Chapter 4: Generating Functions

This chapter looks at Probability Generating Functions (PGFs) for ***discrete*** random variables. Probability generating functions are useful tools for dealing with sums and limits of random variables.

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.

4.1 Common sums

1. Geometric Series

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

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 $\binom{n}{x} = \dfrac{n!}{(n-x)!\,x!}$ *(^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

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

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

$\mathbb{P}(X=x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$ *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: $e^{\lambda} = \lim_{n \to \infty} \left(1 + \dfrac{\lambda}{n}\right)^n$ *for* $\lambda \in \mathbb{R}$.

4.2 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 2.6.

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

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

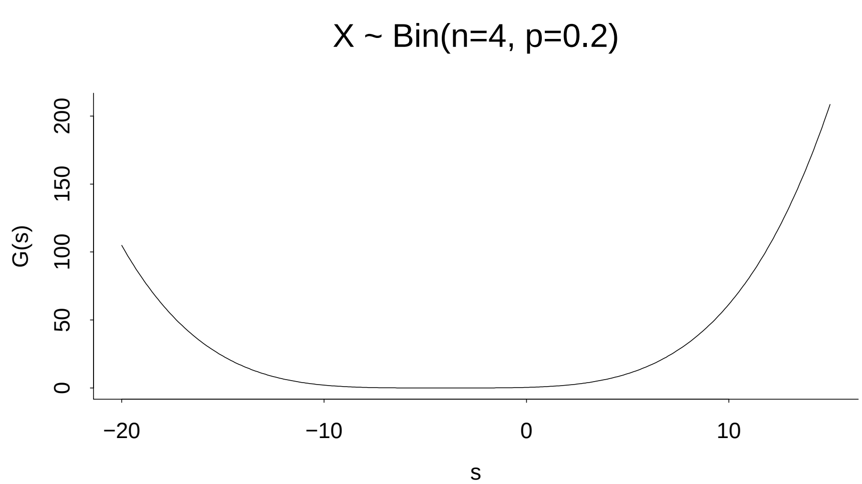
2. $G_X(1) = 1$: $$G_X(1) = \sum_{x=0}^{\infty} 1^x \mathbb{P}(X = x) = \sum_{x=0}^{\infty} \mathbb{P}(X = x) = 1.$$

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 $G_X(1)$:

