Random walks

March 19, 2022

Contents

Random walk

Definition

Let $\{Y_k : k \geqslant 1\}$ be a sequence of independent random variables such that $\mathbb{P}(Y_i = 1) = p$ and $\mathbb{P}(Y_i = -1) = q$, $i \geqslant 1$. Let

$$X_n = \begin{cases} 0, & n = 0 \\ Y_1 + Y_2 + \cdots + Y_n, & n \geqslant 1 \end{cases}$$

Then, $\{X_n : n \geqslant 1\}$ is a discrete time stochastic process called a random walk (with parameter p).

- The state space is $\mathbb{Z}$.
- If $p = q = 1/2$ we say that the random walk is symmetric.

Random walk

A realization of this process can be visualized as the movement of a particle that walks randomly in one dimension, starting from the origin.

If at time n the particle is at coordinate k, then at time $n+1$ it will be

- *either at coordinate $k+1$ with probability p,*

$$\mathbb{P}(X_{n+1} = k+1 \,|\, X_n = k) = p$$

- *or at coordinate $k-1$ with probability q,*

$$\mathbb{P}(X_{n+1} = k-1 \,|\, X_n = k) = q$$

Random walk

Another interpretation is as a game between two players A and B that repeteadly toss a coin showing heads with probability p.

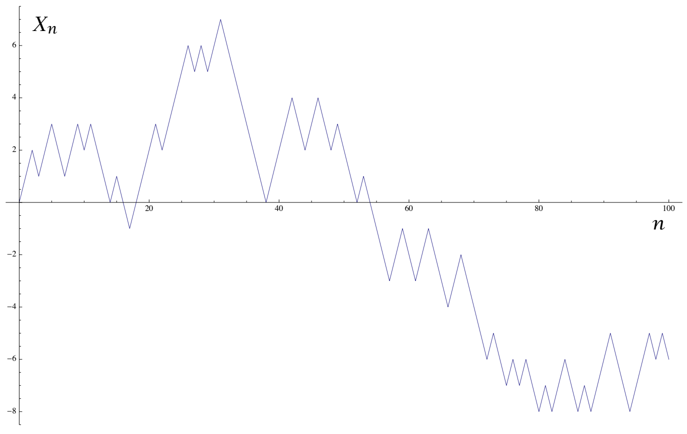
Let H_n and T_n be the number of heads and tails in the first n tosses. Thus,

$$X_n = H_n - T_n, \quad n \geqslant 1$$

If at time n:

- $X_n > 0$, *we say that player A is in the lead;*
- $X_n < 0$, *we say that player B is in the lead;*
- $X_n = 0$, *then the game is in a draw.*

Random walk



Mean and variance

Notice that $\{X_n = k\}$ if and only if in the first n tosses the number r of heads and the number $s = n - r$ of tails satisfy:

$$r - s = 2r - n = k$$

Hence,

$$r = \frac{n+k}{2}, \quad r = 0, 1, \dots, n$$

At time n the particle is at coordinate k with probability

$$\mathbb{P}(X_n = k) = \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}},$$

where $k = -n, -n+2, \dots, n-2, n$.

Mean and variance

We have that

$$\mathbb{E}(X_n) = \mathbb{E}(H_n) - \mathbb{E}(T_n) = n(p-q)$$

$$\text{Var}(X_n) = \text{Var}(2H_n - n) = 4npq$$

Moreover, the Central Limit Theorem implies

$$\frac{X_n - n(p-q)}{\sqrt{4npq}} \xrightarrow{d} \text{N}(0,1)$$

- If the walk is symmetric, then $\mathbb{E}(X_n) = 0$, $\text{Var}(X_n) = n$ and

$$\frac{X_n}{\sqrt{n}} \xrightarrow{d} \text{N}(0,1)$$

First visit to a negative coordinate

Let T be the random time at which the particle visits a negative coordinate for the first time,

$$T = \min\{n \geqslant 1 : \; X_n < 0\}$$

Alternatively, T is the first time at which player B is in the lead.

Let us calculate $\mathbb{P}(T = n)$.

