Chapter 14

Generating functions

Throughout the course I have been emphasizing the idea that discrete probability distributions are specified by the list of possible values and the probabilities attached to those values (and that continuous distributions are specified by density functions). Probability distributions can also be specified by a variety of transforms, that is, by functions that somehow encode the properties of the distributions into a form more convenient for certain kinds of probability calculation. This Chapter will describe a technique that is particularly useful for discrete distributions concentrated on the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ of nonnegative integers.

Definition. *Suppose X is a random variable that takes only nonnegative integer values, with probabilities $p_k = \mathbb{P}\{X = k\}$ for $k \in \mathbb{N}_0$. The* **probability generating function** *$g(\cdot)$ is defined as*

$$g(s) = \mathbb{E}s^X = \sum\nolimits_{k=0}^{\infty} p_k s^k \qquad \text{for } 0 \le s \le 1.$$

The powers of the dummy variable s serve as placeholders for the p_k probabilities that determine the distribution; we could recover the p_k's as the coefficients in a power series expansion of the probability generating function. For example, if an $\mathbb{N}_0$-valued random variable X has probability generating function

$$g(s) = \exp\left(\lambda(s-1)\right) \qquad \text{for } 0 \le s \le 1,$$

with λ a positive constant, then X has a Poisson(λ) distribution, because the coefficient of s^k in the power series expansion

$$\exp\left(\lambda(s-1)\right) = e^{-\lambda} \sum\nolimits_{k=0}^{\infty} \frac{(\lambda s)^k}{k!}$$

equals $e^{-\lambda}\lambda^k/k!$.

Expansion of a probability generating function in a power series is just one way of extracting information about the distribution. Repeated differentiation inside the expectation sign gives

$$g^{(m)}(s) = \frac{\partial^m}{\partial s^m}\mathbb{E}(s^X) = \mathbb{E}\left(X(X-1)\ldots(X-m+1)s^{X-m}\right),$$

whence

$$g^{(m)}(1) = \mathbb{E}\left(X(X-1)\ldots(X-m+1)\right) \qquad \text{for } m = 1, 2, \ldots$$

In particular, we have $\mathbb{E}X = g'(1)$ and $\mathbb{E}(X^2 - X) = g''(1)$. It follows that $\mathrm{var}(X) = g''(1) + g'(1) - g'(1)^2$. With a little more algebra we could recover the higher moments of X.

Example <14.1>: Some properties of the negative binomial distribution, derived using probability generating functions.

The next two Examples show how probability generating functions can be used to solve problems involving the stochastic model called a **branching process**. I start with a warm-up exercise to show how the method works.

Example <14.2>: A careful study of the reproductive behaviour of the royal house of Oz has revealed that each member of the family has probability:

$\frac{1}{6}$ of producing no children;
$\frac{3}{6}$ of producing only one child;
$\frac{2}{6}$ of producing exactly two children.

The present king, Osgood, is only 8 years old. Assuming that family members reproduce independently of each other, according to the stated distribution, find the probability that Osgood eventually has exactly two grandchildren.

Naturally Osgood would like the House of Oz to survive forever. The prospects might appear good, because each member of the family has expected number of offspring equal to $(0\times 1/6)+(1\times 3/6)+(2\times 2/6) = 7/6$. On the average, each generation size should be about $7/6$ times the size of the previous generation size. But averages don't tell the whole story, as the next Example shows.

Example <14.3>: Extinction probability for a branching process: probability 1/2 that the Osgood line dies out eventually.

In general, suppose there is a probability p_k, for $k \in \mathbb{N}_0$, that an individual has k offspring. Write $g(s)$ for the corresponding probability generating function and $\mu = \sum_{k=0}^{\infty} kp_k$ for the mean. Let X_i denote the size of the ith generation, starting from $X_0 = 1$. Assume independence of reproduction. The conditional distribution of X_n given $X_{n-1} = k$ is like that of a sum $\xi_1 + \ldots + \xi_k$, where the ξ_i are independent random variables each with $\mathbb{P}\{\xi_i = k\} = p_k$ for $k \in \mathbb{N}_0$. Thus

$$\begin{aligned} g_n(s) = \mathbb{E}s^{X_n} &= \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\}\mathbb{E}\left(s^{X_n} \mid X_{n-1} = k\right) \\ &= \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\} g(s)^k = g_{n-1}(g(s)) \end{aligned}$$

and

$$\mathbb{E}X_n = \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\}\mathbb{E}\left(X_n \mid X_{n-1} = k\right) = \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\}(k\mu) = \mu\mathbb{E}X_{n-1}.$$

The expression for g_n leads to the conclusion that $\mathbb{P}\{$ process dies out eventually $\} = \theta$, where θ is the smaller solution of the equation $\theta = g(\theta)$. If $\mu \le 1$ then $\theta = 1$.