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

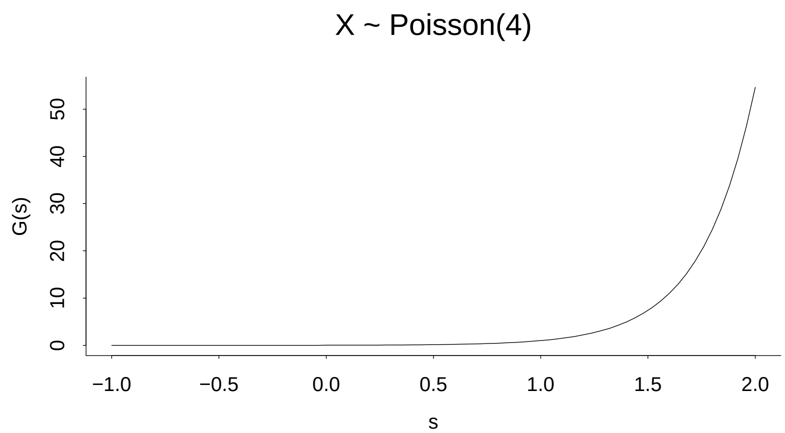
Example 2: Poisson Distribution

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

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

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

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

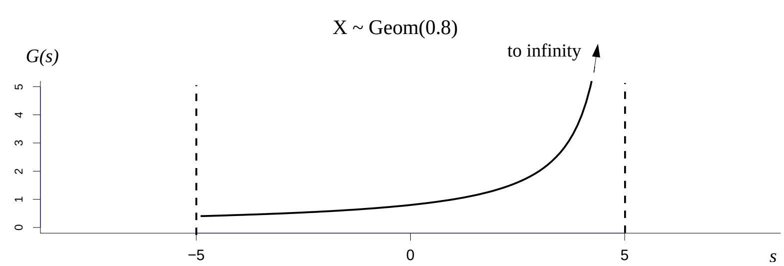


Example 3: Geometric Distribution

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

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

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



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

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

4.3 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

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

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

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

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

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

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

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

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 $G_X(s) = \frac{s}{5}(2 + 3s^2)$. 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 $p_n = \mathbb{P}(X = n) = \left(\frac{1}{n!}\right) G_X^{(n)}(0)$ 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.*

4.4 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 4.4: Let X be a discrete random variable with PGF $G_X(s)$. Then:

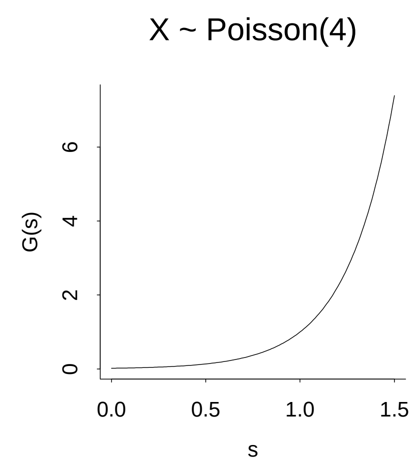
1. $\mathbb{E}(X) = G'_X(1)$.

2. $\mathbb{E}\Big\{X(X-1)(X-2)\dots(X-k+1)\Big\} = G_X^{(k)}(1) = \left.\dfrac{d^k\, G_X(s)}{ds^k}\right|_{s=1}$.

   (This is the kth factorial moment of X.)

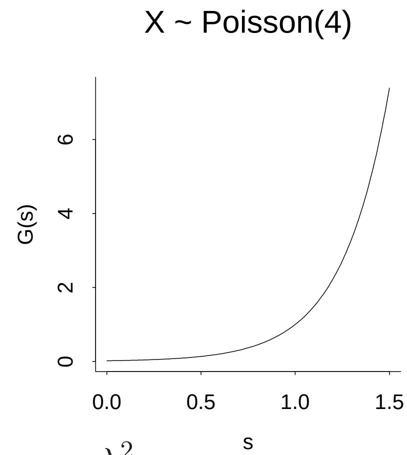
Proof: (Sketch: see Section 4.8 for more details)

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



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

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:

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

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

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

4.5 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 4.5: 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:

$$
\begin{aligned}
G_Y(s) &= \mathbb{E}(s^{(X_1+\ldots+X_n)}) \\
&= \mathbb{E}(s^{X_1}s^{X_2}\ldots s^{X_n}) \\
&= \mathbb{E}(s^{X_1})\mathbb{E}(s^{X_2})\ldots\mathbb{E}(s^{X_n}) \\
&\qquad \textit{(because } X_1,\ldots,X_n \textit{ are independent)} \\
&= \prod_{i=1}^{n} G_{X_i}(s). \quad \textit{as required.} \quad \square
\end{aligned}
$$

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*$(\lambda+\mu)$.

4.6 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 $\mathrm{Var}(T_N) = \sigma^2\,\mathbb{E}(N) + \mu^2\,\mathrm{Var}(N)$, where $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i)$.

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

Theorem 4.6: 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:

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

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: **Heron goes fishing**

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,

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

So

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

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

$$(T \,|\, N = n) \sim \textit{Binomial}(n, p).$$

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

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

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

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 $(\star)$:

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

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

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

$$= \dots ?$$

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* $\mathbb{P}(T = t \mid N = n)$*;*
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.

