Chapter 7: Generating Functions

This chapter looks at Probability Generating Functions (PGFs) for ***discrete*** random variables. PGFs are useful tools for dealing with sums and limits of random variables. For some stochastic processes, they also have a special role in telling us whether a process will *ever* reach a particular state.

By the end of this chapter, you should be able to:

- find the sum of Geometric, Binomial, and Exponential series;
- know the definition of the PGF, and use it to calculate the mean, variance, and probabilities;
- calculate the PGF for Geometric, Binomial, and Poisson distributions;
- calculate the PGF for a randomly stopped sum;
- calculate the PGF for first reaching times in the random walk;
- use the PGF to determine whether a process will *ever* reach a given state.

7.1 Common sums

1. Geometric Series

$$1 + r + r^2 + r^3 + \ldots \quad = \quad \sum_{x=0}^{\infty} r^x \quad = \quad \frac{1}{1-r}, \qquad \textit{when } |r| < 1.$$

This formula proves that $\sum_{x=0}^{\infty} \mathbb{P}(X = x) = 1$ when $X \sim \text{Geometric}(p)$:

$$\mathbb{P}(X = x) = p(1-p)^x \quad \Rightarrow \quad \sum_{x=0}^{\infty} \mathbb{P}(X = x) = \sum_{x=0}^{\infty} p(1-p)^x$$

$$= p \sum_{x=0}^{\infty} (1-p)^x$$

$$= \frac{p}{1-(1-p)} \qquad \textit{(because } |1-p| < 1\textit{)}$$

$$= 1.$$

2. Binomial Theorem *For any $p, q \in \mathbb{R}$, and integer n,*

$$(p+q)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}.$$

Note that $\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$ *(${}^nC_r$ button on calculator.)*

The Binomial Theorem proves that $\sum_{x=0}^{n} \mathbb{P}(X=x) = 1$ when $X \sim \text{Binomial}(n,p)$: $\mathbb{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ *for* $x = 0, 1, \ldots, n$*, so*

$$\begin{aligned} \sum_{x=0}^{n} \mathbb{P}(X=x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \Big(p + (1-p)\Big)^n \\ &= 1^n \\ &= 1. \end{aligned}$$

3. Exponential Power Series

For any $\lambda \in \mathbb{R}$, $\quad \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}.$

This proves that $\sum_{x=0}^{\infty} \mathbb{P}(X=x) = 1$ when $X \sim \text{Poisson}(\lambda)$:

$\mathbb{P}(X=x) = \frac{\lambda^x}{x!} e^{-\lambda}$ *for* $x = 0, 1, 2, \ldots$*, so*

$$\begin{aligned} \sum_{x=0}^{\infty} \mathbb{P}(X=x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} e^{\lambda} \\ &= 1. \end{aligned}$$

Note: Another useful identity is: $\quad e^{\lambda} = \lim_{n \to \infty} \left(1 + \frac{\lambda}{n}\right)^n \quad$ *for* $\lambda \in \mathbb{R}$.

7.2 Probability Generating Functions

The **probability generating function (PGF)** is a useful tool for dealing with ***discrete*** random variables taking values $0, 1, 2, \ldots$. Its particular strength is that it gives us an easy way of characterizing the distribution of $X + Y$ when X and Y are independent. In general it is difficult to find the distribution of a sum using the traditional probability function. The PGF transforms a sum into a product and enables it to be handled much more easily.

Sums of random variables are particularly important in the study of stochastic processes, because many stochastic processes are formed from the sum of a sequence of repeating steps: for example, the Gambler's Ruin from Section 2.7.

The name *probability generating function* also gives us another clue to the role of the PGF. The PGF can be used to generate all the probabilities of the distribution. This is generally tedious and is not often an efficient way of calculating probabilities. However, the fact that it *can* be done demonstrates that *the PGF tells us everything there is to know about the distribution.*

Definition: Let X be a discrete random variable taking values in the non-negative integers $\{0, 1, 2, \ldots\}$. The **probability generating function (PGF)** of X is $G_X(s) = \mathbb{E}(s^X)$, *for all* $s \in \mathbb{R}$ *for which the sum converges.*

Calculating the probability generating function

$$G_X(s) = \mathbb{E}\left(s^X\right) = \sum_{x=0}^{\infty} s^x \mathbb{P}(X = x).$$

Properties of the PGF:

1. $G_X(0) = \mathbb{P}(X = 0)$:

$$\begin{aligned} G_X(0) &= 0^0 \times \mathbb{P}(X = 0) + 0^1 \times \mathbb{P}(X = 1) + 0^2 \times \mathbb{P}(X = 2) + \ldots \\ \therefore \quad G_X(0) &= \mathbb{P}(X = 0). \end{aligned}$$

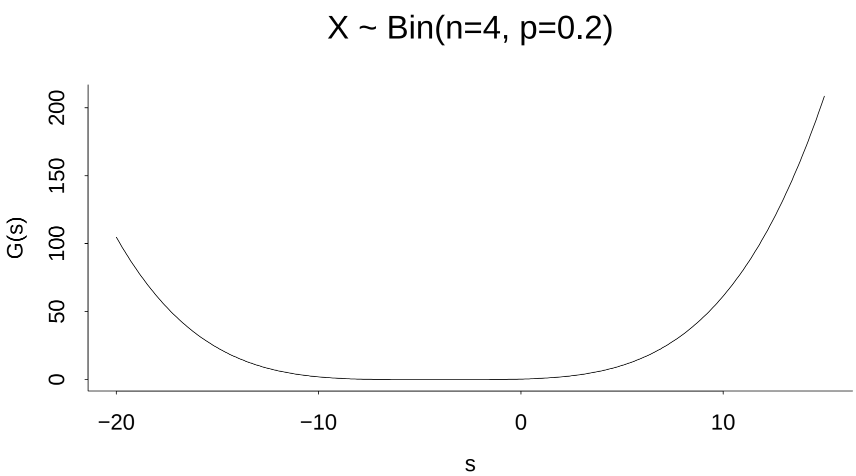
2. $G_X(1) = 1$: $$G_X(1) = \sum_{x=0}^{\infty} 1^x \mathbb{P}(X = x) = \sum_{x=0}^{\infty} \mathbb{P}(X = x) = 1.$$

Example 1: **Binomial Distribution**

Let $X \sim \text{Binomial}(n, p)$, so $\mathbb{P}(X = x) = \binom{n}{x} p^x q^{n-x}$ for $x = 0, 1, \ldots, n$.

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{n} s^x \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (ps)^x q^{n-x} \\ &= (ps+q)^n \quad \textit{by the Binomial Theorem: true for all } s. \end{aligned}$$

Thus $G_X(s) = (ps+q)^n$ *for all* $s \in \mathbb{R}$.



Check $G_X(0)$:

$$\begin{aligned} G_X(0) &= (p \times 0 + q)^n \\ &= q^n \\ &= \mathbb{P}(X = 0). \end{aligned}$$

Check $G_X(1)$:

$$\begin{aligned} G_X(1) &= (p \times 1 + q)^n \\ &= (1)^n \\ &= 1. \end{aligned}$$

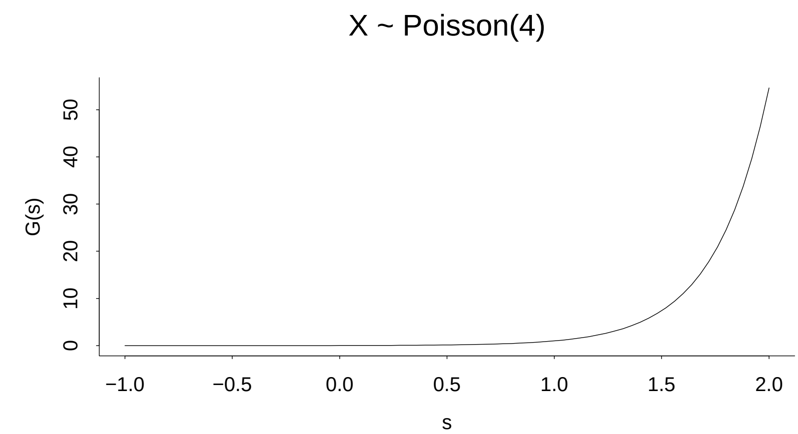
Example 2: **Poisson Distribution**

Let $X \sim \text{Poisson}(\lambda)$, so $\mathbb{P}(X = x) = \dfrac{\lambda^x}{x!}e^{-\lambda}$ for $x = 0, 1, 2, \ldots$.

$$G_X(s) = \sum_{x=0}^{\infty} s^x \frac{\lambda^x}{x!} e^{-\lambda} \quad = \quad e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda s)^x}{x!}$$

$$= \quad e^{-\lambda} e^{(\lambda s)} \quad \textit{for all } s \in \mathbb{R}.$$

Thus $\quad G_X(s) = e^{\lambda(s-1)} \quad$ *for all* $s \in \mathbb{R}$.