By repeated substitution, we get $\mathbb{E}X_n = \mu^n$. If $\mu < 1$ then the expected number of offspring (including the founder) equals

$$\mathbb{E}\left(\sum_{n=0}^{\infty} X_n\right) = \sum_{n=0}^{\infty} \mathbb{E}X_n = \sum_{n=0}^{\infty} \mu^n = 1/(1-\mu).$$

Note that the expected number of offspring is infinite when $\mu = 1$, even though the process dies out eventually with probability one.

Branching processes show up in unexpected places.

Example <14.4>: Behavior of a queue.

Examples for Chapter 14

<14.1> **Example.** Let X have a negative binomial distribution (as defined on Homework 4),

$$\mathbb{P}\{X = k\} = \binom{-\alpha}{k} p^\alpha (-q)^k \qquad \text{for } k \in \mathbb{N}_0, \qquad \text{where } q = 1-p.$$

It has probability generating function

$$g(s) = \mathbb{E}s^X = \sum\nolimits_{k=0}^{\infty} \binom{-\alpha}{k} p^\alpha (-qs)^k = p^\alpha (1-qs)^{-\alpha} \qquad \text{for } 0 \le s \le 1,$$

with derivatives

$$g'(s) = p^\alpha \alpha q (1-qs)^{-\alpha-1} \qquad \text{and} \qquad g''(s) = p^\alpha \alpha(\alpha+1) q^2 (1-qs)^{-\alpha-2}.$$

Thus $\mathbb{E}X = g'(1) = \alpha q/p$ and

$$\text{var}(X) = g''(1) + \mathbb{E}X - (\mathbb{E}X)^2 = \alpha q/p^2.$$

As a check, note that $1 + X$ has a geometric(p) distribution if $\alpha = 1$. The expectation $\mathbb{E}(1+X) = 1 + q/p = 1/p$ agrees with the calculation from Chapter 2.

On Homework 10, you were asked to find the marginal distribution of a random variable X when $X \mid L = \lambda \sim$ Poisson(λ) and $L/\theta \sim$ gamma(α) for some fixed, positive constant θ. You could have solved this Problem by an application of the conditioning formula from Chapter 10. For $k \in \mathbb{N}_0$,

$$\begin{aligned}
\mathbb{P}\{X = k\} &= \int_0^\infty \mathbb{P}\{X = k \mid L/\theta = t\} f(t)\, dt \qquad \text{with } f = \text{gamma}(\alpha) \text{ density} \\
&= \int_0^\infty \frac{e^{-\theta t}(\theta t)^k}{k!} \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)}\, dt = \frac{\theta^k}{k!\Gamma(\alpha)} \int_0^\infty t^{k+\alpha-1} e^{-t(1+\theta)}\, dt \\
&= \frac{\theta^k}{(1+\theta)^{k+\alpha} k! \Gamma(\alpha)} \int_0^\infty y^{k+\alpha-1} e^{-y}\, dy \qquad \text{change: } y = t(1+\theta) \\
&= \left(\frac{\theta}{1+\theta}\right)^k \left(\frac{1}{1+\theta}\right)^\alpha \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)}.
\end{aligned}$$

The last ratio equals

$$\frac{(\alpha+k-1)(\alpha+k-2)\dots(\alpha)}{k!} = \binom{-\alpha}{k}(-1)^k$$

If we choose $\theta = q/p$ then X has the negative binomial distribution with parameters p and α, as defined at the start of the Example.