4.7 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:	$\mathbb{E}(X) = G'_X(1)$	$\mathbb{E}\Big\{X(X-1)\dots(X-k+1)\Big\} = G_X^{(k)}(1)$
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	

4.8 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 4.4? When we said in Section 4.4 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 important notion of 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 $X \sim$ Geometric$(p = 0.8)$. What is the radius of convergence of $G_X(s)$?

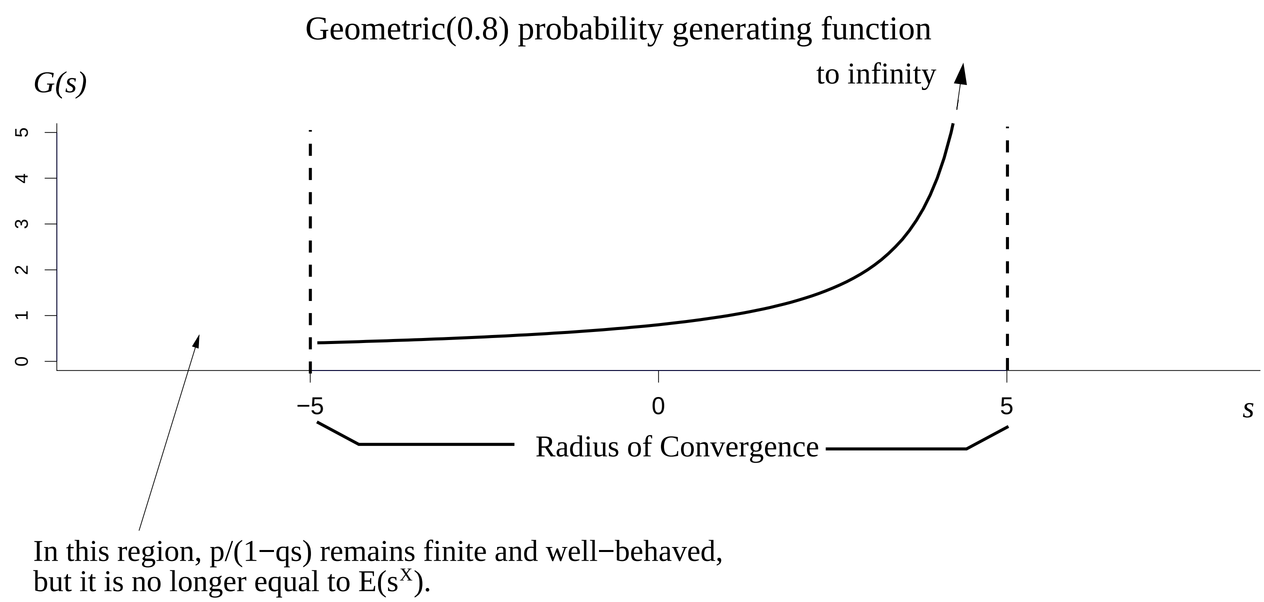
As in Section 4.2,

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

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

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 4.2,

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

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 4.8: Abel's Theorem.

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

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

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

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 4.4. If we assume that term-by-term differentiation is allowed for $G_X(s)$ (see below), then the proof on page 75 gives:

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

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

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).

4.9 Using PGFs for First Reaching Times in the Random Walk

Remember the Drunkard's Walk from Chapter 1: a drunk person staggers to left and right as he walks. This process is also 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.

Our interest is to find the distribution of T_{ij}, 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 demonstrate the procedure with the simplest case of a symmetric random walk on the integers.

