ELEMENTARY STOCHASTIC PROCESSES

William J. Anderson
McGill University

Contents

Chapter 1

Some Probability Background

1.1 Review of Conditional Expectation

Definition. Let X and Y be random variables and suppose $f_{Y|X}(y|x)$ is the conditional density (or conditional probability) function of Y given that $X = x$. Then if x is such that $f_{Y|X}(y|x)$ is defined (that is, if $f_X(x) > 0$), we define

$$E(\phi(Y)|X = x) = \begin{cases} \int_{-\infty}^{\infty} \phi(y) f_{Y|X}(y|x)\, dy, & \text{if } Y \text{ is a continuous random variable;} \\ \sum_y \phi(y) f_{Y|X}(y|x), & \text{if } Y \text{ is a discrete random variable;} \end{cases}$$

to be the *conditional expectation of* $\phi(Y)$ *given that* $X = x$. Temporarily let $h(x) = E(\phi(Y)|X = x)$. Then we define $E(\phi(Y)|X) = h(X)$ to be the conditional expectation of $\phi(Y)$ given X.

Notice that one gets $E(\phi(Y)|X)$ by replacing x in the expression for $E(\phi(Y)|X = x)$ by X. Thus whereas $E(\phi(Y)|X = x)$ is a function of the numerical variable x, $E(\phi(Y)|X)$ is a random variable.

We now examine several examples of computing conditional expectations.

Example 1.1. Suppose that X and Y are jointly continuous random variables with joint density function given by:

$$f(x, y) = \begin{cases} x + y, & \text{if } 0 \le x, y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Find (i) $E(Y^2|X = x)$ and (ii) $E(Y^2|X)$.

Solution. The marginal density function of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy = \begin{cases} \int_0^1 x + y\, dy = x + .5, & \text{if } 0 \le x \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Then the conditional density function of Y given $X = x$ is

$$f_{Y|X}(y|x) = \begin{cases} \text{undefined;} & \text{if } f_X(x) = 0, \\ \frac{f(x,y)}{f_X(x)}, & \text{if } f_X(x) > 0; \end{cases}$$
$$= \begin{cases} \text{undefined;} & \text{if } x < 0 \text{ or } x > 1, \\ \frac{x+y}{x+.5}, & \text{if } 0 \le x, y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

If $x < 0$ or $x > 1$, $E(Y^2|X = x)$ is therefore not defined, while if $0 \le x \le 1$, then

$$E(Y^2|X = x) = \int_{-\infty}^{+\infty} y^2 f_{Y|X}(y|x)\, dy = \int_0^1 y^2 \frac{x + y}{x + .5}\, dy = \frac{\frac{x}{3} + .25}{x + .5}.$$

Finally, replacing x by X gives $E(Y^2|X) = \frac{\frac{X}{3}+.25}{X+.5}$ It is interesting to note (in view of theorem 1.1) that we have

$$E\big[E(Y^2|X)\big] = \int_{-\infty}^{+\infty} \frac{\frac{x}{3}+.25}{x+.5} f_X(x)\,dx = \int_0^1 \frac{\frac{x}{3}+.25}{x+.5}\cdot(x+.5)\,dx = \frac{5}{12},$$

while

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y)\,dy = \int_0^1 y^2(y+.5)\,dy = \frac{5}{12},$$

and so we have $E\big[E(Y^2|X)\big] = E(Y^2)$. We shall see in theorem 1.1 that this type of statement is true in general.

Example 1.2. Let X and Y be independent random variables, each having the geometric distribution with parameter p; that is, assume that X and Y have probability function

$$f(x) = pq^{x-1}, \quad x = 1, 2, 3, \ldots$$

where $0 < p < 1$ and $q = 1 - p$. Find (i) $E(Y|X+Y=n)$ and (ii) $E(Y|X+Y)$.

