Special Proceedings: ISBN #: 978-81-950383-2-9
25th Annual Conference, 15-17 February 2023; pp151-168

# First Collision Time of Three Independent Random Walks

**Arighna Dey**[1], **Kumarjit Saha**[2], **and Anish Sarkar**[3]
[1] *Indian Statistical Institute, Kolkata, West Bengal, India*
[2] *Department of Mathematics, Ashoka University, Sonepat, Haryana, India*
[3] *Indian Statistical Institute, Delhi, India*

Received: 29 July 2023; Revised: 07 August 2023; Accepted: 10 August 2023

**Abstract**

Random walks are mathematical objects for modelling random trajectories where the future of the trajectory does not depend on the past. We take three simple random walk where the increments are distributed as $+1, -1$ valued random variables with probabilities $p$ and $1-p$. We study the expected first collision time of three such random walks. This work is an extension of the work of Coupier *et. al.* (2020) where they studied the case of $p = 1/2$.

*Key words:* Random walks; First collision time; Martingale.

**AMS Subject Classifications:** 62D05

## 1. Introduction

A random walk, denoted by RW, represents a trajectory or collection of trajectories that consists of taking successive random steps, each of which are independent and identically distributed. The most studied example of random walk is the walk on the integers $\mathbb{Z}$, which starts at an integer point and at each step moves by $+1$ or $-1$. This is known as the simple random walk (SRW). When the probabilities of moving to $+1$ and to $-1$ are identical, we call it the simple symmetric random walk (SSRW).

Random walks originate in almost all sciences quite naturally and find applications in various branches of mathematics, computer science, biology, chemistry, physics. In Physics, random walks are used to model the movement of particles in a random environment. The limiting process of the random walk yields the Brownian motion which is central to almost many predictive models. This has connected various branches of Mathematics and physics through the application of random walk.

In biological science, the genetic drift is modelled using random walks, which provide a general idea of the statistical processes involved. In physics, we can random walks to describe an ideal chains of polymers. The concepts of random work has been very crucially used in several fields such as psychology, finance, ecology. In Economics Stock market modelling and pricing are done through the Brownian motion. It is possible to describe fluctuations in the stock market with the random walk concepts. This has resulted several Nobel prizes

Corresponding Author: Anish Sarkar
Email: anish.sarkar@gmail.com

in Economics. Random walks also find application in the Google search engine algorithms, namely the page rank algorithm.

A simple way to construct the random walk is to flip a coin, and if the toss results in a HEAD, move to right by single step, whereas if the toss results in a TAIL, move to left by a single step. To define this walk formally, we take a sequence of independent random variables independent and identically distributed random variables, called the increment sequence, $\{I_k : k \in \mathbb{N}\}$ and an initial state $x \in \mathbb{Z}$. The random walk, starting from $x$, is defined as follows:

$$S_0 = x \text{ and } S_n = x + \sum_{k=1}^{n} I_k \text{ for } n \in \mathbb{N}.$$

This sequence $\{S_n : n \geq 0\}$ is called the random walk on $\mathbb{Z}$.

In this article we deal with three independent simple random walks. Therefore, we will consider three starting points. We note that if the starting positions of two random walks are of different parity, they will never be at the same position at any time point. Thus, we need to consider all starting positions of same parity. Since the intersection times and collision times will not change when we translate all the processes by same amount, we may choose the starting positions so that one random walk starts below the origin (the left random walk), one at the origin (the middle random walk) and the other above the origin (the right random walk). More precisely, we choose $a$ and $b$ positive even numbers and start the random walks at $-a, 0$ and $b$ respectively. We also consider three independent sequences of independent and identically distributed increment random variables $\Big\{I_k^{(L)} : k \geq 1\Big\}, \Big\{I_k^{(M)} : k \geq 1\Big\}$ and $\Big\{I_k^{(R)} : k \geq 1\Big\}$ with

$$\mathbb{P}\Big(I_k^{(s)} = +1\Big) = p = 1 - \mathbb{P}\Big(I_k^{(s)} = -1\Big) \tag{1}$$

where $p \in (0, 1)$ and $s \in \{L, M, R\}$. Now, we consider the random walks represented by

$$S_n^{(L)} = -a + \sum_{k=1}^{n} I_k^{(L)}, \qquad S_n^{(M)} = \sum_{k=1}^{n} I_k^{(M)} \qquad \text{and} \qquad S_n^{(R)} = b + \sum_{k=1}^{n} I_k^{(R)}.$$

By construction, these three random walks $S_n^{(L)}$, $S_n^{(M)}$ and $S_n^{(R)}$, starting from $-a$, 0 and $+b$ respectively, are independent. We define the first collision time of these three random walks by

$$\tau_c = \inf\Big\{n \geq 1 : (S_n^{(M)} - S_n^{(L)})(S_n^{(R)} - S_n^{(M)})(S_n^{(L)} - S_n^{(R)}) = 0\Big\}. \tag{2}$$

In this article we compute the expectation of $\tau_c$. Coupier et. al. (2020) studied the behavior of $\tau_c$ in the case of simple symmetric random walks, i.e., the increment random variables are distributed as random variables taking values $+1$ with probability $\frac{1}{2}$ and $-1$ with probability $\frac{1}{2}$. We extend the result of Coupier et. al. (2020) for any value of $p \in (0, 1)$.

## 2. Collision of two random walks

In Spitzer (1964) it is shown that the first hitting time of a random walk to a state where increment random variables are independent and identically distributed having mean 0 and finite variance is finite almost surely.

We observe that the expectation of the increment random variables and the expectation of the square of the increment random variables are given by : for $s \in \{L, M, R\}$,

$$\mathbb{E}\Big(I_k^{(s)}\Big) = p - (1-p) = 2p - 1 \text{ and}$$
$$\mathbb{E}\left(\Big(I_k^{(s)}\Big)^2\right) = p + 1 - p = 1.$$

Therefore, we have

$$\mathbb{V}\text{ar}\Big(I_k^{(s)}\Big) = \mathbb{E}\left(\Big(I_k^{(s)}\Big)^2\right) - \Big(\mathbb{E}\Big(I_k^{(s)}\Big)\Big)^2 = 4p(1-p).$$

In our particular case, we consider first the collision times of the left random walk and the middle random walk, i.e., set

$$\tau_{L,M} = \inf\Big\{n \geq 1 : S_n^{(L)} = S_n^{(M)}\Big\} = \inf\Big\{n \geq 1 : S_n^{(L)} - S_n^{(M)} = 0\Big\}. \tag{3}$$

Similarly, we may define the first collision time of the middle random walk and the right random walk by

$$\tau_{M,R} = \inf\Big\{n \geq 1 : S_n^{(M)} = S_n^{(R)}\Big\} = \inf\Big\{n \geq 1 : S_n^{(M)} - S_n^{(R)} = 0\Big\}. \tag{4}$$

We consider the collision time of the left and the middle random walk. We set the difference of the two walks by

$$X_n = S_n^{(M)} - S_n^{(L)} \tag{5}$$

for all $n \geq 0$. Similarly set

$$Y_n = S_n^{(R)} - S_n^{(M)} \tag{6}$$

for all $n \geq 0$. Hence, we observe that $X_0 = a$ and $Y_0 = b$.