First visit to a negative coordinate

For instance,

$$\mathbb{P}(T = 1) = q$$
$$\mathbb{P}(T = 2) = 0$$
$$\mathbb{P}(T = 3) = pq^2$$
$$\mathbb{P}(T = 4) = 0$$
$$\mathbb{P}(T = 5) = 2p^2q^3$$
$$\dots\dots\dots\dots\dots$$

Notice that $\mathbb{P}(T = 2k) = 0$ for all $k \geqslant 1$.

- It could be possible that the particle will never visit a negative coordinate, that is to say, $\mathbb{P}(T = \infty) > 0$. In such a case, we say that the random variable T is defective.

First visit to a negative coordinate

For odd $n \geqslant 3$ we have that

$$P(T = n) = \mathbb{P}(T = n, X_1 = 1) = p\,\mathbb{P}(T = n \,|\, X_1 = 1)$$

Let Z be the time of the first return to the origin,

$$Z = \min\{m \geqslant 2 : \; X_m = 0\}$$

Then

$$\mathbb{P}(T = n \,|\, X_1 = 1) = \sum_{m=2}^{n-1} \mathbb{P}(T = n \,|\, Z = m, X_1 = 1)\,\mathbb{P}(Z = m \,|\, X_1 = 1)$$

First visit to a negative coordinate

Notice that

- $\mathbb{P}(Z = m \,|\, X_1 = 1) = \mathbb{P}(T = m - 1)$
- $\mathbb{P}(T = n \,|\, Z = m, X_1 = 1) = \mathbb{P}(T = n - m)$

Therefore, if we denote $\mathbb{P}(T = n)$ as r_n we have the recurrent equation

$$r_n = p \sum_{m=2}^{n-1} r_{m-1} r_{n-m}, \quad n \geqslant 3, \text{ odd}$$

First visit to a negative coordinate

For instance,

$$P(T=1) = r_1 = q$$
$$\mathbb{P}(T=3) = r_3 = p\, r_1^2 = pq^2$$
$$\mathbb{P}(T=5) = r_5 = p\,(r_1 r_3 + r_3 r_1) = 2p^2q^3$$
$$\mathbb{P}(T=7) = r_7 = p\,(r_1 r_5 + r_3^2 + r_5 r_1) = 5p^3q^4$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$\mathbb{P}(T=n) = r_n = p\,(r_1 r_{n-2} + r_3 r_{n-4} + \cdots + r_{n-2} r_1)$$
$$\cdots\cdots\cdots\cdots\cdots$$

First visit to a negative coordinate

Let

$$G_T(s) = \sum_{n \geqslant 0} r_n\, s^n$$

be the generating function of the sequence $\{r_n\}$. (We take $r_0 = 0$.)

Observe that

$$G_T(s) = r_1 s + r_3 s^3 + r_5 s^5 + \cdots$$

First visit to a negative coordinate

Therefore

$$\begin{aligned} p\,s\,(G_T(s))^2 &= ps\,(r_1 s + r_3 s^3 + r_5 s^5 + \cdots)(r_1 s + r_3 s^3 + r_5 s^5 + \cdots) \\ &= p\, r_1^2 s^3 + p(r_1 r_3 + r_3 r_1) s^5 + p\,(r_1 r_5 + r_3^2 + r_5 r_1) s^7 + \cdots \\ &\cdots + p\,\underbrace{(r_1 r_{n-2} + r_3 r_{n-4} + \cdots + r_{n-2} r_1)}_{r_n} s^n + \cdots \\ &= r_3 s^3 + r_5 s^5 + r_7 s^7 + \cdots + r_n s^n + \cdots = G_T(s) - q\,s \end{aligned}$$

First visit to a negative coordinate

Then

$$p\,s\,(G_T(s))^2 - G_T(s) + qs = 0$$

Solving for $G_T(s)$ we obtain

$$G_T(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps} = \frac{2qs}{1 + \sqrt{1 - 4pqs^2}}$$

- We have to choose the root of $\sqrt{}$ which makes $G_T(s)$ finite at $s = 0$.
- The radius of convergence is $R = 1/(2\sqrt{pq})$. Hence, if $p = q = 1/2$, then $R = 1$. Otherwise, $R > 1$.