We can also derive the result by using probability generating functions.

$$\begin{aligned}
\mathbb{E}s^X &= \int_0^\infty \mathbb{E}\left(s^X \mid L/\theta = t\right) \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}\, dt \\
&= \int_0^\infty e^{\theta t(s-1)} \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}\, dt \qquad \text{cf. pgf for Poisson}(\theta t) \\
&= \big(1 + \theta(1-s)\big)^{-\alpha} \int_0^\infty \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\, dy \qquad \text{change: } y = t + \theta t(1-s) \\
&= \left(\frac{1/(1+\theta)}{1 - \theta s/(1+\theta)}\right)^\alpha.
\end{aligned}$$

Putting $p = 1/(1+\theta)$ we get the probability generating function of the negative binomial distribution. A power series expansion (not really necessary) would recover the negative binomial probabilities as coefficients. □

<14.2> **Example.** Write X_n for the size of the nth generation, starting from $X_0 = 1$ for Osgood himself. The question asks for $\mathbb{P}\{X_2 = 2\}$.

The problem is simple enough to yield to straightforward conditioning on X_1, the number of children that Osgood will produce. Clearly

$$\mathbb{P}\{X_2 = 2 \mid X_1 = 0\} = 0 \qquad \text{and} \qquad \mathbb{P}\{X_2 = 2 \mid X_1 = 1\} = \tfrac{2}{6}$$

If $X_1 = 2$ then each of the two children will reproduce according to the stated offspring distribution, so that X_2 can be written as a sum of two (conditionally) independent random variables ξ_1 and ξ_2 with

$$\mathbb{P}\{\xi_i = 0 \mid X_1 = 2\} = \tfrac{1}{6}, \quad \mathbb{P}\{\xi_i = 1 \mid X_1 = 2\} = \tfrac{3}{6}, \quad \mathbb{P}\{\xi_i = 2 \mid X_1 = 2\} = \tfrac{2}{6}.$$

Arguing conditonally, with $X_2 = \xi_1 + \xi_2$, we get

$$\begin{aligned}
&\mathbb{P}\{X_2 = 2 \mid X_1 = 2\} = \mathbb{P}\{\xi_1 + \xi_2 = 2 \mid X_1 = 2\} \\
&\quad = \mathbb{P}\{\xi_1 = 0, \xi_2 = 2 \mid X_1 = 2\} + \mathbb{P}\{\xi_1 = 1, \xi_2 = 1 \mid X_1 = 2\} + \mathbb{P}\{\xi_1 = 2, \xi_2 = 0 \mid X_1 = 2\} \\
&\quad = \left(\tfrac{1}{6} \times \tfrac{2}{6}\right) + \left(\tfrac{3}{6} \times \tfrac{3}{6}\right) + \left(\tfrac{2}{6} \times \tfrac{1}{6}\right) \qquad \text{by conditional independence} \\
&\quad = \frac{13}{36}.
\end{aligned}$$

Average out over the X_1 distribution.

$$\begin{aligned}
\mathbb{P}\{X_2 = 2\} &= \mathbb{P}\{X_2 = 2 \mid X_1 = 0\}\tfrac{1}{6} + \mathbb{P}\{X_2 = 2 \mid X_1 = 1\}\tfrac{3}{6} + \mathbb{P}\{X_2 = 2 \mid X_1 = 2\}\tfrac{2}{6} \\
&= 0 + \left(\tfrac{2}{6} \times \tfrac{3}{6}\right) + \left(\tfrac{13}{36} \times \tfrac{2}{6}\right) \\
&= \tfrac{31}{108}.
\end{aligned}$$

Not so hard.

You would have a lot more work to do—mainly bookkeeping—if I asked for the probability of exactly 7 great-great-great-great-grandchildren. It would be hard to keep track of all the possible ways of getting $X_6 = 7$. For such a task, generating functions come in handy.

Define $g_n(s) = \mathbb{E}s^{X_n}$ for $0 \le s \le 1$. For fixed s, calculate the expected value of a function of a random variable in the usual way:

$$g_1(s) = s^0\mathbb{P}\{X_1 = 0\} + s^1\mathbb{P}\{X_1 = 1\} + s^2\mathbb{P}\{X_1 = 2\} = \tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2.$$

Similarly,

$$g_2(s) = s^0\mathbb{P}\{X_2 = 0\} + s^1\mathbb{P}\{X_2 = 1\} + s^2\mathbb{P}\{X_2 = 1\} + s^3\mathbb{P}\{X_2 = 3\} + s^4\mathbb{P}\{X_2 = 4\} = \quad ?$$

It might appear that calculation of $g_2(s)$ involves five times the sort of work required for the calculation of $\mathbb{P}\{X_2 = 2\}$. Not so.