We may now rephrase the first collision time of two random walks as follows:

$$\tau_{L,M} = \inf\Big\{n \geq 1 : X_n = 0\Big\} \quad \text{and} \quad \tau_{M,R} = \inf\Big\{n \geq 1 : Y_n = 0\Big\}. \tag{7}$$

We observe that, for $n \geq 1$,

$$X_n = S_n^{(M)} - S_n^{(L)} = a + \sum_{k=1}^{n} \Big[I_k^{(M)} - I_k^{(L)}\Big] = a + \sum_{k=1}^{n} D_k^{(M,L)}$$

where $D_k^{(M,L)} = I_k^{(M)} - I_k^{(L)}$ for any any $k \geq 1$. Note that $\mathbb{E}\Big(D_k^{(M,L)}\Big) = \mathbb{E}\Big(I_k^{(M)}\Big) - \mathbb{E}\Big(I_k^{(M)}\Big) = 0$ and $\mathbb{V}\text{ar}\Big(D_k^{(M,L)}\Big) = \mathbb{V}\text{ar}\Big(I_k^{(M)}\Big) + \mathbb{V}\text{ar}\Big(I_k^{(L)}\Big) = 8p(1-p)$. Thus, it is clear that the difference process $\{X_n : n \geq 0\}$ can also be be presented as a random walk with increments having mean 0 with finite variance. Therefore, using the result of Spitzer (1964), we may conclude that is finite almost surely. However, we will provide a direct argument and will actually compute the generating function of the collision time of the middle random walk and left random walk.

**Theorem 1:** Under the Assumption, we have

$$\tau_{L,M} < +\infty \text{ almost surely.}$$

Note that there is nothing special about the middle and left random walks. The result may be applied to any pair of random walks. So, as a corollary, we also have

**Corollary 1:** Under the Assumption, we have

$$\tau_{M,R} < +\infty \text{ almost surely.}$$

We will prove the result using martingale method. The method is inspired by the results in Williams (1991).Let us define the filtration $\left\{\mathcal{F}_n^{(M,L)} : n \geq 0\right\}$, where

$$\mathcal{F}_n^{(M,L)} = \sigma\left(I_k^{(L)}, I_k^{(M)} : k \leq n\right) = \sigma\left(S_k^{(L)}, S_k^{(M)} : k \leq n\right)$$

is the $\sigma$-algebra generated by the increment random variables of the middle random walk and the left random walk up to time $n$. Also, this is same as the $\sigma$-algebra generated by the middle random walk and the left random walk up to time $n$. This is the natural filtration associated with two random walks we are studying.

We have already observed that

$$X_n = a + \sum_{k=1}^{n}\left(I_k^{(M)} - I_k^{(L)}\right)$$

for $n \geq 0$. The random variables $\{I_k^{(M)} - I_k^{(L)} : k \geq 1\}$ is a sequence of independently and identically distributed random variables with common distribution being the same as of a random variable taking values $+2$ with probability $p(1-p)$, $-2$ with probability $p(1-p)$ and 0 with probability $1 - 2p(1-p)$. Let us set $\alpha = p(1-p)$.

For $\lambda \in \mathbb{R}$, let us define, the Laplace transform of the common increment distribution by

$$f(\lambda) = \mathbb{E}\left[\exp\left(-\lambda\left(I_1^{(M)} - I_1^{(L)}\right)\right)\right] = \alpha\left(e^{2\lambda} + e^{-2\lambda}\right) + (1 - 2\alpha). \tag{8}$$

Clearly, we have

$$f(\lambda) = \alpha\left(e^{2\lambda} + e^{-2\lambda} - 2\right) + 1 = \alpha\left(e^{\lambda} - e^{-\lambda}\right)^2 + 1.$$

This implies that $f(\lambda) > 1$ for $\lambda \in \mathbb{R}$ and $f(\lambda) = 1$ for $\lambda = 0$. Also, by continuity of $f$ at 0, $f(\lambda) \downarrow 1$ as $\lambda \to 0$.

Let us define, for $n \geq 0$,

$$Z_n = \exp\left(-\lambda X_n\right)\left(f(\lambda)\right)^{-n}. \tag{9}$$

We first show

**Proposition 1:** The sequence $\{Z_n : n \geq 0\}$ is an $\mathcal{F}_n^{(M,L)}$-martingale.

**Proof:** Clearly $Z_0 = \exp\Big(-\lambda X_0\Big) = \exp(-\lambda a)$. We observe that the $X_n$ is $\mathcal{F}_n^{(M,L)}$ adapted by definition. Since $Z_n$ is a measurable function of $X_n$, $Z_n$ is also $\mathcal{F}_n^{(M,L)}$ adapted. It is easy to check that for each $n \geq 0$, we have $|Z_n| \leq \exp\Big(|\lambda|(a+n)\Big)$ and hence $\mathbb{E}(|Z_n|) < \infty$ for all $n \geq 1$.

Now, to show $\{Z_n : n \geq 0\}$ is a martingale with respect to $\mathcal{F}_n^{(M,L)}$, we note that $X_n$ is measurable with respect to $\mathcal{F}_n^{(M,L)}$. We have

$$\begin{aligned}
&\mathbb{E}\Big(Z_{n+1} \mid \mathcal{F}_n^{(M,L)}\Big) \\
&= \mathbb{E}\Big[\exp\left(-\lambda X_{n+1}\right)\Big(f(\lambda)\Big)^{-n-1} \mid \mathcal{F}_n^{(M,L)}\Big] \\
&= \mathbb{E}\Big[\exp\Big(-\lambda\Big(X_n + I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big)\Big(f(\lambda)\Big)^{-n-1} \mid \mathcal{F}_n^{(M,L)}\Big] \\
&= \exp\Big(-\lambda X_n\Big)\Big(f(\lambda)\Big)^{-n-1}\mathbb{E}\left[\exp\Big(-\lambda\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big)\right] \\
&= \exp\Big(-\lambda X_n\Big)\Big(f(\lambda)\Big)^{-n-1} f(\lambda) = \exp\Big(-\lambda X_n\Big)\Big(f(\lambda)\Big)^{-n} = Z_n.
\end{aligned}$$

This completes the proof of the proposition. □

Now we prove Theorem 1.