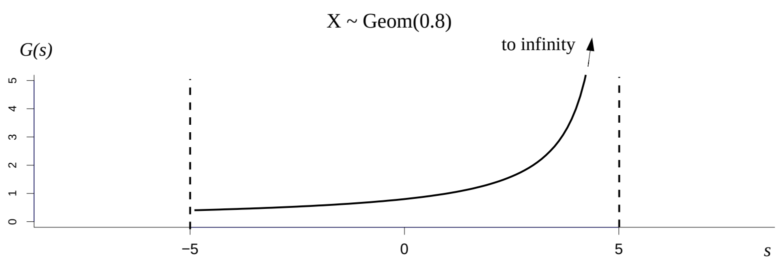
Example 3: **Geometric Distribution**

Let $X \sim \text{Geometric}(p)$, so $\mathbb{P}(X = x) = p(1-p)^x = pq^x$ for $x = 0, 1, 2, \ldots$, where $q = 1 - p$.

$$G_X(s) \quad = \quad \sum_{x=0}^{\infty} s^x p q^x$$

$$= \quad p \sum_{x=0}^{\infty} (qs)^x$$

$$= \quad \frac{p}{1-qs} \quad \textit{for all } s \textit{ such that } |qs| < 1.$$



Thus $\quad G_X(s) = \dfrac{p}{1-qs} \quad$ *for* $\quad |s| < \dfrac{1}{q}$.

7.3 Using the probability generating function to calculate probabilities

The probability generating function gets its name because the power series can be expanded and differentiated to reveal the individual probabilities. Thus, ***given only the PGF*** $G_X(s) = \mathbb{E}(s^X)$***, we can recover all probabilities*** $\mathbb{P}(X = x)$.

For shorthand, write $p_x = \mathbb{P}(X = x)$. Then

$$G_X(s) = \mathbb{E}(s^X) = \sum_{x=0}^{\infty} p_x s^x = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + p_4 s^4 + \dots$$

Thus $p_0 = \mathbb{P}(X = 0) = G_X(0)$.

First derivative: $G'_X(s) = p_1 + 2p_2 s + 3p_3 s^2 + 4p_4 s^3 + \dots$

Thus $p_1 = \mathbb{P}(X = 1) = G'_X(0)$.

Second derivative: $G''_X(s) = 2p_2 + (3 \times 2)p_3 s + (4 \times 3)p_4 s^2 + \dots$

Thus $p_2 = \mathbb{P}(X = 2) = \frac{1}{2} G''_X(0)$.

Third derivative: $G'''_X(s) = (3 \times 2 \times 1)p_3 + (4 \times 3 \times 2)p_4 s + \dots$

Thus $p_3 = \mathbb{P}(X = 3) = \frac{1}{3!} G'''_X(0)$.

In general:

$$p_n = \mathbb{P}(X = n) = \left(\frac{1}{n!}\right) G_X^{(n)}(0) = \left(\frac{1}{n!}\right) \frac{d^n}{ds^n} \left(G_X(s)\right)\Big|_{s=0}.$$

Example: Let X be a discrete random variable with PGF $G_X(s) = \dfrac{s}{5}(2 + 3s^2)$. Find the distribution of X.

$$G_X(s) = \frac{2}{5}s + \frac{3}{5}s^3 : \qquad G_X(0) = \mathbb{P}(X = 0) = 0.$$

$$G'_X(s) = \frac{2}{5} + \frac{9}{5}s^2 : \qquad G'_X(0) = \mathbb{P}(X = 1) = \frac{2}{5}.$$

$$G''_X(s) = \frac{18}{5}s : \qquad \frac{1}{2}G''_X(0) = \mathbb{P}(X = 2) = 0.$$

$$G'''_X(s) = \frac{18}{5} : \qquad \frac{1}{3!}G'''_X(0) = \mathbb{P}(X = 3) = \frac{3}{5}.$$

$$G_X^{(r)}(s) = 0 \ \forall r \geq 4 : \qquad \frac{1}{r!}G_X^{(r)}(s) = \mathbb{P}(X = r) = 0 \ \forall r \geq 4.$$

Thus

$$X = \begin{cases} 1 \text{ with probability } 2/5, \\ 3 \text{ with probability } 3/5. \end{cases}$$

Uniqueness of the PGF

The formula $p_n = \mathbb{P}(X = n) = \left(\dfrac{1}{n!}\right) G_X^{(n)}(0)$ shows that the whole sequence of probabilities $p_0, p_1, p_2, \ldots$ is determined by the values of the PGF and its derivatives at $s = 0$. It follows that the PGF specifies a ***unique*** set of probabilities.

Fact: If two power series agree on any interval containing 0, however small, then all terms of the two series are equal.

Formally: let $A(s)$ and $B(s)$ be PGFs with $A(s) = \sum_{n=0}^{\infty} a_n s^n$, $B(s) = \sum_{n=0}^{\infty} b_n s^n$. If there exists some $R' > 0$ such that $A(s) = B(s)$ for all $-R' < s < R'$, then $a_n = b_n$ for all n.

Practical use: If we can show that two random variables have the same PGF in some interval containing 0, then we have shown that *the two random variables have the same distribution.*

Another way of expressing this is to say that *the PGF of X tells us everything there is to know about the distribution of X.*

7.4 Expectation and moments from the PGF

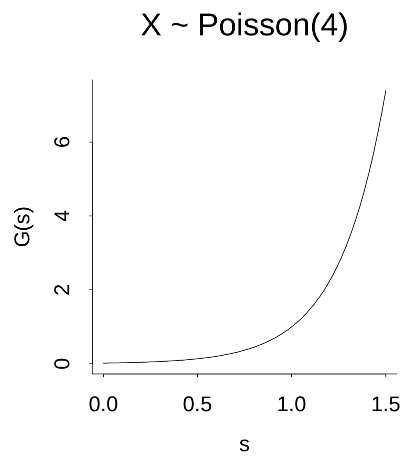
As well as calculating probabilities, we can also use the PGF to calculate the moments of the distribution of X. The moments of a distribution are *the mean, variance, etc.*

Theorem 7.4: Let X be a discrete random variable with PGF $G_X(s)$. Then:

1. $\mathbb{E}(X) = G'_X(1)$.
2. $\mathbb{E}\Big\{X(X-1)(X-2)\ldots(X-k+1)\Big\} = G_X^{(k)}(1) = \left.\dfrac{d^k\, G_X(s)}{ds^k}\right|_{s=1}$.
   (This is the kth factorial moment of X.)