Condition once more on the value of X_1.

$$\begin{aligned}
&\mathbb{E}(s^{X_2} \mid X_1 = 0) = s^0 = 1 = g_1(s)^0 \qquad \text{because } \mathbb{P}\{X_2 = 0 \mid X_1 = 0\} = 1 \\
&\mathbb{E}(s^{X_2} \mid X_1 = 1) = g_1(s) \qquad \text{offspring distribution for one king.}
\end{aligned}$$

Conditional independence of the offspring from each child when $X_1 = 2$ justifies a factorization:

$$\mathbb{E}(s^{X_2} \mid X_1 = 2) = \mathbb{E}(s^{\xi_1}s^{\xi_2} \mid X_1 = 2) = g_1(s)^2.$$

In short,

$$\mathbb{E}(s^{X_2} \mid X_1 = k) = g_1(s)^k \qquad \text{for } k = 0, 1, 2.$$

Average out over the X_1 distribution.

$$\begin{aligned}
g_2(s) = \mathbb{E}s^{X_2} &= \mathbb{E}\left(s^{X_2} \mid X_1 = 0\right)\mathbb{P}\{X_1 = 0\} + \mathbb{E}\left(s^{X_2} \mid X_1 = 1\right)\mathbb{P}\{X_1 = 1\} \\
&\qquad + \mathbb{E}\left(s^{X_2} \mid X_1 = 2\right)\mathbb{P}\{X_1 = 2\} \\
&= g_1(s)^0\mathbb{P}\{X_1 = 0\} + g_1(s)\mathbb{P}\{X_1 = 1\} + g_1(s)^2\mathbb{P}\{X_1 = 2\} \\
&= g_1(g_1(s)) \qquad \text{cf. } g_1(t) \text{ where } t = g_1(s) \\
&= \tfrac{1}{6} + \tfrac{3}{6}g_1(s) + \tfrac{2}{6}g_1(s)^2 \\
&= \tfrac{1}{6} + \tfrac{3}{6}\left(\tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2\right) + \tfrac{2}{6}\left(\tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2\right)^2 \\
&= \left(\tfrac{1}{6} + \tfrac{3}{6} \times \tfrac{1}{6} + \tfrac{2}{6} \times \tfrac{1}{36}\right) + \left(\tfrac{3}{6} \times \tfrac{3}{6} + \tfrac{2}{6} \times \tfrac{6}{36}\right)s \\
&\quad + \left(\tfrac{3}{6} \times \tfrac{2}{6} + \tfrac{2}{6} \times \tfrac{13}{36}\right)s^2 + \left(\tfrac{2}{6} \times \tfrac{12}{36}\right)s^3 + \left(\tfrac{2}{6} \times \tfrac{4}{36}\right)s^4.
\end{aligned}$$

The probability $\mathbb{P}\{X_2 = 2\}$ equals $31/108$, the coefficient of s^2. Not only is the answer the same as before, but also the numerical expression leading to that value is exactly the same as in the direct calculation of $\mathbb{P}\{X_2 = 2\}$. The powers of s have served merely as placeholders around which the algebra has been organized; the powers of s tag the various products of probabilities that go into the sums for calculating each $\mathbb{P}\{X_2 = k\}$ by conditioning. □

The virtue of the generating function as a bookkeeping device becomes clearer if we follow the later generations of the House of Oz. You should check that

$$E(s^{X_n} \mid X_{n-1} = k) = g_1(s)^k,$$

by writing X_n as a sum of k conditionally independent random variables $\xi_1, \dots, \xi_k$ when $X_{n-1} = k$. Averaging out over the X_{n-1} distribution, you would then get

$$\mathbb{E}s^{X_n} = 1 + g_1(s)\mathbb{P}\{X_{n-1} = 1\} + g_1(s)^2\mathbb{P}\{X_{n-1} = 2\} + \dots = g_{n-1}(g_1(s))$$

The same argument repeated $n-2$ more times would then give

$$g_n(s) = g_1(g_1(g_1(\dots g_1(s)))\dots),$$

an n-fold composition of functions.