**Proof:** We note that

$$\{\tau_{L,M} = n\} = \{X_0 = a > 0, X_1 > 0, \ldots, X_{n-1} > 0, X_n = 0\}$$

and hence $\{\tau_{L,M} = n\} \in \mathcal{F}_n^{(M,L)}$. Thus, $\tau_{L,M}$ is a stopping time relative to $\{\mathcal{F}_n^{(M,L)}\}$. Hence, the family $\Big\{Z_{n\wedge\tau_{L,M}} : n \geq 0\Big\}$ is also a $\mathcal{F}_n^{(M,L)}$-martingale. Therefore, we obtain

$$\begin{aligned}
&\mathbb{E}\Big(\exp(-\lambda X_{n\wedge\tau_{L,M}})\Big(f(\lambda)\Big)^{n\wedge\tau_{L,M}}\Big) = \mathbb{E}\left(Z_{n\wedge\tau_{L,M}}\right) \\
&= \mathbb{E}\left(Z_{0\wedge\tau_{L,M}}\right) = \mathbb{E}\left(Z_0\right) = \exp(-\lambda a). \qquad (10)
\end{aligned}$$

Now, we specialize to the case of $\lambda > 0$ and take limit as $n \to \infty$ in equation (10). We have already noted that $f(\lambda) > 1$ for $\lambda \in \mathbb{R}$, in particular for $\lambda > 0$.

- On the event $\{\tau_{L,M} = +\infty\}$, clearly $\Big(f(\lambda)\Big)^{-n\wedge\tau_{L,M}} \to 0$ as $n \to \infty$.
- On the event $\{\tau_{L,M} < \infty\}$, we have $X_{n\wedge\tau_{L,M}} \to X_{\tau_{L,M}} = 0$. Thus, $\exp(-\lambda X_{n\wedge\tau_{L,M}}) \to 1$ as $n \to \infty$ and $\Big(f(\lambda)\Big)^{-n\wedge\tau_{L,M}} \to \Big(f(\lambda)\Big)^{-\tau_{L,M}}$ as $n \to \infty$.

Combining, we have

$$\exp\left(-\lambda X_{n\wedge\tau_{L,M}}\right)\left(f(\lambda)\right)^{-\left(n\wedge\tau_{L,M}\right)} \to \mathbb{I}\left(\tau_{L,M} < \infty\right)\left(f(\lambda)\right)^{-\tau_{L,M}}$$

as $n \to \infty$. Further, we observe that

- For all $n \geq 0$, $X_{n\wedge\tau_{L,M}} \geq 0$. For $\lambda > 0$, this implies that
$$\exp\left(-\lambda X_{n\wedge\tau_{L,M}}\right) \leq 1.$$
- Since $f(\lambda) > 1$ for $\lambda > 0$ and $n \geq 0$, we have
$$(f(\lambda))^{-(n\wedge\tau_{L,M})} \leq 1.$$

Thus, we have
$$\exp(-\lambda X_{n\wedge\tau_{L,M}})\Big(f(\lambda)\Big)^{n\wedge\tau_{L,M}} \leq 1.$$
Thus, we can use DCT in equation (10) to obtain, for all $\lambda > 0$,
$$\mathbb{E}\bigg(\mathbb{I}\left(\tau_{L,M} < \infty\right)\Big(f(\lambda)\Big)^{-\tau_{L,M}}\bigg) = \exp(-\lambda a). \tag{11}$$

Now, we will take limit by letting $\lambda \downarrow 0$ in equation (11). On the event $\{\tau_{L,M} < \infty\}$, using continuity of $f$, we get $\Big(f(\lambda)\Big)^{-\tau_{L,M}} \to 1$ as $\lambda \downarrow 0$. Therefore, we have
$$\mathbb{I}\left(\tau_{L,M} < \infty\right)\Big(f(\lambda)\Big)^{-\tau_{L,M}} \to \mathbb{I}\left(\tau_{L,M} < \infty\right).$$
Furthermore, we have
$$\mathbb{I}\left(\tau_{L,M} < \infty\right)\Big(f(\lambda)\Big)^{-\tau_{L,M}} \leq 1$$
as $f(\lambda) > 1$ for any $\lambda > 0$. Thus, by apply DCT in (11), we have
$$\begin{aligned}\mathbb{P}\Big(\tau_{L,M} < \infty\Big) &= \mathbb{E}\Big(\mathbb{I}\left(\tau_{L,M} < \infty\right)\Big)\\ &= \lim_{\lambda\downarrow 0}\mathbb{E}\bigg(\mathbb{I}\left(\tau_{L,M} < \infty\right)\Big(f(\lambda)\Big)^{-\tau_{L,M}}\bigg)\\ &= \lim_{\lambda\downarrow 0}\exp(-\lambda a) = 1.\end{aligned}$$
This proves that $\tau_{L,M} < \infty$ with probability 1. □

The result in (11) yields more information. Indeed, we may calculate the probability generating function of $\tau_{L,M}$, in in turn provides more information.

**Corollary 2:** The probability generating function of $\tau_{L,M}$ is given by
$$\mathbb{E}\Big(s^{\tau_{L,M}}\Big) = \frac{1}{(2\sqrt{\alpha})^a}\left[\sqrt{\frac{1}{s} - 1 + 4\alpha} - \sqrt{\frac{1}{s} - 1}\right]^a \tag{12}$$
for $-1 < s \leq 1$.