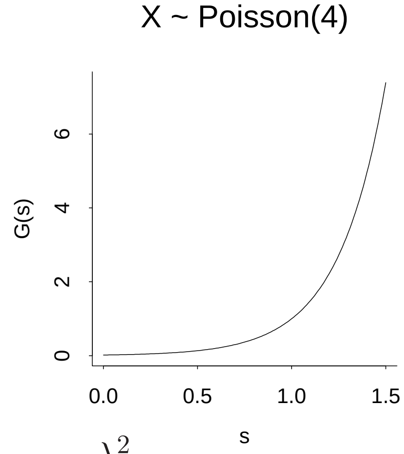
Proof: (Sketch: see Section 7.8 for more details)

1.
$$\begin{aligned} G_X(s) &= \sum_{x=0}^{\infty} s^x\, p_x, \\ \text{so} \quad G'_X(s) &= \sum_{x=0}^{\infty} x s^{x-1} p_x \\ \Rightarrow \quad G'_X(1) &= \sum_{x=0}^{\infty} x p_x = \mathbb{E}(X) \end{aligned}$$



2.
$$\begin{aligned} G_X^{(k)}(s) = \frac{d^k\, G_X(s)}{ds^k} &= \sum_{x=k}^{\infty} x(x-1)(x-2)\ldots(x-k+1)s^{x-k}p_x \\ \text{so} \quad G_X^{(k)}(1) &= \sum_{x=k}^{\infty} x(x-1)(x-2)\ldots(x-k+1)p_x \\ &= \mathbb{E}\Big\{X(X-1)(X-2)\ldots(X-k+1)\Big\}. \qquad \square \end{aligned}$$

Example: Let $X \sim \text{Poisson}(\lambda)$. The PGF of X is $G_X(s) = e^{\lambda(s-1)}$. Find $\mathbb{E}(X)$ and $\text{Var}(X)$.



Solution:

$$\begin{aligned} G_X'(s) &= \lambda e^{\lambda(s-1)} \\ \Rightarrow \quad \mathbb{E}(X) &= G_X'(1) = \lambda. \end{aligned}$$

For the variance, consider

$$\mathbb{E}\Big\{X(X-1)\Big\} = G_X''(1) = \lambda^2 e^{\lambda(s-1)}|_{s=1} = \lambda^2.$$

So

$$\begin{aligned} \mathit{Var}(X) &= \mathbb{E}(X^2) - (\mathbb{E}X)^2 \\ &= \mathbb{E}\Big\{X(X-1)\Big\} + \mathbb{E}X - (\mathbb{E}X)^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

7.5 Probability generating function for a sum of independent r.v.s

One of the PGF's greatest strengths is that it turns a sum into a product:

$$\mathbb{E}\left(s^{(X_1+X_2)}\right) = \mathbb{E}\left(s^{X_1} s^{X_2}\right).$$

This makes the PGF useful for finding the probabilities and moments of *a sum of independent random variables.*

Theorem 7.5: Suppose that $X_1, \ldots, X_n$ are ***independent*** random variables, and let $Y = X_1 + \ldots + X_n$. Then

$$G_Y(s) = \prod_{i=1}^{n} G_{X_i}(s).$$

Proof:

$$\begin{aligned} G_Y(s) &= \mathbb{E}(s^{(X_1+\ldots+X_n)}) \\ &= \mathbb{E}(s^{X_1}s^{X_2}\ldots s^{X_n}) \\ &= \mathbb{E}(s^{X_1})\mathbb{E}(s^{X_2})\ldots\mathbb{E}(s^{X_n}) \end{aligned}$$

(because $X_1, \ldots, X_n$ are independent)

$$= \prod_{i=1}^{n} G_{X_i}(s). \quad \textit{as required.} \quad \square$$

Example: Suppose that X and Y are independent with $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$. Find the distribution of $X + Y$.

Solution:

$$\begin{aligned} G_{X+Y}(s) &= G_X(s) \cdot G_Y(s) \\ &= e^{\lambda(s-1)}e^{\mu(s-1)} \\ &= e^{(\lambda+\mu)(s-1)}. \end{aligned}$$

But this is the PGF of the Poisson$(\lambda + \mu)$ distribution. So, by the uniqueness of PGFs, $X + Y \sim$ Poisson$(\lambda + \mu)$.

7.6 Randomly stopped sum

Remember the randomly stopped sum model from Section 3.4. A random number N of events occur, and each event i has associated with it a cost or reward X_i. The question is to find the distribution of the total cost or reward: $T_N = X_1 + X_2 + \ldots + X_N$. T_N is called a *randomly stopped sum* because it has a random number of terms.

Example: Cash machine model. N customers arrive during the day. Customer i withdraws amount X_i. The total amount withdrawn during the day is $T_N = X_1 + \ldots + X_N$.

In Chapter 3, we used the Laws of Total Expectation and Variance to show that $\mathbb{E}(T_N) = \mu\,\mathbb{E}(N)$ and $\text{Var}(T_N) = \sigma^2\,\mathbb{E}(N) + \mu^2\,\text{Var}(N)$, where $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \text{Var}(X_i)$.

In this chapter we will now use probability generating functions to investigate the *whole distribution of* T_N.

Theorem 7.6: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common PGF G_X. Let N be a random variable, independent of the X_i's, with PGF G_N, and let $T_N = X_1 + \ldots + X_N = \sum_{i=1}^{N} X_i$. Then the PGF of T_N is:

$$G_{T_N}(s) = G_N\Big(G_X(s)\Big).$$

Proof:

$$\begin{aligned}
G_{T_N}(s) &= \mathbb{E}(s^{T_N}) = \mathbb{E}\left(s^{X_1+\ldots+X_N}\right) \\
&= \mathbb{E}_N\Big\{\mathbb{E}\left(s^{X_1+\ldots+X_N} \,\Big|\, N\right)\Big\} \quad \textit{(conditional expectation)} \\
&= \mathbb{E}_N\Big\{\mathbb{E}\left(s^{X_1} \ldots s^{X_N} \mid N\right)\Big\} \\
&= \mathbb{E}_N\Big\{\mathbb{E}\left(s^{X_1} \ldots s^{X_N}\right)\Big\} \quad (X_i\textit{'s are indept of } N) \\
&= \mathbb{E}_N\Big\{\mathbb{E}\left(s^{X_1}\right) \ldots \mathbb{E}\left(s^{X_N}\right)\Big\} \quad (X_i\textit{'s are indept of each other}) \\
&= \mathbb{E}_N\Big\{(G_X(s))^N\Big\} \\
&= G_N\Big(G_X(s)\Big) \quad \textit{(by definition of } G_N\textit{)}. \qquad \square
\end{aligned}$$

Example: Let $X_1, X_2, \ldots$ and N be as above. Find the mean of T_N.

$$\begin{aligned}
\mathbb{E}(T_N) = G'_{T_N}(1) &= \frac{d}{ds} G_N(G_X(s))\Big|_{s=1} \\
&= G'_N(G_X(s)) \cdot G'_X(s)\Big|_{s=1} \\
&= G'_N(1) \cdot G'_X(1) \qquad \textit{Note: } G_X(1) = 1 \textit{ for any r.v. } X \\
&= \mathbb{E}(N) \cdot \mathbb{E}(X_1), \qquad \textit{— same answer as in Chapter 3.}
\end{aligned}$$

Example: <u>**Heron goes fishing**</u>

My aunt was asked by her neighbours to feed the prize goldfish in their garden pond while they were on holiday. Although my aunt dutifully went and fed them every day, she never saw a single fish for the whole three weeks. It turned out that all the fish had been eaten by a heron when she wasn't looking!

Let N be the number of times the heron visits the pond during the neighbours' absence. Suppose that $N \sim \text{Geometric}(1 - \theta)$, so $\mathbb{P}(N = n) = (1 - \theta)\theta^n$, for $n = 0, 1, 2, \ldots$. When the heron visits the pond it has probability p of catching a prize goldfish, independently of what happens on any other visit. (This assumes that there are infinitely many goldfish to be caught!) Find the distribution of

$$T = \text{total number of goldfish caught.}$$

Solution:

$$\textit{Let} \quad X_i = \begin{cases} 1 & \textit{if heron catches a fish on visit } i, \\ 0 & \textit{otherwise.} \end{cases}$$

Then $T = X_1 + X_2 + \ldots + X_N$ *(randomly stopped sum), so*

$$G_T(s) = G_N(G_X(s)).$$

Now

$$G_X(s) = \mathbb{E}(s^X) = s^0 \times \mathbb{P}(X=0) + s^1 \times \mathbb{P}(X=1) = 1 - p + ps.$$

Also,

$$\begin{aligned} G_N(r) = \sum_{n=0}^{\infty} r^n \mathbb{P}(N=n) &= \sum_{n=0}^{\infty} r^n (1-\theta)\theta^n \\ &= (1-\theta)\sum_{n=0}^{\infty}(\theta r)^n \\ &= \frac{1-\theta}{1-\theta r}. \qquad\qquad (r < 1/\theta). \end{aligned}$$