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

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

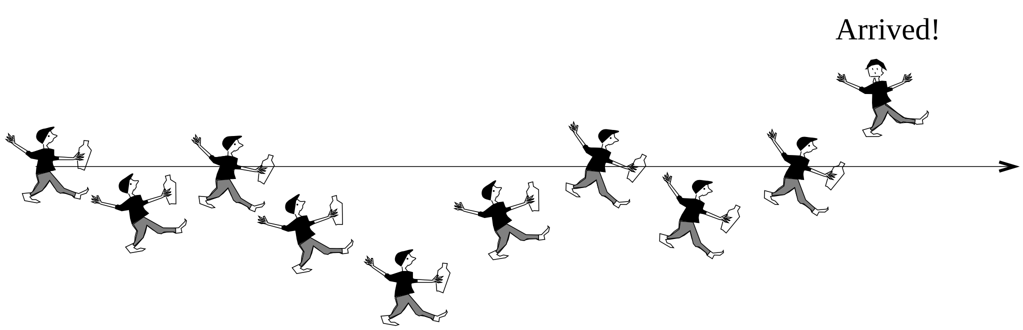
Y_i represents the step to left or right taken by the drunkard at time i.

Let $X_0 = 0$, and $X_t = \sum_{i=1}^{t} Y_i$ for $t > 0$, and consider the stochastic process $\{X_0, X_1, X_2, \ldots\}$.

X_t represents the drunkard's *position at time t.*

Define T_{01} = *number of steps (i.e. time taken) to get from state 0 to state 1.*

What is the PGF of T_{01}, $\mathbb{E}\left(s^{T_{01}}\right)$?



Solution:

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

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

Then

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

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

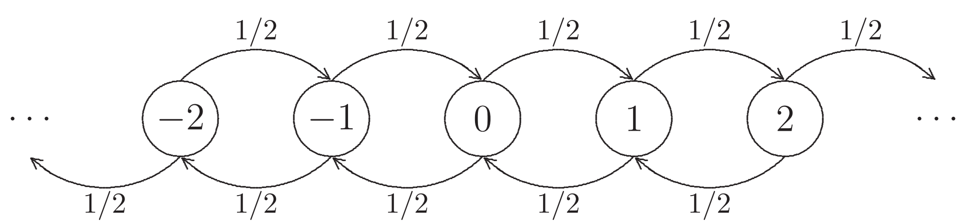
If $Y_1 = -1$*, then* $T_{01} = 1 + T_{-1,1}$*:*

$\rightarrow$ *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: see later). Also, they have the same distribution because the process is translation invariant:*



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:*

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

4.10 Defective, or improper, random variables

Defective random variables often arise in stochastic processes. They are random variables that *may take the value* ∞*.*

For example, consider the first reaching time T_{ij} from the random walk in Section 4.9. There might be a chance that *we never reach state* j*, starting from state* i*.*

If so, then T_{ij} *can take the value* ∞*.*

Alternatively, the process might guarantee that *we will always reach state* j *eventually, starting from state* i*.*

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

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

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

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*

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

This means that:

- the term for $\mathbb{P}(T = \infty)s^{\infty}$ has been missed out;
- $G_T(s)$ is *only* equal to $\mathbb{E}(s^T)$ for $|s| < 1$, because $s^{\infty} = 0$ when $|s| < 1$ so the missing term doesn't matter;
- for $|s| > 1$, $\mathbb{E}(s^T)$ is infinite or undefined because s^{∞} is infinite or undefined. The function $G_T(s)$ might still exist and be well-behaved for $|s| > 1$, but it is no longer equal to $\mathbb{E}(s^T)$.

- By Abel's Theorem, the function $G_T(s)$ is continuous at $s = 1$, although it no longer equals $\mathbb{E}(s^T)$ when $s = 1$. (Note that the function $\mathbb{E}(s^T)$ is *not* continuous at $s = 1$.) In fact,

$$\lim_{s\uparrow 1} G_T(s) = G_T(1) = \sum_{t=0}^{\infty} \mathbb{P}(T = t) = 1 - \mathbb{P}(T = \infty),$$

  from above.

This last fact gives us our **test for defectiveness** in a random variable.

> The random variable T is defective if and only if $G_T(1) < 1$.

Summary of defective random variables:

The concept of defective random variables might seem hard to understand. However, you only have to remember the following points.

