Generating Function Approach for Simple Random Walk

Fangda Xu, Xin Zhao

Notes: All contents below are coded by ourselves.

Background: Definition and Property of the Generating Function

Definition of the generating function:

\[ G(s) = \sum_{n=0}^{\infty} s^n a_n \]

If the sequence \( \{a_n\} \) is the probability mass function of a random variable \( X \) on the nonnegative integers (i.e. \( P(x = n) = a_n \)), then we call the generating function the probability generating function of \( X \), and we can write it as:

\[ G(s) = E[s^X] \]

It is obvious that if \( G_x(s) \) is the generating function of a random variable \( X \), then,

\[ G'_x(1) = E[x] \]

since \( G'(s) = \frac{d}{ds} (a_0 + a_1 s + a_2 s^2 + \cdots) \)

\[ = a_1 + 2a_2 s + 3a_3 s^2 + \cdots \]
then \[ G'(1) = a_1 + 2a_2 + 3a_3 + \cdots = E[x] \]

based on the assumption that \( P(x = n) = a_n \)

Application of Generating Function on Simple Random Walk

The following content will be confined to the field of Simple Random Walk.

Random Walk: A random walk is a stochastic sequence \( \{S_n\} \) with \( S_0=0 \), defined by

\[
S_n = \sum_{k=1}^{n} X_k
\]

where \( \{X_k\} \) are independent and identically distributed random variables.

Simple Random Walk: The random walk is simple if \( X_k = \pm 1 \) with \( P(X_k=1)=p \) and \( P(X_k=-1)=q=1-p \).

![Figure 1: Simple random walk](image)

Let us suppose that the random walk starts in state 0 at time 0:
\(T_r\) = time(steps) that the walk first reaches state \(r\), for \(r \geq 1\)

\(T_0\) = time(steps) that the walk first returns to state 0

\(f_r(n) = P(T_r = n|X_0 = 0)\) for \(r \geq 0\) and \(n \geq 0\) and we let

\[
G_r(s) = \sum_{n=0}^{\infty} s^n f_r(n)
\]

Then the key point is how we can obtain the \(G_r(s)\) for \(r > 1\). We will start from \(G_1(s)\) based on the Markov property below.

\[
f_r(n) = \sum_{k=0}^{\infty} P(T_r = n|T_1 = k)f_1(k)
\]

\[
= \sum_{k=0}^{n} P(T_r = n|T_1 = k)f_1(k)
\]

Now the focus turns to the \(P(T_r = n|T_1 = k)\) (we truncate the sum at \(n\) since \(P(T_r = n|T_1 = k) = 0\) for \(k > n\)). By applying temporal and spatial homogeneity (this is the same as the probability that the first time we reach state \(r-1\) is at time \(n-k\) given that we start in state 0 at time 0), that is

\[
P(T_r = n|T_1 = k) = f_{r-1}(n-k)
\]
and so

\[ f_r(n) = \sum_{k=0}^{n} f_{r-1}(n-k) f_1(k) \]

Right now \( \{f_r(n)\} \) depends on two other sequences: \( \{f_{r-1}(n)\} \) and \( \{f_1(n)\} \). By applying the convolution property of generating function (which states that \( G_c(s) = G_a(s)G_b(s) \), \( G_a(s) \) is the generating function of \( \{a_n\} \) and \( G_b(s) \) is the generating function of \( \{b_n\} \), where \( c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = \sum_{i=0}^{n} a_i b_{n-i} \), we have

\[ G_r(s) = G_{r-1}(s)G_1(s) \]

Keep doing the decomposition on \( G_{r-1}(s) \), we will finally reach:

\[ G_r(s) = G_1(s)^r \]

Now the problem reduces to how to decompose the \( G_1(s) \).

\[ f_1(n) = P(T_1 = n|X_1 = 1)p + P(T_1 = n|X_1 = -1)q \]

where \( P(T_1 = n|X_1 = -1) = f_2(n - 1) \) (applying temporal and spatial homogeneity again), therefore \( f_1(n) \) can be written as

\[ f_1(n) = q f_2(n - 1) \]

Keep in mind that \( f_1(1) = p \) and \( f_1(0) = 0 \), then we can write
$G_1(s)$ in the form of:

$$G_1(s) = \sum_{n=0}^{\infty} s^n f_1(n) = s f_1(1) + \sum_{n=2}^{\infty} s^n f_1(n)$$

$$= ps + \sum_{n=2}^{\infty} s^n f_2(n-1) = ps + qs \sum_{n=2}^{\infty} s^{n-1} f_2(n-1)$$

$$= ps + qs \sum_{n=1}^{\infty} s^n f_2(n) = ps + q s G_2(s)$$

But based on the previous result that $G_r(s) = G_1(s)^r$,

$$G_2(s) = G_1(s)^2$$

thus,

$$G_1(s) = ps + q s G_1(s)^2$$

then we can decompose the $G_1(s)$ in the form

$$G_1(s) = \frac{1 \pm \sqrt{1 - 4pq s^2}}{2qs}$$

with the boundary condition that $G_1(0) = f_1(0) = 0$, the correct form of $G_1(s)$ should be

$$G_1(s) = \frac{1 - \sqrt{1 - 4pq s^2}}{2qs}.$$ 

If we set $s=1$, we have

$$G_1(1) = \frac{1 - \sqrt{1 - 4pq}}{2q}.$$
However, after so many steps, we should always keep in mind that after so many steps, we haven’t solved the $G_0(s)$! The leftover part will focus on how to decompose the $G_0(s)$.