First visit to a negative coordinate

For a symmetric random walk ($p = q = 1/2$):

$$G_T(s) = \frac{s}{1+\sqrt{1-s^2}}$$
$$= \frac{1}{2}s + \frac{1}{8}s^3 + \frac{1}{16}s^5 + \frac{5}{128}s^7 + \frac{7}{256}s^9 + \frac{21}{1024}s^{11} + \cdots$$

In general,

$$G_T(s) = \frac{2qs}{1+\sqrt{1-4pqs^2}}$$
$$= qs + pq^2s^3 + 2p^2q^3s^5 + 5p^3q^4s^7 + 14p^4q^5s^9 + 42p^5q^6s^{11} + \cdots$$

Eventual visit to a negative coordinate

The probability of an eventual visit to a negative coordinate is

$$\sum_{n\geqslant 0} \mathbb{P}(T = n) = \sum_{n\geqslant 0} r_n$$
$$= G_T(1) = \frac{1-\sqrt{1-4pq}}{2p} = \frac{1-|q-p|}{2p}$$
$$= \begin{cases} 1, & p \leqslant q \\ \frac{q}{p}, & p > q \end{cases}$$

Eventual visit to a negative coordinate

Therefore

- If $p \leqslant q$, the particle will visit a negative coordinate with probability 1. In this case, T is a proper random variable, and

  $$r_n = \mathbb{P}(T = n), \quad n \geqslant 0$$

  is its probability function.
- If $p > q$, the variable T is defective. There is a nonzero probability that the particle never visits a negative coordinate:

  $$\mathbb{P}(T = \infty) = 1 - \sum_{n\geqslant 0} \mathbb{P}(T = n) = 1 - \frac{q}{p}$$

Expected time to a negative coordinate

If T is not defective ($p \leqslant q$), then

$$\mathbb{E}(T) = G_T'(1) = \begin{cases} \frac{1}{q-p}, & p < q \\ \infty, & p = q \end{cases}$$

For instance,

- for $p = 0.4$ one has $\mathbb{E}(T) = 5$,
- for $p = 0.49$ one has $\mathbb{E}(T) = 50$,
- for $p = 0.499$ one has $\mathbb{E}(T) = 500$ ("long excursions").

Explicit value of r_n

The probability $r_n = \mathbb{P}(T = n)$ is the coefficient of s^n in the series expansion of

$$G_T(s) = \frac{1 - \sqrt{1 - 4pqs^2}}{2ps}$$

Therefore

$$r_{2k-1} = -\frac{1}{2p} c_{2k},$$

where c_{2k} is the coefficient of s^{2k} in the expansion of the function

$$(1 - 4pqs^2)^{1/2}$$

Explicit value of r_n

To find c_{2k} we can use the binomial series

$$(1 + x)^\alpha = \sum_{m=0}^{\infty} \binom{\alpha}{m} x^m,$$

where

$$\binom{\alpha}{m} = \frac{\alpha(\alpha - 1)\cdots(\alpha - m + 1)}{m!}, \quad \alpha \in \mathbb{R}$$

Thus

$$(1 - 4pqs^2)^{1/2} = \sum_{k=0}^{\infty} \binom{1/2}{k} (-4pqs^2)^k$$

Explicit value of r_n

Therefore

$$\begin{aligned} r_{2k-1} &= -\frac{1}{2p}\binom{1/2}{k}(-4pq)^k = -\frac{1}{2}\binom{1/2}{k}(-4)^k p^{k-1} q^k \\ &= -\frac{1}{2}\,\frac{(1/2)(-1/2)(-3/2)\cdots(1/2 - k + 1)}{k!}(-4)^k p^{k-1} q^k \\ &= \frac{1}{2} 2^k \frac{1 \cdot 3 \cdots (2k-3)}{k!} p^{k-1} q^k \\ &= \frac{1 \cdot 2 \cdot 3 \cdots (2k-3)(2k-2)}{(k-1)!\,k!} p^{k-1} q^k \\ &= \frac{1}{2k-1}\binom{2k-1}{k} p^{k-1} q^k, \quad k \geqslant 1 \end{aligned}$$