Solution. Let $Z = X + Y$. We will begin by computing $f_{Y|Z}(y|z)$. We have

$$f_{Y|Z}(y|z) = \frac{\Pr\{Y=y, Z=z\}}{\Pr\{Z=z\}} = \frac{\Pr\{Y=y, X=z-x\}}{\Pr\{Z=z\}} = \frac{\Pr\{Y=y\}\Pr\{X=z-x\}}{\Pr\{Z=z\}}.$$

Before going further, we have to compute the probability function of Z which appears in the denominator in the last term. This could be done by the convolution method, or by the method of moment generating functions, but let us take a more "probabilistic"approach. X can be thought of as the number of tosses of a coin with probability p of getting heads, until the first head occurs, and Y the number of additional tosses until the second head occurs. Then Z is the number of tosses required to get two heads, and so

$$\Pr\{Z=z\} = \underbrace{\binom{z-1}{1} pq^{z-2}}_{\text{one head in } z-1 \text{ tosses}} \cdot \underbrace{p}_{\text{head on } z\text{th toss}} = (z-1)p^2q^{z-2}, \quad z = 2, 3, \ldots$$

(Note that Z has the negative binomial distribution). Continuing on then,

$$f_{Y|Z}(y|z) = \frac{pq^{y-1}\cdot pq^{z-y-1}}{(z-1)p^2q^{z-2}} = \frac{1}{z-1}, \quad y = 1, 2, \ldots, z-1;\ z = 2, 3, \ldots$$

Thus, conditional on $Z = z$, the random variable Y is uniformly distributed on the integers $1, 2, \ldots, z-1$. For the conditional expectation, we then have

$$E(Y|X+Y=z) = \sum_{y=1}^{z-1} y\cdot\frac{1}{z-1} = \frac{z}{2}, \quad z = 2, 3, \ldots$$

Finally, substituting $X + Y$ for z, we have

$$E(Y|X+Y) = \frac{X+Y}{2}.$$

Again, we note that

$$E\big[E(Y|X+Y)\big] = E\left[\frac{X+Y}{2}\right] = \frac{E(X)+E(Y)}{2} = E(Y)$$

(since $E(X) = E(Y)$), the same type of result as in example 1. This is not an accident, as the following theorem shows.

Theorem 1.1.1 (The Law of Total Expectation) *For any random variables X and Y, we have*

$$E(Y) = E\big[E(Y|X)\big],$$

or in less succint notation,

$$E(Y) = \begin{cases} \int_{-\infty}^{\infty} E(Y|X=x) f_X(x)\,dx, & \text{if } X \text{ is a continuous random variable;} \\ \sum_x E(Y|X=x) f_X(x), & \text{if } X \text{ is a discrete random variable;} \end{cases}$$

where $f_X(x)$ is the density function, or probability function of X, depending on whether X is continuous or discrete.

Proof. Let us suppose that X is a discrete random variable. Then

$$\begin{aligned} E\big[E(Y|X)\big] = \sum_x E(Y|X=x) f_X(x) &= \begin{cases} \sum_x \left(\sum_y y f_{Y|X}(y|x) \right) f_X(x); & \text{if } Y \text{ is discrete,} \\ \sum_x \left(\int_{-\infty}^{+\infty} y f_{Y|X}(y|x)\,dy \right) f_X(x); & \text{if } Y \text{ is continuous,} \end{cases} \\ &= \begin{cases} \sum_x \sum_y y f(x,y); & \text{if } Y \text{ is discrete,} \\ \sum_x \int_{-\infty}^{+\infty} y f(x,y)\,dy; & \text{if } Y \text{ is continuous,} \end{cases} \\ &= \begin{cases} \sum_y y \sum_x f(x,y); & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{+\infty} y \sum_x f(x,y)\,dy; & \text{if } Y \text{ is continuous,} \end{cases} \\ &= \begin{cases} \sum_y y f_Y(y); & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{+\infty} y f_Y(y)\,dy; & \text{if } Y \text{ is continuous,} \end{cases} \\ &= E(Y). \end{aligned}$$