So

$$G_T(s) = \frac{1-\theta}{1-\theta\, G_X(s)} \qquad \textit{(putting } r = G_X(s)\textit{)},$$

giving:

$$\begin{aligned} G_T(s) &= \frac{1-\theta}{1-\theta(1-p+ps)} \\ &= \frac{1-\theta}{1-\theta+\theta p-\theta ps} \end{aligned}$$

[could this be Geometric? $G_T(s) = \dfrac{1-\pi}{1-\pi s}$ *for some* π*?]*

$$\begin{aligned} &= \frac{1-\theta}{(1-\theta+\theta p)-\theta ps} \\ &= \frac{\left(\dfrac{1-\theta}{1-\theta+\theta p}\right)}{\left(\dfrac{(1-\theta+\theta p)-\theta ps}{1-\theta+\theta p}\right)} \end{aligned}$$

$$= \frac{\left(\dfrac{1-\theta+\theta p-\theta p}{1-\theta+\theta p}\right)}{1-\left(\dfrac{\theta p}{1-\theta+\theta p}\right)s}$$

$$= \frac{1-\left(\dfrac{\theta p}{1-\theta+\theta p}\right)}{1-\left(\dfrac{\theta p}{1-\theta+\theta p}\right)s}.$$

This is the PGF of the Geometric $\left(1-\dfrac{\theta p}{1-\theta+\theta p}\right)$ *distribution, so by uniqueness of PGFs, we have:*

$$T \sim \textit{Geometric}\left(\frac{1-\theta}{1-\theta+\theta p}\right).$$

Why did we need to use the PGF?

We could have solved the heron problem without using the PGF, but it is much more difficult. PGFs are very useful for dealing with sums of random variables, which are difficult to tackle using the standard probability function.

Here are the first few steps of solving the heron problem without the PGF. Recall the problem:

- Let $N \sim \text{Geometric}(1-\theta)$, so $\mathbb{P}(N=n) = (1-\theta)\theta^n$;
- Let $X_1, X_2, \ldots$ be independent of each other and of N, with $X_i \sim \text{Binomial}(1, p)$ (remember $X_i = 1$ with probability p, and 0 otherwise);
- Let $T = X_1 + \ldots + X_N$ be the randomly stopped sum;
- Find the distribution of T.

Without using the PGF, we would tackle this by looking for an expression for $\mathbb{P}(T = t)$ for any t. Once we have obtained that expression, we might be able to see that T has a distribution we recognise (e.g. Geometric), or otherwise we would just state that T is defined by the probability function we have obtained.

To find $\mathbb{P}(T = t)$, we have to ***partition over different values of*** N***:***

$$\mathbb{P}(T = t) = \sum_{n=0}^{\infty} \mathbb{P}(T = t \,|\, N = n)\mathbb{P}(N = n). \qquad (\star)$$

Here, we are ***lucky*** that we can write down the distribution of $T \,|\, N = n$:

- if $N = n$ is fixed, then $T = X_1 + \ldots + X_n$ is a sum of n independent Binomial$(1, p)$ random variables, so $(T \,|\, N = n) \sim$ ***Binomial***(n, p).

For most distributions of X, ***it would be difficult or impossible to write down the distribution of*** $X_1 + \ldots + X_n$***:***

we would have to use an expression like

$$\mathbb{P}(X_1 + \ldots + X_N = t \,|\, N = n) = \sum_{x_1=0}^{t} \sum_{x_2=0}^{t-x_1} \ldots \sum_{x_{n-1}=0}^{t-(x_1+\ldots+x_{n-2})} \Big\{ \mathbb{P}(X_1 = x_1) \times$$

$$\mathbb{P}(X_2 = x_2) \times \ldots \times \mathbb{P}(X_{n-1} = x_{n-1}) \times \mathbb{P}[X_n = t - (x_1 + \ldots + x_{n-1})] \Big\}.$$

Back to the heron problem: we are lucky in this case that we know the distribution of $(T \,|\, N = n)$ is Binomial$(N = n,\ p)$, so

$$\mathbb{P}(T = t \,|\, N = n) = \binom{n}{t} p^t (1 - p)^{n-t} \quad \text{for } t = 0, 1, \ldots, n.$$

Continuing from $(\star)$:

$$\mathbb{P}(T = t) \;=\; \sum_{n=0}^{\infty} \mathbb{P}(T = t \,|\, N = n)\mathbb{P}(N = n)$$

$$= \sum_{n=t}^{\infty} \binom{n}{t} p^t (1-p)^{n-t} (1-\theta)\theta^n$$

$$= (1-\theta) \left(\frac{p}{1-p}\right)^t \sum_{n=t}^{\infty} \binom{n}{t} \Big[\theta(1-p)\Big]^n \qquad (\star\star)$$

$$= \ldots ?$$

As it happens, we can evaluate the sum in $(\star\star)$ using the fact that Negative Binomial probabilities sum to 1. You can try this if you like, but it is quite tricky. [Hint: use the Negative Binomial $(t+1, 1-\theta(1-p))$ distribution.]

Overall, we obtain the same answer that $T \sim \text{Geometric}\left(\frac{1-\theta}{1-\theta+\theta p}\right)$, but hopefully you can see why the PGF is so useful.

Without the PGF, we have two major difficulties:

1. *Writing down* $\mathbb{P}(T = t \,|\, N = n)$*;*
2. *Evaluating the sum over* n *in* $(\star\star)$.

For a general problem, both of these steps might be too difficult to do without a computer. The PGF has none of these difficulties, and even if $G_T(s)$ does not simplify readily, it still tells us everything there is to know about the distribution of T.

7.7 Summary: Properties of the PGF

Definition:	$G_X(s) = \mathbb{E}(s^X)$	
Used for:	Discrete r.v.s with values 0, 1, 2, ...	
Moments:	$\mathbb{E}(X) = G'_X(1)$	$\mathbb{E}\Big\{X(X-1)\ldots(X-k+1)\Big\} = G_X^{(k)}(1)$
Probabilities:	$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0)$	
Sums:	$G_{X+Y}(s) = G_X(s) G_Y(s)$ for independent X, Y	

7.8 Convergence of PGFs

We have been using PGFs throughout this chapter without paying much attention to their mathematical properties. For example, are we sure that the power series $G_X(s) = \sum_{x=0}^{\infty} s^x \mathbb{P}(X = x)$ converges? Can we differentiate and integrate the infinite power series term by term as we did in Section 7.4? When we said in Section 7.4 that $\mathbb{E}(X) = G'_X(1)$, can we be sure that $G_X(1)$ and its derivative $G'_X(1)$ even exist?

This technical section introduces the **radius of convergence** of the PGF. Although it isn't obvious, it is always safe to assume convergence of $G_X(s)$ at least for $|s| < 1$. Also, there are results that assure us that $\mathbb{E}(X) = G'_X(1)$ will work for all non-defective random variables X.

Definition: The **radius of convergence** of a probability generating function is *a number $R > 0$, such that the sum $G_X(s) = \sum_{x=0}^{\infty} s^x \mathbb{P}(X = x)$ converges if $|s| < R$ and diverges ($\to \infty$) if $|s| > R$.*

(No general statement is made about what happens when $|s| = R$.)

Fact: For any PGF, the radius of convergence exists.

It is always ≥ 1: every PGF converges for at least $s \in (-1, 1)$.

The radius of convergence could be anything from $R = 1$ to $R = \infty$.

Note: This gives us the surprising result that the set of s for which the PGF $G_X(s)$ converges is symmetric about 0: the PGF converges for all $s \in (-R, R)$, and for no $s < -R$ or $s > R$.

This is surprising because the PGF itself is not usually symmetric about 0: i.e. $G_X(-s) \neq G_X(s)$ in general.

Example 1: **Geometric distribution**

Let $X \sim \text{Geometric}(p = 0.8)$. What is the radius of convergence of $G_X(s)$?