Your algebraic abilities might be up to multiplying out polynomials of polynomials, but mine aren't. Luckily, there are computer packages, such as Mathematica, that make short work of such algebra. For $g_6(s)$, the Mathematica code

```
g[s_] := 1/6 + 3/6 * s + 2/6 * s * s
gn[n_] := Expand[Nest[g, s, n]]
gg[n_] := N[gn[n], 3]
```

gave me the polynomial[1]

$$0.412 + 0.0824s + 0.107s^2 + 0.0934s^3 + 0.0808s^4 + 0.0624s^5 + 0.0483s^6 + 0.0354s^7 +$$
$$0.0254s^8 + 0.0178s^9 + 0.0122s^{10} + 0.00819s^{11} + 0.00539s^{12} + 0.00348s^{13} + 0.00221s^{14} +$$
$$0.00137s^{15} + 0.000838s^{16} + 0.000503s^{17} + 0.000296s^{18} + 0.000171s^{19} + 0.0000969s^{20} +$$
$$0.0000539s^{21} + 0.0000294s^{22} + 0.0000157s^{23} + \qquad \text{(Are you still checking?)}$$
$$8.22 \times 10^{-6}s^{24} + 4.21 \times 10^{-6}s^{25} + 2.11 \times 10^{-6}s^{26} + 1.04 \times 10^{-6}s^{27} + 4.99 \times 10^{-7}s^{28} +$$
$$2.34 \times 10^{-7}s^{29} + 1.08 \times 10^{-7}s^{30} + 4.82 \times 10^{-8}s^{31} + 2.11 \times 10^{-8}s^{32} + 9.02 \times 10^{-9}s^{33} + 3.75 \times 10^{-9}s^{34} +$$
$$1.52 \times 10^{-9}s^{35} + 6. \times 10^{-10}s^{36} + 2.3 \times 10^{-10}s^{37} + 8.57 \times 10^{-11}s^{38} + 3.1 \times 10^{-11}s^{39} + 1.09 \times 10^{-11}s^{40} +$$
$$3.68 \times 10^{-12}s^{41} + 1.21 \times 10^{-12}s^{42} + 3.81 \times 10^{-13}s^{43} + 1.16 \times 10^{-13}s^{44} + 3.4 \times 10^{-14}s^{45} + 9.56 \times 10^{-15}s^{46} +$$
$$2.57 \times 10^{-15}s^{47} + 6.61 \times 10^{-16}s^{48} + 1.62 \times 10^{-16}s^{49} + 3.75 \times 10^{-17}s^{50} + 8.22 \times 10^{-18}s^{51} + 1.7 \times 10^{-18}s^{52} +$$
$$3.27 \times 10^{-19}s^{53} + 5.88 \times 10^{-20}s^{54} + 9.75 \times 10^{-21}s^{55} + 1.48 \times 10^{-21}s^{56} + 2.03 \times 10^{-22}s^{57} + 2.49 \times 10^{-23}s^{58} +$$
$$2.66 \times 10^{-24}s^{59} + 2.43 \times 10^{-25}s^{60} + 1.82 \times 10^{-26}s^{61} + 1.05 \times 10^{-27}s^{62} + 4.19 \times 10^{-29}s^{63} + 8.74 \times 10^{-31}s^{64}$$

Just for the record, the probability that Osgood has exactly 7 great-great-great-great-grandchildren is $\mathbb{P}\{X_6 = 7\} =$ coefficient of $s^7 \approx 0.0354$. *You should also notice that* $\mathbb{P}\{X_6 = 0\} \approx 0.412$. *There is a 41% chance that the House of Oz will have died out by the 6th generation. It's tough to keep the family name alive, even if each family member works hard at keeping the birth rate up.*

[1] *The last line rounds the coefficients off to 3 decimal places: I got tired of looking at fractions like* 2317562/89725362 *in the output.*

<14.3> **Example.** (The House of Oz, continued—maybe.) What is the probability that the House never dies out?