Returns to the origin

The probability that the particle is visiting the origin at time n is

$$\mathbb{P}(X_n = 0) = \begin{cases} 0, & n \text{ odd} \\ \binom{2k}{k} p^k q^k, & n = 2k \end{cases}$$

Returns to the origin

By Stirling's approximation, $m! \sim \sqrt{2\pi m}\,(m/e)^m$, we have that

$$\mathbb{P}(X_{2k}=0) \sim \frac{\sqrt{2\pi 2k}\,\left(\frac{2k}{e}\right)^{2k}}{2\pi k\,\left(\frac{k}{e}\right)^{2k}} p^k\, q^k = \frac{(4pq)^k}{\sqrt{\pi k}}$$

- In particular, for a symmetric random walk:

$$\mathbb{P}(X_{2k}=0) \sim \frac{1}{\sqrt{\pi k}}$$

Returns to the origin

Let $A^\star$ be the limit superior of the sequence of events

$$A_k = \{X_{2k}=0\}\,, \quad k \geqslant 1$$

That is to say, $A^\star$ is the event "the return to the origin happens infinitely often"

Remember that

$$\mathbb{P}(A_k) \sim \frac{(4pq)^k}{\sqrt{\pi k}}$$

Returns to the origin

If the random walk is not symmetric ($p \neq q$), then $4pq < 1$.

In this case, the series $\sum_k ((4pq)^k)/\sqrt{\pi k}$ converges. Therefore, $\sum_k \mathbb{P}(A_k) < \infty$, and the first Borel-Cantelli lemma implies $\mathbb{P}(A^\star)=0$:

$$\mathbb{P}\left(\{X_{2k}=0\} \text{ happens infinitely often}\right) = 0$$

(Remark)

If the random walk is not symmetric, then (with probability one) $\{X_{2k}=0\}$ *occurs only a finite number of times.*

First return to the origin

Let $f_n = \mathbb{P}(Z=n)$, where $Z = \min\{n \geqslant 1 : \ X_n = 0\}$.

That is,

$$f_0 = f_1 = 0$$

$$f_n = \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{n-1} \neq 0, X_n = 0), \quad n \geqslant 2$$

For instance,

$$f_2 = \mathbb{P}(Z=2) = 2pq$$

$$f_3 = \mathbb{P}(Z=3) = 0$$

$$f_4 = \mathbb{P}(Z=4) = 2p^2q^2$$

....

First return to the origin

For $n \geqslant 2$ we have that

$$\begin{aligned} f_n = P(Z = n) & \\ &= p\,\mathbb{P}(Z = n \,|\, X_1 = 1) + q\,\mathbb{P}(Z = n \,|\, X_1 = -1) \\ &= p\, r_{n-1} + q\, r'_{n-1} \end{aligned}$$

where r'_{n-1} is as r_{n-1} exchanging p for q.

- Notice that r'_n is the probability that the first visit to a positive coordinate happens at time n.

First return to the origin

Then

$$\begin{aligned} G_Z(s) &= p\,s \sum_{n\geqslant 2} r_{n-1} s^{n-1} + q\,s \sum_{n\geqslant 2} r'_{n-1} s^{n-1} \\ &= p\,s \sum_{n\geqslant 0} r_n s^n + q\,s \sum_{n\geqslant 0} r'_n s^n \\ &= p\,s\,G_T(s) + q\,s\,G_{T'}(s) \\ &= p\,s\,\frac{1-\sqrt{1-4pqs^2}}{2ps} + q\,s\,\frac{1-\sqrt{1-4pqs^2}}{2qs} \\ &= 1-\sqrt{1-4pqs^2} \end{aligned}$$