**Proof:** Since $\tau_{L,M} < \infty$ almost surely, we can rewrite equation (11), for all $\lambda > 0$
$$\mathbb{E}\left((f(\lambda))^{-\tau_{L,M}}\right) = \exp(-\lambda a).$$

This formula may be used to get the probability generating function of $\tau_{L,M}$. Letting $s = \big(f(\lambda)\big)^{-1}$ for $\lambda > 0$ and solving $\lambda$ in terms of $s$, we have

$$\mathbb{E}\left(s^{\tau_{L,M}}\right) = \exp(-\lambda a) = \frac{1}{(2\sqrt{\alpha})^a}\left[\sqrt{\frac{1}{s} - 1 + 4\alpha} - \sqrt{\frac{1}{s} - 1}\right]^a.$$

This proves the corollary. □

This may be used to show that the expectation is infinite. Indeed, we have

$$\frac{d}{ds}\mathbb{E}\left(s^{\tau_{L,M}}\right) = \frac{a}{(2\sqrt{q})^a}\left[\sqrt{\frac{1}{s} - 1 + 4q} - \sqrt{\frac{1}{s} - 1}\right]^{a-1} \times \frac{1}{2s^2}\left[\frac{1}{\sqrt{\frac{1}{s} - 1}} - \frac{1}{\sqrt{\frac{1}{s} - 1 + 4q}}\right].$$

So, when $s \uparrow 1$, the right hand side diverges to $\infty$. Thus, $\mathbb{E}\left(\tau_{L,M}\right) = \infty$. Similarly we can also prove that $\mathbb{E}\left(\tau_{M,R}\right) = \infty$. We may also obtain the tail behaviour of the stopping time.

## 3. Collision time of three random walks : simulation

Before we go into the theoretical derivation, we carry out some simulation studies. Here we use a cutoff, to stop the process the process if the the simulation has not resulted in a value. Our cutoff is 10000000 and we have simulated for 10000000 times. We have also taken different values of $a$ and $b$ where $a$ and $b$ are both even positive integers. We have carried out the simulation using 3 different values of $p$, which are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{5}{7}$ respectively.

For $p = \frac{1}{2}$, $y_1$ is the observed mean of the first collision time of three random walks after simulating it 10000000 times, For $p = \frac{1}{3}$, $y_2$ is the observed mean of the first collision time of three random walks after simulating it 10000000 times, For $p = \frac{5}{7}$, $y_3$ is the observed mean of the first collision time of three random walks after simulating it 10000000 times. Now we will look at the scatter plots of $(ab,y_1)$, $(ab,y_2)$ and $(ab,y_3)$ and also we will find and plot regression lines of $y_1$ on $ab$, $y_2$ on $ab$ and $y_3$ on $ab$. Here $S_0^{(L)}$, $S_0^{(M)}$ and $S_0^{(R)}$ are $-a$, 0 and $+b$ respectively.

**Simulation output**

**Table 1: Simulation of expected collision times**

| $-a$ | $+b$ | $y_1$ | $y_2$ | $y_3$ | $ab$ |
|---|---|---|---|---|---|
| -2 | 2 | 3.9987 | 4.4956 | 4.8801 | 4 |
| -2 | 4 | 7.9961 | 9.0174 | 9.7983 | 8 |
| -2 | 6 | 12.0102 | 13.4858 | 14.6516 | 12 |
| -2 | 8 | 15.9895 | 17.9811 | 19.6139 | 16 |
| -2 | 10 | 20.0246 | 22.5134 | 24.4733 | 20 |
| -2 | 12 | 24.0198 | 27.0171 | 29.3998 | 24 |
| -2 | 14 | 28.0139 | 31.4881 | 34.3256 | 28 |
| -2 | 16 | 31.9907 | 35.9944 | 39.2114 | 32 |
| -2 | 18 | 35.9821 | 40.4913 | 44.0897 | 36 |

**Table 1: Simulation of expected collision times**

| $-a$ | $+b$ | $y_1$ | $y_2$ | $y_3$ | $ab$ |
|---|---|---|---|---|---|
| -2 | 20 | 40.0912 | 44.9591 | 48.9771 | 40 |
| -4 | 4 | 15.9931 | 17.5619 | 19.5812 | 16 |
| -4 | 6 | 23.9978 | 27.0127 | 29.3665 | 24 |
| -4 | 8 | 31.9914 | 36.0223 | 39.1997 | 32 |
| -4 | 10 | 40.0297 | 45.0136 | 49.0315 | 40 |
| -4 | 12 | 47.9956 | 53.9889 | 58.7969 | 48 |
| -4 | 14 | 55.9992 | 62.9156 | 68.5899 | 56 |
| -4 | 16 | 63.9958 | 71.9929 | 78.3878 | 64 |
| -4 | 18 | 72.0154 | 81.0147 | 88.2156 | 72 |
| -4 | 20 | 80.0083 | 90.0396 | 97.9089 | 80 |
| -6 | 6 | 35.9841 | 40.5069 | 44.0989 | 36 |
| -6 | 8 | 47.9892 | 53.9574 | 58.7899 | 48 |
| -6 | 10 | 59.9946 | 67.4998 | 73.5017 | 60 |
| -6 | 12 | 71.9839 | 80.9758 | 88.1898 | 72 |
| -6 | 14 | 84.0629 | 94.5195 | 102.8761 | 84 |
| -6 | 16 | 95.9779 | 108.0251 | 117.6112 | 96 |
| -6 | 18 | 108.0022 | 121.5245 | 132.2893 | 108 |
| -6 | 20 | 119.9141 | 134.9596 | 146.9674 | 120 |
| -8 | 8 | 63.9951 | 72.0212 | 78.3894 | 64 |
| -8 | 10 | 79.9917 | 90.0018 | 97.9825 | 80 |
| -8 | 12 | 95.9679 | 107.9786 | 117.5997 | 96 |
| -8 | 14 | 112.0091 | 125.9925 | 137.2119 | 112 |
| -8 | 16 | 127.9899 | 143.9512 | 156.7898 | 128 |
| -8 | 18 | 143.9769 | 161.9213 | 176.2996 | 144 |
| -8 | 20 | 160.0998 | 179.9621 | 196.0176 | 160 |
| -10 | 10 | 100.0518 | 112.5185 | 122.4886 | 100 |
| -10 | 12 | 119.9371 | 135.0121 | 147.0259 | 120 |
| -10 | 14 | 139.9145 | 157.4852 | 171.4966 | 140 |
| -10 | 16 | 159.9159 | 179.9597 | 195.9979 | 160 |
| -10 | 18 | 180.0263 | 202.5096 | 220.3999 | 180 |
| -10 | 20 | 199.9564 | 224.9917 | 244.9732 | 200 |
| -12 | 12 | 143.9768 | 161.9129 | 176.3993 | 144 |
| -12 | 14 | 168.0459 | 189.0432 | 205.7915 | 168 |
| -12 | 16 | 192.0091 | 216.0278 | 235.2112 | 192 |
| -12 | 18 | 215.9316 | 242.9841 | 264.5889 | 216 |
| -12 | 20 | 239.9089 | 269.9124 | 294.0113 | 240 |
| -14 | 14 | 195.9989 | 220.5398 | 240.1376 | 196 |
| -14 | 16 | 223.9388 | 251.9492 | 274.2998 | 224 |
| -14 | 18 | 251.9164 | 283.4919 | 308.6779 | 252 |
| -14 | 20 | 279.9936 | 315.0154 | 342.9547 | 280 |
| -16 | 16 | 255.9989 | 287.9754 | 313.6291 | 256 |
| -16 | 18 | 287.9669 | 324.0478 | 352.7959 | 288 |
| -16 | 20 | 319.9799 | 359.9954 | 391.9286 | 320 |