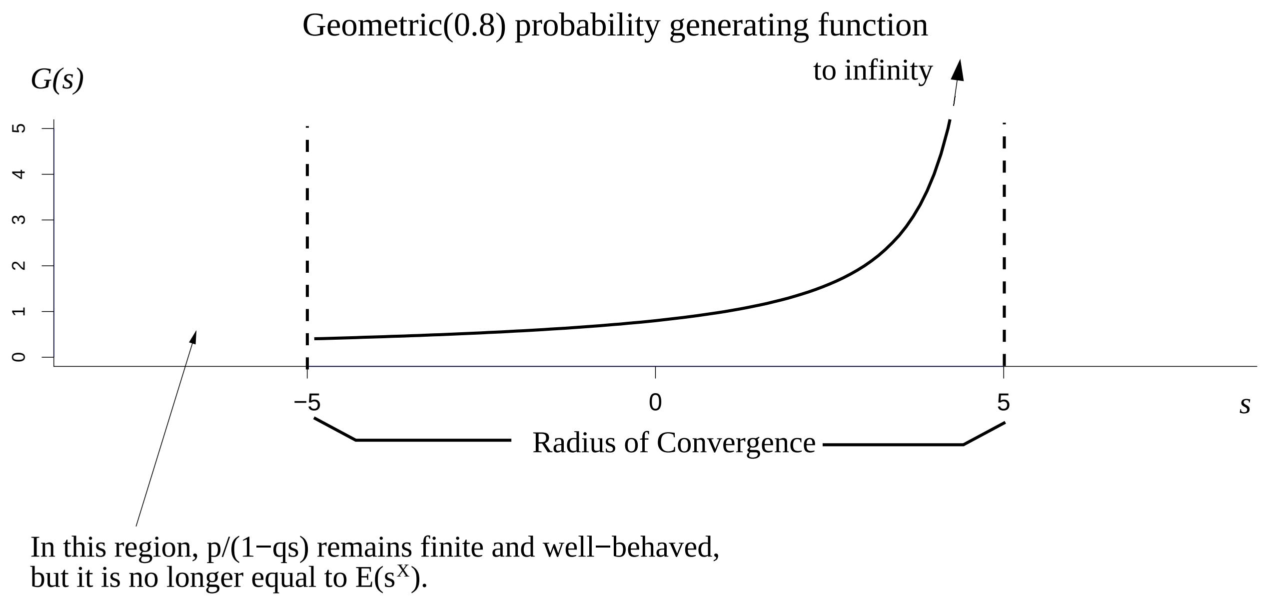
As in Section 7.2,

$$G_X(s) = \sum_{x=0}^{\infty} s^x (0.8)(0.2)^x = 0.8 \sum_{x=0}^{\infty} (0.2s)^x$$

$$= \frac{0.8}{1 - 0.2s} \quad \textit{for all } s \textit{ such that } |0.2s| < 1.$$

This is valid for all s *with* $|0.2s| < 1$*, so it is valid for all* s *with* $|s| < \frac{1}{0.2} = 5$.
(i.e. $-5 < s < 5$.*)*
The radius of convergence is $R = 5$.

The figure shows the PGF of the Geometric($p = 0.8$) distribution, with its radius of convergence $R = 5$. Note that although the convergence set $(-5, 5)$ is symmetric about 0, the function $G_X(s) = p/(1 - qs) = 4/(5 - s)$ is not.



At the limits of convergence, strange things happen:

- At the positive end, as $s \uparrow 5$, both $G_X(s)$ and $p/(1 - qs)$ approach infinity. So the PGF *is* (left)-continuous at $+R$:

$$\lim_{s \uparrow 5} G_X(s) = G_X(5) = \infty.$$

However, the PGF does *not* converge at $s = +R$.

- At the negative end, as $s \downarrow -5$, the function $p/(1-qs) = 4/(5-s)$ is continuous and passes through 0.4 when $s=-5$. However, when $s \leq -5$, this function no longer represents $G_X(s) = 0.8\sum_{x=0}^{\infty}(0.2s)^x$, because $|0.2s| \geq 1$.

  Additionally, when $s=-5$, $G_X(-5) = 0.8\sum_{x=0}^{\infty}(-1)^x$ does not exist. Unlike the positive end, this means that $G_X(s)$ is *not* (right)-continuous at $-R$:

$$\lim_{s\downarrow -5} G_X(s) = 0.4 \neq G_X(-5).$$

  Like the positive end, this PGF does *not* converge at $s=-R$.

Example 2: **Binomial distribution**

Let $X \sim \text{Binomial}(n,p)$. What is the radius of convergence of $G_X(s)$?

As in Section 7.2,

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{n} s^x \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (ps)^x q^{n-x} \\ &= (ps+q)^n \quad \textit{by the Binomial Theorem: true for all } s. \end{aligned}$$

This is true for all $-\infty < s < \infty$, so the radius of convergence is $R=\infty$.

Abel's Theorem for continuity of power series at $s=1$

Recall from above that if $X \sim \text{Geometric}(0.8)$, then $G_X(s)$ is not continuous at the negative end of its convergence ($-R$):

$$\lim_{s\downarrow -5} G_X(s) \neq G_X(-5).$$

Abel's theorem states that this sort of effect can never happen at $s=1$ (or at $+R$). In particular, $G_X(s)$ is always left-continuous at $s=1$:

$$\lim_{s\uparrow 1} G_X(s) = G_X(1) \;\text{ always, even if } G_X(1)=\infty.$$

Theorem 7.8: Abel's Theorem.

$$\text{Let} \quad G(s) = \sum_{i=0}^{\infty} p_i s^i \quad \text{for any } p_0, p_1, p_2, \ldots \text{ with } p_i \geq 0 \text{ for all } i.$$

Then $G(s)$ is left-continuous at $s = 1$:

$$\lim_{s \uparrow 1} G(s) = \sum_{i=0}^{\infty} p_i \quad = \quad G(1)\,,$$

whether or not this sum is finite.

Note: Remember that the radius of convergence $R \geq 1$ for any PGF, so Abel's Theorem means that even in the worst-case scenario when $R = 1$, we can still trust that the PGF will be continuous at $s = 1$. (By contrast, we can not be sure that the PGF will be continuous at the the lower limit $-R$).

Abel's Theorem means that for any PGF, we can write $G_X(1)$ as shorthand for $\lim_{s \uparrow 1} G_X(s)$.

It also clarifies our proof that $\mathbb{E}(X) = G'_X(1)$ from Section 7.4. If we assume that term-by-term differentiation is allowed for $G_X(s)$ (see below), then the proof on page 136 gives:

$$G_X(s) \;=\; \sum_{x=0}^{\infty} s^x\, p_x,$$

$$\text{so} \qquad G'_X(s) \;=\; \sum_{x=1}^{\infty} x s^{x-1} p_x \qquad \text{(term-by-term differentiation: see below).}$$

Abel's Theorem establishes that $\mathbb{E}(X)$ is equal to $\lim_{s \uparrow 1} G'_X(s)$:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x=1}^{\infty} x p_x \\ &= G'_X(1) \\ &= \lim_{s \uparrow 1} G'_X(s), \end{aligned}$$

because Abel's Theorem applies to $G'_X(s) = \sum_{x=1}^{\infty} x s^{x-1} p_x$, establishing that $G'_X(s)$ is left-continuous at $s = 1$. Without Abel's Theorem, we could not be sure that the limit of $G'_X(s)$ as $s \uparrow 1$ would give us the correct answer for $\mathbb{E}(X)$.

Absolute and uniform convergence for term-by-term differentiation

We have stated that the PGF converges for all $|s| < R$ for some R. In fact, the probability generating function converges *absolutely* if $|s| < R$. Absolute convergence is stronger than convergence alone: it means that the sum of absolute values, $\sum_{x=0}^{\infty} |s^x \mathbb{P}(X = x)|$, also converges. When two series both converge absolutely, the product series also converges absolutely. This guarantees that $G_X(s) \times G_Y(s)$ is absolutely convergent for any two random variables X and Y. This is useful because $G_X(s) \times G_Y(s) = G_{X+Y}(s)$ if X and Y are independent.