First return to the origin

The series expansion of $G_Z(s)$ is

$$G_Z(s) = 2pq\,s^2 + 2p^2q^2\,s^4 + 4p^3q^3\,s^6 + 10p^4q^4\,s^8 + \cdots$$

Thus,

$$\begin{aligned} \mathbb{P}(Z=2) &= f_2 = 2pq \\ \mathbb{P}(Z=4) &= f_4 = 2p^2q^2 \\ \mathbb{P}(Z=6) &= f_6 = 4p^3q^3 \\ \mathbb{P}(Z=8) &= f_8 = 10p^4q^4 \\ \cdots & \quad \cdots \end{aligned}$$

Eventual return to the origin

The probability of an eventual return to the origin is:

$$\begin{aligned} \sum_{n\geqslant 0} \mathbb{P}(Z=n) = \sum_{n\geqslant 0} f_n & \\ &= G_Z(1) = 1-\sqrt{1-4pq} = 1-|q-p| \end{aligned}$$

- If $p \neq q$, there is a nonzero probability $|q-p|$ that the particle never returns to the origin.
- If $p = q$ (symmetric walk), then Z is not defective and the return to the origin happens with probability one. However,

$$\begin{aligned} \mathbb{E}(Z) = G'_Z(1) & \\ &= \frac{d}{ds}\left(1-\sqrt{1-s^2}\right)\Big|_{s=1} = \frac{s}{\sqrt{1-s^2}}\Big|_{s=1} = \infty \end{aligned}$$

Other properties

The following property holds:

$$\begin{aligned}\mathbb{P}(Z=2k) &= f_{2k} = p\, r_{2k-1} + q\, r'_{2k-1}\\ &= p\,\frac{1}{2k-1}\binom{2k-1}{k} p^{k-1}q^{k} + q\,\frac{1}{2k-1}\binom{2k-1}{k} p^{k}q^{k-1}\\ &= \frac{2}{2k-1}\binom{2k-1}{k} p^{k}q^{k} = \frac{1}{2k-1}\binom{2k}{k} p^{k}q^{k}\\ &= \frac{1}{2k-1}\,\mathbb{P}(X_{2k}=0)\end{aligned}$$

Other properties

Moreover, for a symmetric random walk it can be proved that

$$\mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m} \neq 0) = \mathbb{P}(X_{2m} = 0)$$

Proof:

Let $f_{2k} = \mathbb{P}(Z = 2k)$ and $u_{2k} = \mathbb{P}(X_{2k} = 0)$. If the random walk is symmetric, then

$$f_{2k} = \frac{1}{2k-1}\binom{2k}{k}\frac{1}{2^{2k}}, \quad u_{2k} = \binom{2k}{k}\frac{1}{2^{2k}}$$

It can be checked that

$$u_{2k-2} - u_{2k} = f_{2k}, \quad k \geqslant 1$$

Other properties

Hence

$$\begin{aligned}&\mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m} \neq 0)\\ &\qquad = 1 - \mathbb{P}(X_{2k} = 0 \text{ for some } 1 \leqslant k \leqslant m)\\ &\qquad = 1 - \sum_{k=1}^{m} f_{2k} = 1 - \sum_{k=1}^{m}(u_{2k-2} - u_{2k}) = u_{2m}\\ &\qquad = \mathbb{P}(X_{2m} = 0)\end{aligned}$$

Last return to the origin

Theorem

The probability $\alpha_{2k,2m}$ that a symmetric random walk of length $2m$ has a last return to the origin at time $2k$ equals

$$\begin{aligned}\alpha_{2k,2m} &= \binom{2k}{k}\binom{2m-2k}{m-k}\frac{1}{2^{2m}}\\ &= \mathbb{P}(X_{2k}=0)\,\mathbb{P}(X_{2m-2k}=0), \quad 0 \leqslant k \leqslant m\end{aligned}$$

(The case $k = 0$ corresponds to $\{X_i \neq 0 :\ 1 \leqslant i \leqslant 2m\}$.)

Last return to the origin

Proof.
Let $A_{2k,2m}$ denote the event "the symmetric random walk of length $2m$ has a last return to the origin at time $2k$", $0 \leqslant k \leqslant m$.