Still, we start from the $f_0(n)$ to generate the $G_0(s)$.

$$f_0(n) = P(T_0 = n|X_0 = 0)$$

$$= P(\text{walk first returns to 0 at time } n|X_0 = 0)$$

Based on the Markov property, we have that

$$f_0(n) = P(T_0 = n|X_1 = 1)p + P(T_0 = n|X_0 = -1)q$$

Still, applying the temporal and spatial homogeneity again, we can have

$$P(T_0 = n|X_0 = -1) = P(T_1 = n - 1|X_0 = 0) = f_1(n - 1)$$

Similarly we can have

$$P(T_0 = n|X_0 = 1) = P(T_{-1} = n - 1|X_0 = 0) = f_{-1}(n - 1)$$

Combine the above formulas, we obtain

$$f_0(n) = f_{-1}(n - 1)p + f_1(n - 1)q$$

if we denote $f_1^*(n) = f_{-1}(n)$, we will have
\[ f_0(n) = f_1^*(n - 1)p + f_1(n - 1)q \]

Therefore, we will have

\[
G_0(s) = \sum_{n=1}^{\infty} s^n f_1^*(n - 1)p + \sum_{n=1}^{\infty} s^n f_1(n - 1)q
\]

\[
= ps \sum_{n=1}^{\infty} s^{n-1} f_1^*(n - 1) + \sum_{n=1}^{\infty} s^{n-1} f_1(n - 1)
\]

\[
= psG_1^*(s) + qsG_1(s)
\]

where \( G_1^*(s) \) is the same function as \( G_1(s) \) except with \( p \) and \( q \) inter-changed.

With the result we already obtained, \( G_1(s) = \frac{1 \pm \sqrt{1 - 4pq}}{2qs} \), we will have

\[
G_1^*(s) = \frac{1 \pm \sqrt{1 - 4pq}}{2ps}
\]

combine those two formulas, we obtain

\[
G_0(s) = psG_1^*(s) + qsG_1(s) = 1 - \sqrt{1 - 4pq}
\]

Again, we set \( s=1 \), we obtain

\[
G_0(1) = \sum_{n=1}^{\infty} f_0(n) = P(\text{walk ever returns to 0}|X_0 = 0) = 1 - \sqrt{1 - 4pq}
\]
Finally we summarize those key formulas we have reached,

$$G_0(1) = 1 - \sqrt{1 - 4pq}$$

$$G_1(1) = \frac{1 - \sqrt{1 - 4pq}}{2q}$$

$$G_r(1) = G_1(1)^r$$

Here we set $s=1$ then we obtain the meaning behind the $G_r(1)$:

$$G_r(1) = P(\text{walk ever reaches to } r | X_0 = 0)$$

Distribution of Steps from Level 0 to Level R: up till now, we used the generating function to derive the possibility of random walk reach to a certain level. It’s necessary to explore how the steps can be distributed from level 0 to level $r$. Here we define $n$ the steps a simple random walk to take to reach position $r$. Here we define $n$ the steps a simple random walk to take to reach position $r$. One can easily get the formulas below: $a$ represents the step to the right(or up) and $b$ represents the step to the left(or down).

$$a - b = r$$

$$a + b = n$$

$$p + q = 1$$
After solving the equations above, one can easily obtain the expressions for a and b respectively below:

\[
a = \frac{n + r}{2}
\]

\[
b = \frac{n - r}{2}
\]

The possibility of using n steps to reach position r is:

\[
P_r(n) = C_{a+b}^{n} p^a q^b = C_n^{\frac{n+r}{2}} \left(\frac{p}{2}\right)^{\frac{n+r}{2}} \left(\frac{q}{2}\right)^{\frac{n-r}{2}} \text{ (for } n + r = \text{ even)}
\]

\[
P_r(n) = \begin{cases} 
0; & \text{ } n + r = odd \\
\frac{n+r}{2} \left(\frac{p}{2}\right)^{\frac{n+r}{2}} \left(\frac{q}{2}\right)^{\frac{n-r}{2}}; & \text{ } n + r = even
\end{cases}
\]

we can use the matlab to plot such relationship. See fig 1. Here we set r=2, nMax=1000, which means the maximum step is confined to 1000.
Fig 1. The distribution of steps from 0 to 2. The maximum steps are confined to 1000.

However, we haven’t used the generating function approach to give a explicit expression of the distribution. But it is worth trying in the future work.