The PGF also converges *uniformly* on any set $\{s : |s| \leq R'\}$ where $R' < R$. Intuitively, this means that the speed of convergence does not depend upon the value of s. Thus a value n_0 can be found such that for all values of $n \geq n_0$, the *finite* sum $\sum_{x=0}^{n} s^x \mathbb{P}(X = x)$ is *simultaneously* close to the converged value $G_X(s)$, for all s with $|s| \leq R'$. In mathematical notation: $\forall \epsilon > 0$, $\exists n_0 \in \mathbb{Z}$ such that $\forall s$ with $|s| \leq R'$, and $\forall n \geq n_0$,

$$\left| \sum_{x=0}^{n} s^x \mathbb{P}(X = x) - G_X(s) \right| < \epsilon.$$

Uniform convergence allows us to differentiate or integrate the PGF term by term.

Fact: Let $G_X(s) = \mathbb{E}(s^X) = \sum_{x=0}^{\infty} s^x \mathbb{P}(X = x)$, and let $s < R$.

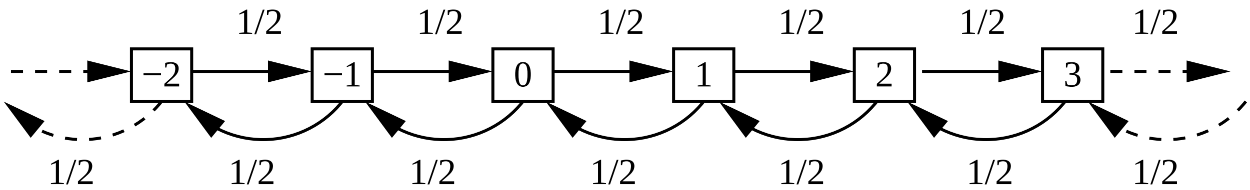
1. $$G'_X(s) = \frac{d}{ds}\left(\sum_{x=0}^{\infty} s^x \mathbb{P}(X = x)\right) = \sum_{x=0}^{\infty} \frac{d}{ds}\left(s^x \mathbb{P}(X = x)\right) = \sum_{x=0}^{\infty} x s^{x-1} \mathbb{P}(X = x).$$
(term by term differentiation).

2. $$\int_a^b G_X(s)\, ds = \int_a^b \left(\sum_{x=0}^{\infty} s^x \mathbb{P}(X = x)\right) ds = \sum_{x=0}^{\infty}\left(\int_a^b s^x \mathbb{P}(X = x)\, ds\right)$$
$$= \sum_{x=0}^{\infty} \frac{s^{x+1}}{x+1} \mathbb{P}(X = x) \quad \text{for } -R < a < b < R.$$
(term by term integration).

7.9 Special Process: the Random Walk

We briefly saw the Drunkard's Walk in Chapter 1: a drunk person staggers to left and right as he walks. This process is called the ***Random Walk*** in stochastic processes. Probability generating functions are particularly useful for processes such as the random walk, because the process is defined as the sum of a single repeating step. The repeating step is a move of one unit, left or right at random. The sum of the first t steps gives the position at time t.

The transition diagram below shows the *symmetric random walk (all transitions have probability* $p = 1/2$*.)*



Question:

What is the key difference between the random walk and the gambler's ruin?

The random walk has an INFINITE state space: it never stops. The gambler's ruin stops at both ends.

This fact has two important consequences:

- The random walk is hard to tackle using first-step analysis, because we would have to solve an ***infinite*** number of simultaneous equations. In this respect it might seem to be more difficult than the gambler's ruin.
- Because the random walk never stops, *all states are equal.*

  In the gambler's ruin, states are not equal: the states closest to 0 are more likely to end in ruin than the states closest to winning. By contrast, the random walk has no end-points, so (for example) the distribution of the time to reach state 5 starting from state 0 is exactly the same as the distribution of the time to reach state 1005 starting from state 1000. We can exploit this fact to solve some problems for the random walk that would be much more difficult to solve for the gambler's ruin.

PGFs for finding the distribution of reaching times

For random walks, we are particularly interested in ***reaching times:***

- How long will it take us to reach state j, starting from state i?
- Is there a chance that we will ***never*** reach state j, starting from state i?

In previous chapters we have seen how to find ***expected reaching times:*** the expected number of steps taken to reach a particular state. We used *the law of total expectation and first-step analysis (Section 3.5).*

However, the *expected* or *average* reaching time doesn't tell the whole story. Think back to the model for gene spread in Section 3.7. If there is just one animal out of 100 with the harmful allele, the expected number of generations to fixation is quite large at 10.5: even though the allele will usually die out after one or two generations. The high average is caused by a small chance that the allele will take hold and grow, requiring a very large number of generations before it either dies out or saturates the population. In most stochastic processes, the average is of limited use by itself, without having some idea about the *variance and skew of the distribution.*

With our tool of PGFs, we can characterise the *whole distribution* of the time T taken to reach a particular state, by finding its PGF. This will give us the mean, variance, and skew by differentiation. In principle the PGF could even give us the full set of probabilities, $\mathbb{P}(T = t)$ for all possible $t = 0, 1, 2, \ldots$, though in practice it may be computationally infeasible to find more than the first few probabilities by repeated differentiation.

However, there is a new and very useful piece of information that the PGF can tell us quickly and easily:

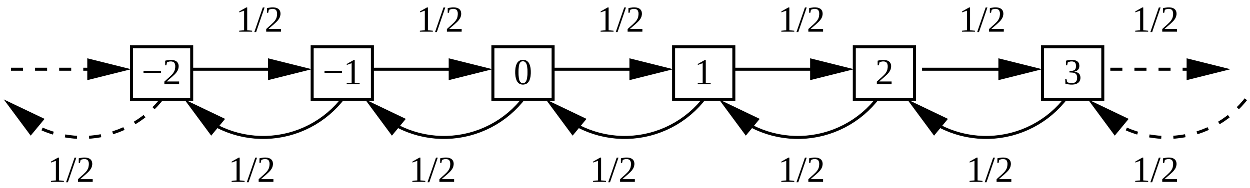
what is the probability that we NEVER reach state j, starting from state i?

For example, imagine that the random walk represents the share value for an investment. The current share price is i dollars, and we might decide to sell when it reaches j dollars. Knowing how long this might take, and whether there is a chance we will never succeed, is fundamental to managing our investment.

To tackle this problem, we define the random variable T to be the time taken (number of steps) to reach state j, starting from state i. We find the PGF of T, and then use the PGF to discover $\mathbb{P}(T = \infty)$. If $\mathbb{P}(T = \infty) > 0$, there is a positive chance that we will NEVER reach state j, starting from state i.

We will see how to determine the probability of never reaching our goal in Section 7.11. First we will see how to calculate the PGF of a reaching time T in the random walk.

Finding the PGF of a reaching time in the random walk

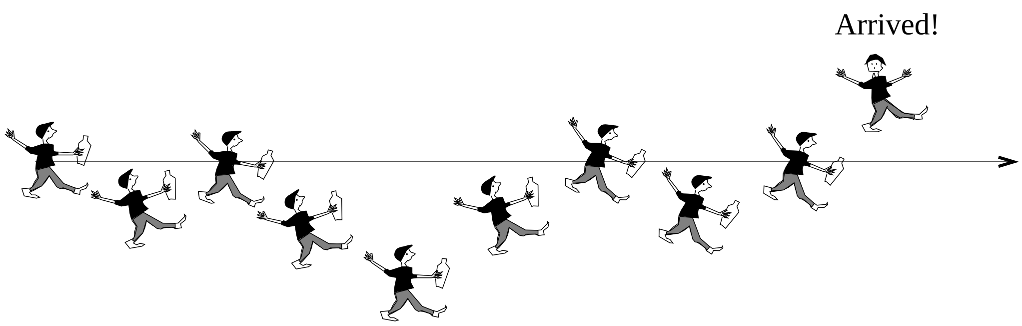


Define T_{ij} to be the *number of steps taken to reach state j, starting at state i.*

T_{ij} is called the *first reaching time from state i to state j.*

We will focus on T_{01} = *number of steps to get from state* 0 *to state* 1.

