->u(n)^2:
  ubartimesz:= (n,z)->ubar(n)*z^n:
  Ubar:= z->add(ubartimesz(n,z),n=0..300):
  Fbar:= z->1-1/Ubar(z):
  evalf(series(Fbar(z),z=0,N+1),7):
  sum(%,z=1);
```