1. The random variable T is defective if and only if $G_T(1) < 1$.
2. If $G_T(1) < 1$, then the probability that T takes the value ∞ is $\mathbb{P}(T = \infty) = 1 - G_T(1)$.
3. When you are asked to find $\mathbb{E}(T)$ in a context where T might be defective (e.g. $T = T_{ij}$ is a first reaching time in a random walk), you must do the following:
   - First check whether T is defective: *is* $G_T(1) < 1$ *or* $= 1$*?*
   - If T is *not* defective ($G_T(1) = 1$), then $\mathbb{E}(T) = G'_T(1)$ *as usual.*
   - If T *is* defective ($G_T(1) < 1$), then 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.

Example of defective random variables: finding $\mathbb{E}(T_{01})$ in the random walk

Recall the random walk example in Section 4.9. We defined the first reaching time, T_{01}, as the number of steps taken to get from state 0 to state 1.

In Section 4.9 we found the PGF of T_{01}, which contains all information about the distribution of T_{01}:

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

We now wish to use the PGF to find $\mathbb{E}(T_{01})$. We must bear in mind that T_{01} *is a reaching time, so it might be defective: there might be a possibility that we* *never* *reach state 1, starting from state 0.*

Questions:

a) What is $\mathbb{E}(T_{01})$?

b) What happens if we try to use the Law of Total Expectation instead of the PGF to find $\mathbb{E}(T_{01})$?

a) First check if T_{01} is defective.

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

But $\lim_{s\uparrow 1} H(s) = H(1) = \frac{1-\sqrt{1-1^2}}{1} = 1$. *So* T_{01} *is not defective, and we can proceed with finding* $\mathbb{E}(T_{01}) = \lim_{s\to 1} H'(s)$.

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

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

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 (although $\mathbb{P}(T_{01} = \infty) = 0$*), the expectation is infinite:* $\mathbb{E}(T_{01}) = \infty$.

b) Using the law of total expectation, conditioning on the outcome of the first step Y_1*:*

$$
\begin{aligned}
\mathbb{E}(T_{01}) &= \mathbb{E}\left\{\mathbb{E}(T_{01} \,|\, Y_1)\right\} \\
&= \mathbb{E}(T_{01} \,|\, Y_1 = 1)\mathbb{P}(Y_1 = 1) + \mathbb{E}(T_{01} \,|\, Y_1 = -1)\mathbb{P}(Y_1 = -1) \\
&= 1 \times \tfrac{1}{2} + \{1 + \mathbb{E}(T_{-1,1})\} \times \tfrac{1}{2} \\
&= 1 + \tfrac{1}{2}\mathbb{E}(T_{01} + T_{01}) \\
&= 1 + \tfrac{1}{2} \times 2\mathbb{E}(T_{01}) \\
\Rightarrow \quad \mathbb{E}(T_{01}) &= 1 + \mathbb{E}(T_{01}). \qquad \textit{So 0 = 1 ???}
\end{aligned}
$$

What's wrong? The expectations were not finite, so the law of total expectation was invalid. We can deduce from $\mathbb{E}(T_{01}) = 1 + \mathbb{E}(T_{01})$ *that the expectation is infinite.*

Note: **(Non-examinable)** If T_{ij} is defective in the random walk, $\mathbb{E}(s^{T_{ij}})$ is not continuous at $s = 1$. In Section 4.9 we saw that we had to solve a quadratic equation to find $H(s) = \mathbb{E}(s^{T_{ij}})$. The negative root solution for $H(s)$ generally represents $\mathbb{E}(s^{T_{ij}})$ for $s < 1$. At $s = 1$, the required solution for $\mathbb{E}(s^{T_{ij}})$ suddenly flips from the $-$ root to the $+$ root of the quadratic. This explains how $\mathbb{E}(s^{T_{ij}})$ 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 4.9 still applies for $s = 1$. The explanation 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 4.9 no longer applies.