Problem: Let $H(s) = \mathbb{E}\left(s^{T_{01}}\right)$ be the PGF of T_{01}. Find $H(s)$.



Solution:

Let Y_n be the step taken at time n: up or down. For the symmetric random walk,

$$Y_n = \begin{cases} 1 & \textit{with probability } 0.5, \\ -1 & \textit{with probability } 0.5, \end{cases}$$

and $Y_1, Y_2, \ldots$ are independent.

Recall T_{ij} = number of steps to get from state i to state j for any i, j,

and $H(s) = \mathbb{E}\left(s^{T_{01}}\right)$ is the PGF required.

Use first-step analysis, partitioning over the first step Y_1:

$$\begin{aligned} H(s) &= \mathbb{E}\left(s^{T_{01}}\right) \\ &= \mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = 1\right) \mathbb{P}(Y_1 = 1) + \mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = -1\right) \mathbb{P}(Y_1 = -1) \\ &= \frac{1}{2}\left\{\mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = 1\right) + \mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = -1\right)\right\}. \qquad \spadesuit \end{aligned}$$

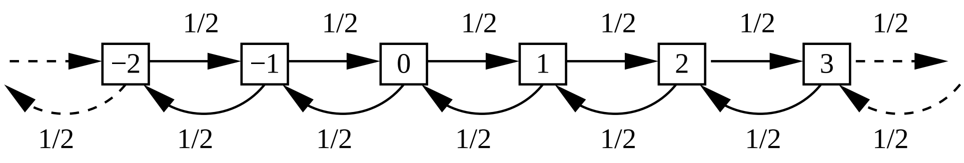
Now if $Y_1 = 1$, then $T_{01} = 1$ definitely, so $\mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = 1\right) = s^1 = s$.

If $Y_1 = -1$, then $T_{01} = 1 + T_{-1,1}$:

- $\rightarrow$ *one step from state 0 to state -1,*
- $\rightarrow$ *then $T_{-1,1}$ steps from state -1 to state 1.*

But $T_{-1,1} = T_{-1,0} + T_{01}$, because the process must pass through 0 to get from -1 to 1.

Now $T_{-1,0}$ and T_{01} are independent (Markov property). Also, they have the same distribution because the process is translation invariant (i.e. all states are the same):



Thus

$$\begin{aligned}\mathbb{E}\left(s^{T_{01}} \,|\, Y_1 = -1\right) &= \mathbb{E}\left(s^{1+T_{-1,1}}\right) \\ &= \mathbb{E}\left(s^{1+T_{-1,0}+T_{0,1}}\right) \\ &= s\mathbb{E}\left(s^{T_{-1,0}}\right)\mathbb{E}\left(s^{T_{01}}\right) \quad \textit{by independence} \\ &= s(H(s))^2 \quad \textit{because identically distributed.}\end{aligned}$$

Thus

$$H(s) = \frac{1}{2}\left\{s + s(H(s))^2\right\} \qquad \textit{by } \spadesuit.$$

This is a quadratic in $H(s)$*:*

$$\begin{aligned}\frac{1}{2}s(H(s))^2 - H(s) + \frac{1}{2}s &= 0 \\ \Rightarrow H(s) &= \frac{1 \pm \sqrt{1 - 4\frac{1}{2}s\frac{1}{2}s}}{s} = \frac{1 \pm \sqrt{1-s^2}}{s}.\end{aligned}$$

Which root? We know that $\mathbb{P}(T_{01} = 0) = 0$*, because it must take at least one step to go from 0 to 1. With the positive root,* $\lim_{s\to 0} H(0) = \lim_{s\to 0}\left(\frac{2}{s}\right) = \infty$*; so we take the negative root instead.*

$$\textit{Thus} \quad H(s) = \frac{1 - \sqrt{1-s^2}}{s}.$$

Check this has $\lim_{s\to 0} H(s) = 0$ *by L'Hospital's Rule:*

$$\begin{aligned}\lim_{s\to 0}\left(\frac{f(s)}{g(s)}\right) &= \lim_{s\to 0}\left(\frac{f'(s)}{g'(s)}\right) \\ &= \lim_{s\to 0}\left\{\frac{\frac{1}{2}(1-s^2)^{-1/2}\times 2s}{1}\right\} \\ &= 0.\end{aligned}$$

Notation for quick solutions of first-step analysis for finding PGFs

As with first-step analysis for finding hitting probabilities and expected reaching times, setting up a good notation is extremely important. Here is a good notation for finding $H(s) = \mathbb{E}\left(s^{T_{01}}\right)$.

Let $T = T_{01}$. *Seek* $H(s) = \mathbb{E}(s^T)$.
Now

$$T = \begin{cases} 1 & \textit{with probability } 1/2, \\ 1 + T' + T'' & \textit{with probability } 1/2, \end{cases}$$

where $T' \sim T'' \sim T$ *and* T', T'' *are independent.*

Taking expectations:

$$H(s) = \mathbb{E}(s^T) = \begin{cases} \mathbb{E}\left(s^1\right) & \textit{w. p. } 1/2 \\ \mathbb{E}\left(s^{1+T'+T''}\right) & \textit{w. p. } 1/2 \end{cases}$$

$$\Rightarrow \quad H(s) = \begin{cases} s & \textit{w. p. } 1/2 \\ s\mathbb{E}\left(s^{T'}\right)\mathbb{E}\left(s^{T''}\right) & \textit{w. p. } 1/2 \qquad \textit{(by independence of } T' \textit{ and } T''\textit{)} \end{cases}$$

$$\Rightarrow \quad H(s) = \begin{cases} s & \textit{w. p. } 1/2 \\ sH(s)H(s) & \textit{w. p. } 1/2 \qquad \textit{(because } T' \sim T'' \sim T\textit{)} \end{cases}$$

$$\Rightarrow \quad H(s) = \tfrac{1}{2}s + \tfrac{1}{2}sH(s)^2.$$

Thus:

$$sH(s)^2 - 2H(s) + s = 0.$$

Solve the quadratic and select the correct root as before, to get

$$H(s) = \frac{1 - \sqrt{1 - s^2}}{s} \qquad \text{for } |s| < 1.$$

7.10 Defective random variables

A random variable is said to be ***defective*** if *it can take the value* ∞*.*

In stochastic processes, *a reaching time* T_{ij} *is defective if there is a chance that we NEVER reach state* j*, starting from state* i*.*

The probability that we never reach state j, starting from state i, is the same as the probability that the time taken is infinite: $T_{ij} = \infty$:

$$\mathbb{P}(T_{ij} = \infty) = \mathbb{P}(\textit{we NEVER reach state } j\textit{, starting from state } i).$$

In other cases, *we will always reach state* j *eventually, starting from state* i*.*

In that case, T_{ij} *can not take the value* ∞*:*

$$\mathbb{P}(T_{ij} = \infty) = 0 \quad \textit{if we are CERTAIN to reach state } j\textit{, starting from state } i.$$

Definition: A random variable T is **defective**, or **improper**, if *it can take the value* ∞*. That is,*

$$T \textit{ is defective if } \mathbb{P}(T = \infty) > 0.$$

Thinking of $\sum_{t=0}^{\infty} \mathbb{P}(T = t)$ as $1 - \mathbb{P}(T = \infty)$

Although it seems strange, when we write $\sum_{t=0}^{\infty} \mathbb{P}(T = t)$, *we are not including the value $t = \infty$.*

The sum $\sum_{t=0}^{\infty}$ continues without ever stopping: at no point can we say we have 'finished' all the finite values of t so we will now add on $t = \infty$. We simply *never get to $t = \infty$ when we take $\sum_{t=0}^{\infty}$.*

For a defective random variable T, this means that

$$\sum_{t=0}^{\infty} \mathbb{P}(T = t) < 1,$$

because we are missing the positive value of $\mathbb{P}(T = \infty)$.