With the same notation as in the previous Example, the probability of survival to at least the nth generation is

$$\mathbb{P}\{X_n > 0\} = 1 - \mathbb{P}\{X_n = 0\} = 1 - g_n(0).$$

Write θ_n for $\mathbb{P}\{X_n = 0\}$. As n increases, θ_n increases. Why? It must have a limiting value, which we can denote by θ. Thus $\mathbb{P}\{\text{survive forever}\} = 1 - \theta$. How do we calculate θ?

Notice that

$$\theta_n = g_n(0) = g_1(g_1(g_1(\ldots g_1(0)\ldots))) = g_1(g_{n-1}(0)) = g_1(\theta_{n-1}).$$

As n increases, the θ_n on the left-hand side increases to θ and the θ_{n-1} on the right hand side also increases to θ. In the limit we have $\theta = g_1(\theta)$. That is,

$$\theta = \tfrac{1}{6} + \tfrac{3}{6}\theta + \tfrac{2}{6}\theta^2.$$

The quadratic equation has two roots, $\theta = 1$ and $\theta = 1/2$. Which one is the value we seek?

Here is an argument to show that $\theta = 1/2$ is the root that solves the extinction problem. By direct substitution,

$$\theta_1 = g_1(0) = 1/6 < 1/2.$$

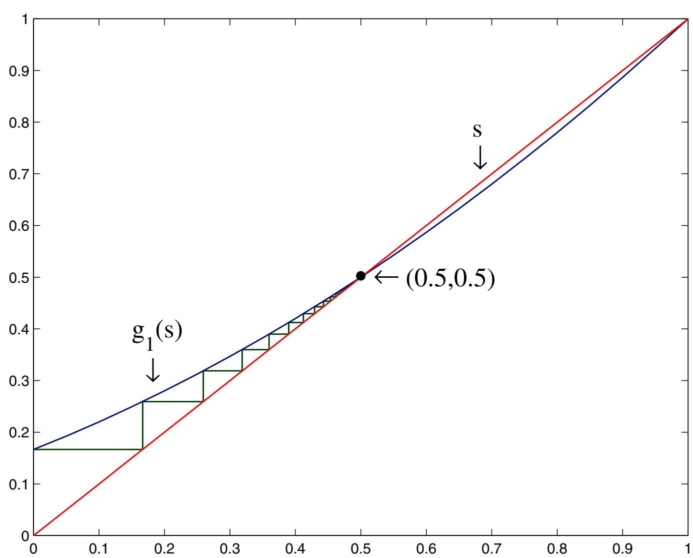
Apply the increasing function $g_1(.)$ to both sides to get

$$\theta_2 = g_1(\theta_1) = g_1(1/6) < g_1(1/2) = 1/2.$$

Apply g_1 again:

$$\theta_3 = g_1(\theta_2) < g_1(1/2) = 1/2.$$

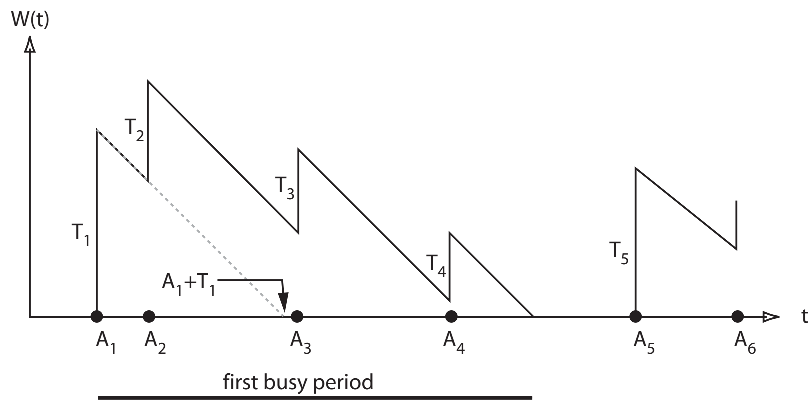
And so on. For every n, we have $\theta_n < 1/2$. The θ_n values cannot increase to 1; they must increase to the other root: $\theta = 1/2$. There is a probability 1/2 that the Osgood line eventually dies out.