**Table 1: Simulation of expected collision times**

| $-a$ | $+b$ | $y_1$ | $y_2$ | $y_3$ | $ab$ |
|---|---|---|---|---|---|
| -18 | 18 | 323.9193 | 364.3991 | 396.8777 | 324 |
| -18 | 20 | 359.9899 | 404.9145 | 441.1223 | 360 |
| -20 | 20 | 399.9918 | 449.7982 | 489.8979 | 400 |

The scatter plots of the above data is very instructive as they clearly bring out the relation between $ab$ and the expected time of the first collision time $\tau_c$.

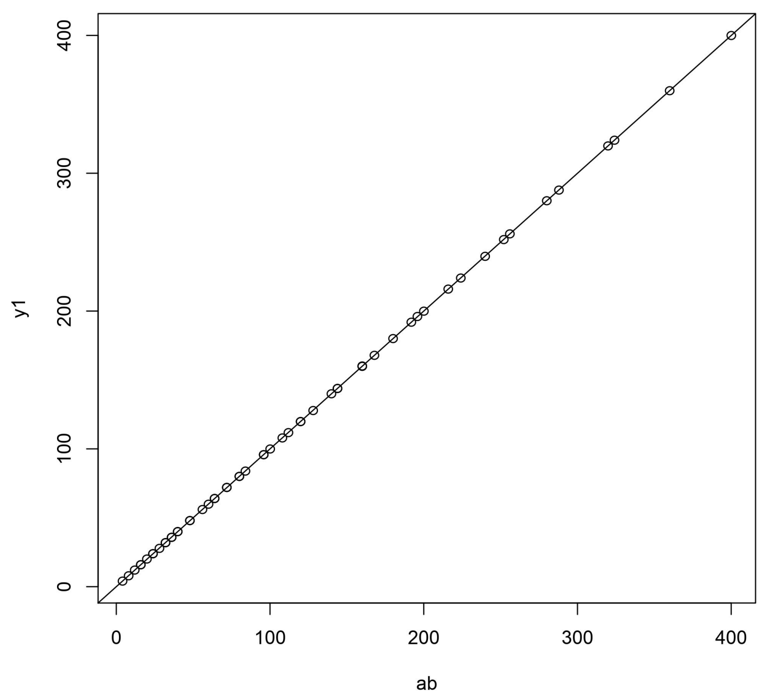


**Figure 1: Scatter plot of ($ab$,$y_1$) and regression line of $y_1$ on $ab$**

**Table 2: Summary statistics of simulation**

| Statistics | Estimate | T statistics | P value |
|---|---|---|---|
| $Constant_1$ | 0.00727396818 | 0.8787564386 | 0.3835000647 |
| $Slope_1$ | 0.9998608551 | 19192.1238453652 | 0 |
| $Constant_2$ | -0.0096277426 | -0.6188423746 | 0.5386710900 |
| $Slope_2$ | 1.1248997399 | 11488.2891168571 | 0 |
| $Constant_3$ | -0.0077084624 | -0.9309990213 | 0.3560755895 |
| $Slope_3$ | 1.2249623703 | 23506.6195548538 | 0 |