All probabilities of T must still sum to 1, so we have

$$1 = \sum_{t=0}^{\infty} \mathbb{P}(T = t) + \mathbb{P}(T = \infty),$$

in other words

$$\sum_{t=0}^{\infty} \mathbb{P}(T = t) = 1 - \mathbb{P}(T = \infty).$$

PGFs for defective random variables

When T is defective, the PGF of T is *defined as* the power series

$$H(s) = \sum_{t=0}^{\infty} \mathbb{P}(T = t)s^t \quad \text{for } |s| < 1.$$

The term for $\mathbb{P}(T = \infty)s^{\infty}$ is missed out. The PGF is defined as the generating function of the probabilities for finite values only.

Because $H(s)$ is a power series satisfying the conditions of Abel's Theorem, we know that:

- $H(s)$ is left-continuous at $s = 1$, i.e. $\lim_{s \uparrow 1} H(s) = H(1)$.

This is different from the behaviour of $\mathbb{E}(s^T)$, if T is defective:

- $\mathbb{E}(s^T) = H(s)$ for $|s| < 1$ because the missing term is zero: i.e. because $s^\infty = 0$ when $|s| < 1$.
- $\mathbb{E}(s^T)$ is NOT left-continuous at $s = 1$. There is a sudden leap (discontinuity) at $s = 1$ because $s^\infty = 0$ as $s \uparrow 1$, but $s^\infty = 1$ when $s = 1$.

Thus $H(s)$ does NOT represent $\mathbb{E}(s^T)$ at $s = 1$. It is as if $H(s)$ is a 'train' that $\mathbb{E}(s^T)$ rides on between $-1 < s < 1$. At $s = 1$, the train keeps going (i.e. $H(s)$ is continuous) but $\mathbb{E}(s^T)$ jumps off the train.

We test whether T is defective by testing whether or not $\mathbb{E}(s^T)$ 'jumps off the train' — that is, we test whether or not $H(s)$ is equal to $\mathbb{E}(s^T)$ when $s = 1$.

We ***know*** what $\mathbb{E}(s^T)$ is when $s = 1$:

- $\mathbb{E}(s^T)$ is always 1 when $s = 1$, whether T is defective or not:
$$\mathbb{E}(1^T) = 1 \qquad \text{for ANY random variable } T.$$

But the function $H(s) = \sum_{t=0}^{\infty} s^t \mathbb{P}(T = t)$ may or may not be 1 when $s = 1$:

- If T is defective, $H(s)$ is missing a term and $H(1) < 1$.
- If T is not defective, $H(s)$ is not missing anything so $H(1) = 1$.

Test for defectiveness:

> Let $H(s) = \sum_{t=0}^{\infty} s^t \mathbb{P}(T = t)$ be the power series representing the PGF of T for $|s| < 1$. Then T is defective if and only if $H(1) < 1$.

Using defectiveness to find the probability we never get there

The simple test for defectiveness tells us whether there is a positive probability that we NEVER reach our goal. Here are the steps.

1. We want to know the probability that we will NEVER reach state j, starting from state i.
2. Define T to be the random variable giving the *number of steps taken* to get from state i to state j.
3. The event that we never reach state j, starting from state i, is the same as the event that $T = \infty$. (If we wait an infinite length of time, we never get there.) So

$$\mathbb{P}(\textit{never reach state } j \mid \textit{start at state } i) = \mathbb{P}(T = \infty).$$

4. Find $H(s) = \sum_{t=0}^{\infty} s^t \mathbb{P}(T = t)$, using a calculation like the one we did in Section 7.9. $H(s)$ is the PGF of T for $|s| < 1$. We only need to find it for $|s| < 1$. The calculation in Section 7.9 only works for $|s| \leq 1$ because the expectations are infinite or undefined when $|s| > 1$.
5. The random variable T is defective if and only if $H(1) < 1$.
6. If $H(1) < 1$, then the probability that T takes the value ∞ is the missing piece: $\mathbb{P}(T = \infty) = 1 - H(1)$.

Overall:

$$\boxed{\mathbb{P}(\textit{ never reach state } j \mid \textit{start at state } i) = \mathbb{P}(T = \infty) = 1 - H(1).}$$

Expectation and variance of a defective random variable

If T is defective, there is a positive chance that $T = \infty$. This means that $\mathbb{E}(T) = \infty$, *Var*$(T) = \infty$, *and* $\mathbb{E}(T^a) = \infty$ *for any power* a.

$\mathbb{E}(T)$ and $\text{Var}(T)$ can not be found using the PGF when T is defective: you will get the wrong answer.

When you are asked to find $\mathbb{E}(T)$ in a context where T might be defective:

- First check whether T is defective: *is* $H(1) < 1$ *or* $= 1$*?*
- If T is defective, then $\mathbb{E}(T) = \infty$.
- If T is not defective ($H(1) = 1$), then $\mathbb{E}(T) = H'(1)$ *as usual.*

7.11 Random Walk: the probability we never reach our goal

In the random walk in Section 7.9, we defined the first reaching time T_{01} as the number of steps taken to get from state 0 to state 1.

In Section 7.9 we found the PGF of T_{01} to be:

$$\textit{PGF of } T_{01} = H(s) = \frac{1 - \sqrt{1 - s^2}}{s} \quad \textit{for } |s| < 1.$$

Questions:

a) What is the probability that we *never* reach state 1, starting from state 0?

b) What is expected number of steps to reach state 1, starting from state 0?

Solutions:

a) We need to know whether T_{01} *is defective.*

T_{01} *is defective if and only if* $H(1) < 1$.

Now $H(1) = \frac{1-\sqrt{1-1^2}}{1} = 1$. *So* T_{01} *is not defective.*

Thus

$$\mathbb{P}(\textit{never reach state 1} \mid \textit{ start from state 0}) = 0.$$

We will DEFINITELY reach state 1 eventually, even if it takes a very long time.

b) Because T_{01} *is not defective, we can find* $\mathbb{E}(T_{01})$ *by differentiating the PGF:* $\mathbb{E}(T_{01}) = H'(1)$*.*

$$H(s) = \frac{1-\sqrt{1-s^2}}{s} = s^{-1} - \left(s^{-2}-1\right)^{1/2}$$

$$\text{So} \quad H'(s) = -s^{-2} - \frac{1}{2}\left(s^{-2}-1\right)^{-1/2}\left(-2s^{-3}\right)$$

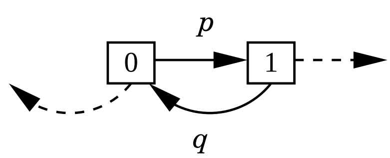
Thus

$$\mathbb{E}(T_{01}) = \lim_{s\uparrow 1} H'(s) = \lim_{s\uparrow 1}\left(-\frac{1}{s^2} + \frac{1}{s^3\sqrt{\frac{1}{s^2}-1}}\right) = \infty.$$

So the expected number of steps to reach state 1 starting from state 0 is infinite: $\mathbb{E}(T_{01}) = \infty$.

This result is striking. Even though we will ***definitely*** reach state 1, the expected time to do so is infinite! In general, we can prove the following results for random walks, starting from state 0:

Property	Reach state 1?	$\mathbb{P}(T_{01} = \infty)$	$\mathbb{E}(T_{01})$
$p > q$	Guaranteed	0	finite
$p = q = \frac{1}{2}$	Guaranteed	0	∞
$p < q$	Not guaranteed	> 0	∞



Note: **(Non-examinable)** If T is defective in the random walk, $\mathbb{E}(s^T)$ is not continuous at $s = 1$. In Section 7.9 we had to solve a quadratic equation to find $H(s) = \mathbb{E}(s^T)$. The negative root solution for $H(s)$ generally represents $\mathbb{E}(s^T)$ for $s < 1$. At $s = 1$, the solution for $\mathbb{E}(s^T)$ suddenly flips from the $-$ root to the $+$ root of the quadratic. This explains how $\mathbb{E}(s^T)$ can be discontinuous as $s \uparrow 1$, even though the negative root for $H(s)$ is continuous as $s \uparrow 1$ and all the working of Section 7.9 still applies for $s = 1$. The reason is that we suddenly switch from the $-$ root to the $+$ root at $s = 1$.

When $|s| > 1$, the conditional expectations are not finite so the working of Section 7.9 no longer applies.