Another way to understand the convergence of θ_n to 1/2 is to plot the functions $g_1(s) = 1/6+3s/6+2s^2/6$ and s on the same graph. They cross at 1/2 and 1. The successive values $\theta_1, \theta_2, \ldots$ correspond to a zig-zag path with alternating horizontal and vertical steps, starting from the point $(0, 1/6)$. The path jams itself into the narrow spike between s and $g(s)$; the zig-zag converges to the tip of the spike at $(1/2, 1/2)$. □

<14.4> **Example.** Consider customers waiting in a line (a queue) for the attention of a single server. Suppose the customers arrive according to a Poisson Process with rate λ. Suppose also that the service times for the customers are independent, identically distributed random variables $T_1, T_2, \ldots$ each with expected value τ.

At each time $t \geq 0$ the server would have to work for a time $W(t)$ just to take care of all the customers already in the waiting line. The random variable $W(t)$ is a rather complicated function of the arrival times $A_1, A_2, \ldots$ of customers during $[0, t]$ and their service times. It can be represented by a function that increases by an amount T_i at the ith arrival and decreases at a 45° slope between arrivals. Whenever $W(t) = 0$ the server is not busy.



The times between arrivals are independent, exponentially distributed random variables, each with expected value $1/\lambda$. Roughly speaking, for large n we should have $A_n \approx n/\lambda$. The service time for these n customers should be about $n\tau$. If $n\tau > n/\lambda$ then the server will be in trouble: on average, customers arriving faster than they can be handled. If $n\tau < n/\lambda$, we might expect that the server will eventually fall idle, until the arrival of a new bunch of customers. With the help of a branching process, we can make this intuition more precise.

Think of the first customer as the Osgood who begins a new royal line, with the X_1 customers who arrive in the time interval $(A_1, A_1 + T_1]$ while he is being served as his children. Think of the customers who arrive during a service period of a child as his children, grandchildren of Osgood. And so on.

Notice that $X_1 \mid T_1 = t \sim \text{Poisson}(\lambda t)$. Suppose T_1 has a continuous distribution with density function f. Define $H(r) = \mathbb{E}e^{-rT_1}$ for each $r \geq 0$. Then

$$g_1(s) = \mathbb{E}s^{X_1} = \int_0^\infty \mathbb{E}(s^{X_1} \mid T_1 = t)\, f(t)\, dt = \int_0^\infty e^{\lambda t(s-1)} f(t)\, dt = H\left(\lambda(1-s)\right)$$

and

$$\mathbb{E}X_1 = \int_0^\infty \mathbb{E}(X_1 \mid T_1 = t)\, f(t)\, dt = \int_0^\infty \lambda t f(t)\, dt = \lambda \mathbb{E}T_1 = \lambda\tau.$$

If $\lambda\tau \leq 1$ the branching process dies out with probability one. In that case, the server eventually falls idle. If $\lambda\tau > 1$ there is a nonzero probability that the server never gets a rest. In fact, the probability θ of getting a rest eventually is given by the smaller root of the equation $\theta = H(\lambda - \lambda\theta)$.

Suppose $\lambda\tau < 1$. Let L denote the length (in time) of the first busy period and let B_i denote the event that customer i arrives during that first busy period. Then

$$L = \sum\nolimits_{i=1}^\infty T_i \mathbb{I}_{B_i}$$

The service time T_i is independent of the event B_i. Why? Thus

$$\mathbb{E}L = \sum\nolimits_{i=1}^\infty \mathbb{E}\left(T_i \mathbb{I}_{B_i}\right) = \sum\nolimits_{i=1}^\infty \mathbb{E}\left(T_i\right)\left(\mathbb{E}\mathbb{I}_{B_i}\right) = \tau\mathbb{E}\left(\sum\nolimits_{i=1}^\infty \mathbb{I}_{B_i}\right).$$

The sum $\sum_{i=1}^{\infty} \mathbb{I}_{B_i}$ counts the number of customers served during the first busy period, which is the same as $\sum_{n=0}^{\infty} X_n$, the total number of individuals generated by the branching process. Thus

$$\mathbb{E}L = \tau\mathbb{E}\left(\sum\nolimits_{n=0}^{\infty} X_n\right) = \frac{\tau}{1-\lambda\tau} \qquad \text{when } \lambda\tau < 1.$$

Who would have guessed that within a queue there lurks a branching process? □