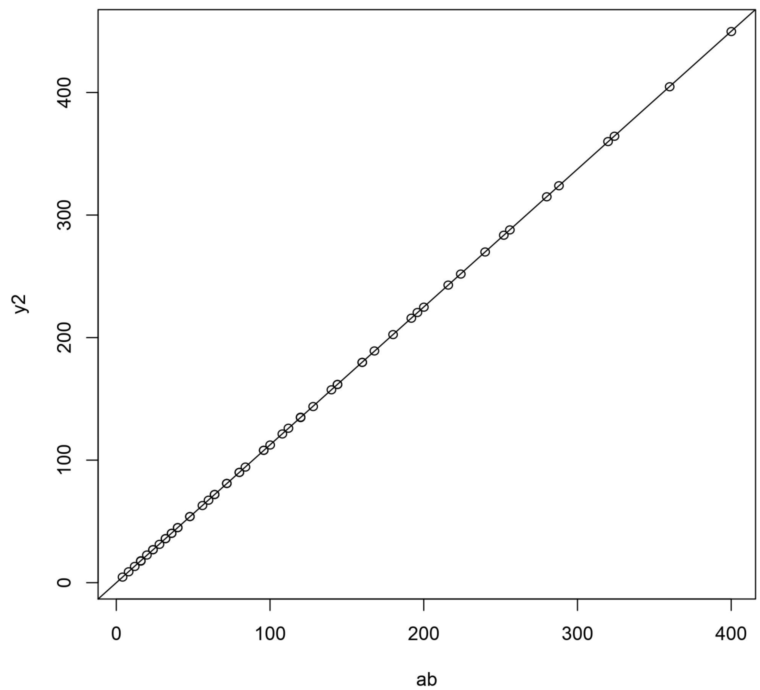


**Figure 2: Scatter plot of ($ab$,$y_2$) and regression line of $y_2$ on $ab$**

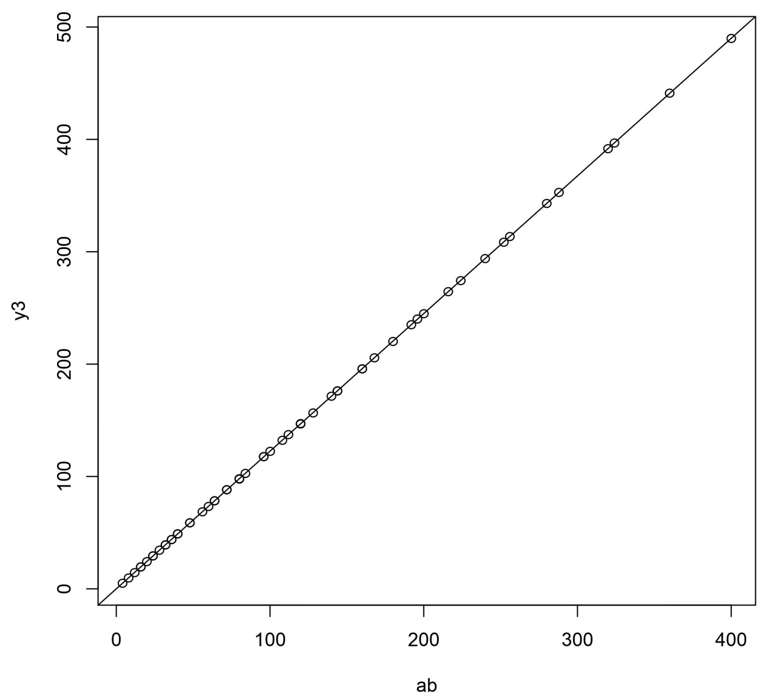


**Figure 3: Scatter plot of ($ab$,$y_3$) and regression line of $y_3$ on $ab$**

The regression lines on $ab$ are for different values of $p$:

$$\widehat{y_1} = 0.00727396818 + 0.9998608551 \times ab$$
$$\widehat{y_2} = -0.0096277426 + 1.1248997399 \times ab$$
$$\widehat{y_3} = -0.0077084624 + 1.2249623703 \times ab.$$

The correlation coefficients are 0.9999999281, 0.9999997992, 0.9999999952 respectively. In each of the three cases the correlation coefficient is very close to +1, so here we can observe near perfect positive correlation.

The summary statistics of the above data, which from the above scatter plots is quite expected, clearly shows that there should be a linear relationship between the expected time and the product of the initial distances $ab$. In each of the three cases the estimate of the constant is very close to 0 and the estimate of the slope is very close to $\Big(4p(1-p)\Big)^{-1}$. Also in each of the three cases the p-value of the intercept is greater than 0.05, so the intercept is not significant. From these observations we postulate that the expectation of $\tau_c$ should be $ab\Big(4p(1-p)\Big)^{-1}$. In the next section we derive these theoretical results.

## 4. Theoretical results

We first note that we are working with random walks having steps size of $\pm 1$ with the starting points are on even lattice. Therefore, these independent random walks do not cross each other before intersecting. So, we can write the first collision time of these three random walks $\tau_c$ as,

$$\tau_c = \min\Big\{\tau_{L,M}, \tau_{M,R}\Big\}. \tag{13}$$

As an immediate consequence of Theorem 1, we have

$$\tau_c < +\infty \text{ with probability 1.}$$

Further from the above observation, it is easy to conclude that at $\tau_c$ either the pair of left random walk and the middle random walk collides or the pair of middle random walk and the right random walk collides. So, we can rephrase the definition of $\tau_c$ (see equation (2)) as follows:

$$\begin{aligned}\tau_c &= \inf\Big\{n \geq 1 : (S_n^{(M)} - S_n^{(L)})(S_n^{(R)} - S_n^{(M)})(S_n^{(L)} - S_n^{(R)}) = 0\Big\}\\ &= \inf\Big\{n \geq 1 : (S_n^{(M)} - S_n^{(L)})(S_n^{(R)} - S_n^{(M)}) = 0\Big\}\\ &= \inf\Big\{n \geq 1 : X_n Y_n = 0\Big\}. \end{aligned} \tag{14}$$

We will use this identification to justify these results.

We will again use the martingale method. Let us define the filtration $\Big\{\mathcal{F}_n : n \geq 0\Big\}$, where

$$\mathcal{F}_n = \sigma\Big(I_k^{(L)}, I_k^{(M)}, I_k^{(R)} : k \leq n\Big) = \sigma\Big(S_k^{(L)}, S_k^{(M)}, S_k^{(R)} : k \leq n\Big)$$

is the $\sigma$-algebra generated by the increment random variables of all the random walks. Also, this is same as the $\sigma$-algebra generated by the all the random walk up to time $n$. This is the natural filtration associated with all three random walks we are studying.

**Proposition 2:** The family $\{X_nY_n + 4np(1-p) : n \geq 0\}$ is an $\mathcal{F}_n$-martingale.

**Proof:** It is easy to see that random variable $X_nY_n+4np(1-p)$ is $\mathcal{F}_n$-adapted for any $n \geq 0$. Further, for any $n \geq 0$,

$$|X_nY_n| \leq (a+2n)(b+2n)$$

Thus, we have $\mathbb{E}\Big(|X_nY_n + 4np(1-p)|\Big) < \infty$ for all $n \geq 0$.

Now, we have

$$\begin{aligned}
&X_{n+1}Y_{n+1} + 4(n+1)p(1-p)\\
&= \Big(X_n + \Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big)\Big(Y_n + \Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)\Big) + 4(n+1)p(1-p)\\
&= X_nY_n + X_n\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) + Y_n\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\\
&\quad + \Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big) + 4(n+1)p(1-p).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&(X_{n+1}Y_{n+1} + 4(n+1)p(1-p)) - (X_nY_n + 4np(1-p))\\
&= X_n\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) + Y_n\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big) + \Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) + 4p(1-p).
\end{aligned}$$

Note that $X_n$ and $Y_n$ are $\mathcal{F}_n$-measurable and the random variables $\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)$, $\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)$ are independent of $\mathcal{F}_n$ with expectation 0. Further, the random variables $I_{n+1}^{(L)}$, $I_{n+1}^{(M)}$ and $I_{n+1}^{(R)}$ are also independent of $\mathcal{F}_n$ and are independent with expectation $2p-1$ and variance $4p(1-p)$.