Then,

$$\alpha_{2k,2m} = \mathbb{P}(A_{2k})$$
$$= \mathbb{P}\left(\{X_{2k} = 0\} \cap \{X_{2k+1} \neq 0, X_{2k+2} \neq 0, \ldots, X_{2m} \neq 0\}\right)$$
$$= \mathbb{P}(X_{2k} = 0)\mathbb{P}(X_{2k+1} \neq 0, X_{2k+2} \neq 0, \ldots, X_{2m} \neq 0 \mid X_{2k} = 0)$$

Last return to the origin

A random walk is temporally homogeneous. Therefore,

$$\mathbb{P}(X_{2k+1} \neq 0, X_{2k+2} \neq 0, \ldots, X_{2m} \neq 0 \mid X_{2k} = 0)$$
$$= \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m-2k} \neq 0 \mid X_0 = 0)$$
$$= \mathbb{P}(X_0 = 0, X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m-2k} \neq 0)$$
$$= \mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m-2k} \neq 0)$$

Moreover, we know that

$$\mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2r} \neq 0) = \mathbb{P}(X_{2r} = 0)$$

Last return to the origin

Hence,

$$\alpha_{2k,2m} = \mathbb{P}(A_{2k})$$
$$= \mathbb{P}(X_{2k} = 0)\mathbb{P}(X_1 \neq 0, X_2 \neq 0, \ldots, X_{2m-2k} \neq 0)$$
$$= \mathbb{P}(X_{2k} = 0)\mathbb{P}(X_{2m-2k} = 0)$$

For large k, the asymptotic behaviour of $\mathbb{P}(X_{2k} = 0)$ is

$$\mathbb{P}(X_{2k} = 0) \sim \frac{1}{\sqrt{\pi k}}$$

Hence, if k and m go to infinity, then

$$\alpha_{2k,2m} \sim \frac{1}{\pi\sqrt{k(m-k)}}$$

Last return to the origin

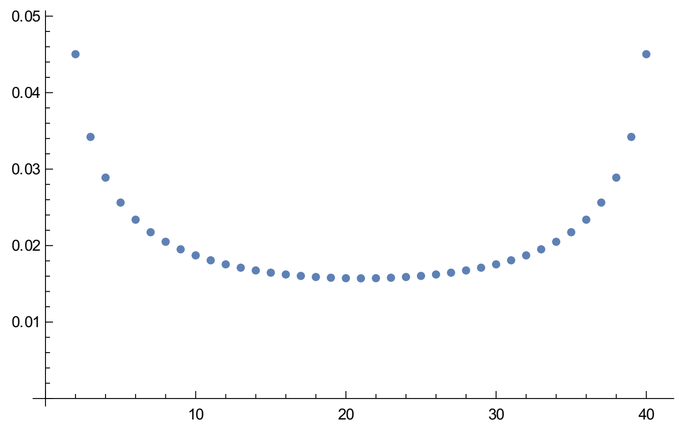


Figure: A plot of the probability $\alpha_{2k,80}$ for $k = 1, 2, \ldots, 40$

Last return to the origin

Observe that

$$\alpha_{2k,2m} = \alpha_{2m-2k,2m}$$

(Remark)

Since each possible trajectory of length $2m$ occurs with probability $1/2^{2m}$, this symmetry means that the number of paths that have a last return to 0 at time $2k$ equals the number of paths that have a last return to 0 at time $2m - 2k$.

Hence, the probability that a symmetric random walk of length $2m$ has no return to the origin during the last m steps is $1/2$.

Last return to the origin

Another curious result is the following one.

Theorem

If players A and B play a game of Heads and Tails of length $2m$, the probability that A will be in the lead exactly $2k$ times is equal to $\alpha_{2k,2m}$.

Last return to the origin

If we ask what is the most probable number of times that player A is in the lead, many people will say that the answer is m. But this is not at all the case.

In the figure, the values of $\alpha_{2m,2k}$ are plotted for $2m = 80$ and $k = 1, 2, \ldots, 40$.

(Remark)

The less probable value of the number of times that A is in the lead occurs for $2k = m$ ($m = 40$ in the figure); whereas (unexpectedly) the most probable one occurs for $2k = 0$ and $2k = 2m$