Now, we take conditional expectation with respect to $\mathcal{F}_n$. Observe that

- $\mathbb{E}\Big[X_n\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) \mid \mathcal{F}_n\Big] = X_n\mathbb{E}\Big[\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) \mid \mathcal{F}_n\Big] = X_n\mathbb{E}\Big[\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)\Big] = 0$ where we have used the fact that $X_n$ is $\mathcal{F}_n$-measurable and the increments random variables are independent of $\mathcal{F}_n$.
- Similarly we have $\mathbb{E}\Big[Y_n\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big) \mid \mathcal{F}_n\Big] = 0$ .
- Finally, using the fact that the increments are independent of $\mathcal{F}_n$, we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) \mid \mathcal{F}_n\Big]\\
&= \mathbb{E}\Big[\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)\Big]\\
&= \mathbb{E}\Big[\Big((I_{n+1}^{(M)} - (2p-1)) - (I_{n+1}^{(L)} - (2p-1)\Big)\Big((I_{n+1}^{(R)} - (2p-1)) - (I_{n+1}^{(M)} - (2p-1))\Big)\Big]\\
&= -\mathbb{V}\mathrm{ar}\Big(I_{n+1}^{(M)}\Big) = -4p(1-p).
\end{aligned}$$

Combing the above and the fact that $X_nY_n$ is measurable with respect to $\mathcal{F}_n$, we now have

$$\mathbb{E}\Big(X_{n+1}Y_{n+1} + 4(n+1)p(1-p) \mid \mathcal{F}_n\Big) = X_nY_n + 4np(1-p).$$

This proves the proposition. □

Next we proves another similar proposition.

**Proposition 3:** The family $\{X_nY_n\,(X_n+Y_n) : n \geq 0\}$ is an $\mathcal{F}_n$-martingale.

**Proof:** The adaptedness of $X_nY_n(X_n+Y_n)$ with respect $\mathcal{F}_n$ is again straightforward. Further, it is also obvious that $|X_nY_n(X_n+Y_n)| \leq (a+2n)(b+2n)(a+b+4n)$ and hence $\mathbb{E}\Big(|X_nY_n(X_n+Y_n)|\Big) < \infty$ for any $n \geq 0$.

As in the earlier proposition, we have

$$\begin{aligned}
&X_{n+1}Y_{n+1}\,(X_{n+1}+Y_{n+1}) \\
&= \left(X_n + \left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right)\right)\left(Y_n + \left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right)\right)\left(X_n + Y_n + \left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right)\right) \\
&= X_nY_n\,(X_n+Y_n) + X_nY_n\left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right) + X_n\,(X_n+Y_n)\left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right) \\
&\quad + Y_n\,(X_n+Y_n)\left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right) + X_n\left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right)\left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right) \\
&\quad + Y_n\left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right)\left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right) + (X_n+Y_n)\left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right)\left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right) \\
&\quad \left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right)\left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right)\left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right).
\end{aligned}$$

As in the previous proposition, we have $X_n$ and $Y_n$ are $\mathcal{F}_n$-measurable and the random variables $\left(I_{n+1}^{(R)} - I_{n+1}^{(M)}\right)$, $\left(I_{n+1}^{(M)} - I_{n+1}^{(L)}\right)$ and $\left(I_{n+1}^{(R)} - I_{n+1}^{(L)}\right)$ are independent of $\mathcal{F}_n$ with expectation 0. Thus, same arguments as above, apply to show that

- $\mathbb{E}\Big[X_nY_n\Big(I_{n+1}^{(R)} - I_{n+1}^{(L)}\Big) \mid \mathcal{F}_n\Big] = \mathbb{E}\Big[X_n(X_n+Y_n)\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) \mid \mathcal{F}_n\Big] = \Big[Y_n(X_n+Y_n)\Big(I_{n+1}^{(R)} - I_{n+1}^{(L)}\Big) \mid \mathcal{F}_n\Big] = 0.$

- Same arguments as above, yield

$$\begin{aligned}
&(a)\quad \mathbb{E}\Big[X_n\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(L)}\Big) \mid \mathcal{F}_n\Big] = X_n\mathbb{V}\mathrm{ar}\Big(I_{n+1}^{(R)}\Big) = 4p(1-p)X_n \\
&(b)\quad \mathbb{E}\Big[Y_n\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(L)}\Big) \mid \mathcal{F}_n\Big] = Y_n\mathbb{V}\mathrm{ar}\Big(I_{n+1}^{(L)}\Big) = 4p(1-p)Y_n \\
&(c)\quad \mathbb{E}\Big[\Big(X_n+Y_n\Big)\Big(I_{n+1}^{(M)} - I_{n+1}^{(L)}\Big)\Big(I_{n+1}^{(R)} - I_{n+1}^{(M)}\Big) \mid \mathcal{F}_n\Big] = -(X_n+Y_n)\mathbb{V}\mathrm{ar}\Big(I_{n+1}^{(M)}\Big) \\
&\qquad\qquad = -4p(1-p)(X_n+Y_n).
\end{aligned}$$

- We also have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(I^{(M)}_{n+1} - I^{(L)}_{n+1}\Big)\Big(I^{(R)}_{n+1} - I^{(M)}_{n+1}\Big)\Big(I^{(R)}_{n+1} - I^{(L)}_{n+1}\Big) \mid \mathcal{F}_n\Big] \\
&= \mathbb{E}\Big[\Big(I^{(M)}_{n+1} - I^{(L)}_{n+1}\Big)\Big(I^{(R)}_{n+1} - I^{(M)}_{n+1}\Big)\Big(I^{(R)}_{n+1} - I^{(L)}_{n+1}\Big)\Big] \\
&= \mathbb{E}\Big[\Big((I^{(M)}_{n+1} - (2p-1)) - (I^{(L)}_{n+1} - (2p-1))\Big)\Big((I^{(R)}_{n+1} - (2p-1)) - (I^{(M)}_{n+1} - (2p-1))\Big) \\
&\qquad \times \Big((I^{(R)}_{n+1} - (2p-1)) - (I^{(L)}_{n+1} - (2p-1))\Big)\Big] = 0
\end{aligned}$$

by independence of the random variables and the fact that they have expectation 0.

Combining the above and the fact that $X_nY_n(X_n + Y_n)$ is $\mathcal{F}_n$-measurable, we have

$$\mathbb{E}\Big(X_{n+1}Y_{n+1}\Big(X_{n+1} + Y_{n+1}\Big) \mid \mathcal{F}_n\Big) = X_nY_n\Big(X_n + Y_n\Big).$$

This completes the proof. □

Now, we are in a position to state and prove our main result.

**Theorem 2:** We have

$$\mathbb{E}\Big(\tau_c\Big) = ab\Big(4p(1-p)\Big)^{-1}. \tag{15}$$

**Proof:** We observe that, from equation (13), that

$$\{\tau_c = n\} = \Big\{X_0Y_0 > 0, X_1Y_1 > 0, \ldots, X_{n-1}Y_{n-1} > 0, X_nY_n = 0\Big\}.$$

Clearly $\{\tau_c = n\} \in \mathcal{F}_n$, which implies that $\tau_c$ is also stopping time relative to $\{\mathcal{F}_n\}$.

By using Proposition 2, we get that, $\Big\{X_{n\wedge\tau_c}Y_{n\wedge\tau_c} + 4p(1-p)\,(n \wedge \tau_c) : n \geq 0\Big\}$ is a martingale and hence for any $n \geq 1$,

$$\begin{aligned}
&\mathbb{E}\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c} + 4p(1-p)\,(n \wedge \tau_c)\Big) \\
&= \mathbb{E}\Big(X_{0\wedge\tau_c}Y_{0\wedge\tau_c} + 4p(1-p)\,(0 \wedge \tau_c)\Big) \\
&= \mathbb{E}\Big(X_0Y_0\Big) = ab
\end{aligned} \tag{16}$$

since $\tau_c \geq 0$.

Now, we will take limit in equation (16) as $n \to \infty$. Since $\tau_c < \infty$ almost surely, $n \wedge \tau_c \uparrow \tau_c$ as $n \to \infty$. By MCT, we obtain

$$\mathbb{E}\Big(n \wedge \tau_c\Big) \to \mathbb{E}\Big(\tau_c\Big)$$

as $n \to \infty$.

To complete the proof we show that $\mathbb{E}\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big) \to 0$ as $n \to \infty$. Since $\tau_c < \infty$ almost surely, we have that

$$X_{n\wedge\tau_c}Y_{n\wedge\tau_c} \to X_{\tau_c}Y_{\tau_c} = 0 \tag{17}$$

as $n \to \infty$.

In order to show that the expected value also converges to 0, we will use Theorem 26.13 of Billingsley (1986). For this we require to show that the sequence of random variable $\{X_{n\wedge\tau_c}Y_{n\wedge\tau_c} : n \geq 0\}$ is an uniformly integrable family. A sufficient condition for a family of random variables to be uniformly integrable (see Billingsley (1986)) is given by

$$\sup_{n\geq 0} \mathbb{E}\Big[\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big)^{1+\epsilon}\Big] < \infty$$

for some $\epsilon > 0$.

By using Proposition 3, we get that $\Big\{X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\left(X_{n\wedge\tau_c} + Y_{n\wedge\tau_c}\right) : n \geq 0\Big\}$ is also a martingale. Hence, for any $n \geq 1$,

$$\begin{aligned}
&\mathbb{E}\Big[X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big(X_{n\wedge\tau_c} + Y_{n\wedge\tau_c}\Big)\Big]\\
&= \mathbb{E}\Big[X_{0\wedge\tau_c}Y_{0\wedge\tau_c}\Big(X_{0\wedge\tau_c} + Y_{0\wedge\tau_c}\Big)\Big]\\
&= \mathbb{E}\Big[X_0Y_0\Big(X_0 + Y_0\Big)\Big]\\
&= ab(a+b).
\end{aligned}$$

For non-negative $u, v \geq 0$, using AM-GM inequality, we have $(uv)^{3/2} \leq \frac{1}{2}uv(u+v)$. Since $X_{n\wedge\tau_c}$ and $Y_{n\wedge r_c}$ are both non negative, we have, for any $n \geq 0$

$$\mathbb{E}\Big[\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big)^{3/2}\Big] \leq \frac{1}{2}\mathbb{E}\Big[X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big(X_{n\wedge\tau_c} + Y_{n\wedge\tau_c}\Big)\Big] = \frac{1}{2}ab(a+b).$$

Therefore,

$$\sup_{n\geq 0} \mathbb{E}\Big[\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big)^{1+1/2}\Big] \leq \frac{1}{2}ab(a+b) < \infty.$$

Hence, we conclude that $\{X_{n\wedge\tau_c}Y_{n\wedge\tau_c} : n \geq 0\}$ is an uniformly integrable family. Therefore, we have

$$\mathbb{E}\Big(X_{n\wedge\tau_c}Y_{n\wedge\tau_c}\Big) \to 0 \text{ as } n \to \infty.$$

This completes the proof of the Theorem. □

**Acknowledgements**

Most of the work was completed when Arighna Dey was pursuing the M. Stat. degree from the Indian Statistical Institute and he wishes to the thank ISI for this opportunity. Kumarjit Saha wishes to thank the Ashoka University for the support. Anish Sarkar wishes to the Indian Statistical Institute for support and also the organizers of the conference for giving the an opportunity to present the work.

## References

Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons.

Coupier, D., Saha, K., Sarkar, A., and Tran, V. C. (2020). Collision times of random walks and applications to the Brownian web. *Genealogies of interacting particle systems, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, World Scientific ISBN - 978-981-120-608-5.*

Spitzer, F. (1964). *Principles of Random Walk*. Van Nostrand.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

## ANNEXURE

### R code for simulation

```
Increment<-function (uniform, p)
{
    # uniform := uniform variable
    # p := probability of increment of +1,

    # output := the increment with probability distribution

    if (uniform <= p)
    {
        return (1)
    }
    return (-1)

}

FindCollision<-function(
    startright,
    startmid,
    startleft,
    p,
    cutoff)
{
    # startright := starting position of right random walk
    # startmid := starting position of mid random walk
    # startleft := starting position of left random walk
    # cutoff := the max length of random walk to be considered

    # output := First Collision time of 3 random walks

    # initialize starting positions
    rightpos = startright
```

```
    midpos = startmid
    leftpos = startleft

    # set time for collision to cutoff + 1
    time = cutoff+1

    # run the loop until cutoff time
    for (i in 1:cutoff)
    {

        # Get three uniforms
        uniforms = runif(3)

        # update the random walk positions
        rightpos = rightpos + Increment(uniforms[1], p)
        midpos = midpos + Increment(uniforms[2], p)
        leftpos = leftpos + Increment(uniforms[3], p)

        # Check for collision
        if ( (rightpos - midpos)*(midpos-leftpos) == 0 )
        {
            # Collision has happened
            # set time to this collision time
            time = i

            # stop the simulation
            break
        }
    }

    # return the time
    return (time)
}

RW<-function (
    startright,
    startmid,
    startleft,
    p,
    cutoff,
    num)
{
    # output := mean of First Collision times of num repeatation
    W = rep (0, num)

    # run loop for repeatations of times
    for (i in 1:num)
```

```
    {
        W[i] = FindCollision(
            startright,
            startmid,
            startleft,
            p,
            cutoff)
    }
    ava = c(mean(W))
    return (ava)
}
```