∎

The proof is identical, except for the obvious changes, if X is a continuous random variable. Note that in the above proof, $\sum_x$ means "sum over all values x in the range of the random variable X". It is possible that $f_{Y|X}(y|x)$ might be undefined for such a value x, but this can happen only if $f_X(x) = 0$, in which case the term corresponding to this x contributes nothing to the sum. The case where Y is continuous in the above proof might appear strange, for then $f(x,y)$ is the joint "density"function of a discrete random variable X and a continuous random variable Y. Handling discrete and continuous random variables jointly is a common occurrence. The best and most intuitive method is to specify their joint distribution by means of the conditional density function, as in the example below.

Example 1.3. 25 per cent of the transistors in a large bin are type 1 transistors, 25 per cent are type 2 transistors, and the remainder are type 3. The lifetime of a type 1 transistor is exponentially distributed with mean 1 hour. The lifetimes of type 2 and 3 transistors are exponential with means 30 minutes and 20 minutes respectively. What is the lifetime distribution of a transistor chosen randomly from the bin? What is the mean lifetime of such a transistor?

Solution. Let X = the type (1,2, or 3) of the chosen transistor, and let Y denote its lifetime. Then X has probability function

$$f_X(x) = \begin{cases} .25 & \text{if } x = 1, \\ .25 & \text{if } x = 2, \\ .50 & \text{if } x = 3, \end{cases}$$

and the conditional density of Y given X is

$$f_{Y|X}(y|x) = \begin{cases} x e^{-xy} & \text{if } y \geq 0, \\ 0 & \text{if } y < 0, \end{cases}$$

for $x = 1, 2, 3$. From the formula $f(x, y) = f_{Y|X}(y|x)f_X(x)$, we easily find

$$f(1, y) = .25e^{-y}, \quad f(2, y) = .50e^{-2y}, \quad f(3, y) = 1.5e^{-3y}$$

for $y \geq 0$, and $f(x, y) = 0$ when $y < 0$. Then, in the usual way,

$$f_Y(y) = \sum_x f(x, y) = \begin{cases} .25e^{-y} + .50e^{-2y} + 1.5e^{-3y} & \text{if } y \geq 0, \\ 0 & \text{if } y < 0. \end{cases}$$

This answers the first question. The answer to the second may now be directly computed as 13/24. However, $E(Y)$ may also be computed from the law of total expectation, without doing the first part of this problem. We have

$$\begin{aligned} E(Y) = E\big[E(Y|X)\big] &= .25E(Y|X = 1) + .25E(Y|X = 2) + .50E(Y|X = 3) \\ &= (.25 \times 1) + (.25 \times \frac{1}{2}) + (.50 \times \frac{1}{3}) \\ &= \frac{13}{24}. \end{aligned}$$

Example 1.3. Mrs. Brown owns a candy store which is frequented by the neighbourhood children. Suppose that the amount of time Mrs. Brown requires to serve a typical customer is a random variable with mean 2 minutes, and that in any period of time of length t minutes, the number of customers who will enter the store is a random variable with mean $3t$.

Linus, a steady customer, arrives at the store to purchase his usual fare. What is the expected number of customers to arrive during the time that Linus is served by Mrs. Brown?

Solution. Let T = the service time of Linus, and N = the number of customers to arrive while Linus is being served. Note that T is continuous and N is discrete. There is not enough information to determine the joint distribution, but we do not need this anyway.

If we knew the exact value of T, the problem would be easy. If, say, $T = t$, then the answer would be $3t$. In other words, $E(N|T = t) = 3t$. This means that $E(N|T) = 3T$. But then by the law of total expectation, we have

$$E(N) = E\big[E(N|T)\big] = E(3T) = 3E(T) = 6.$$

Proposition 1.1.2

(1) For any function $\phi(x, y)$, we have $E(\phi(X, Y)|X = x) = E(\phi(x, Y)|X = x)$.

(2) If X and Y are independent, $E(Y|X = x) = E(Y)$ for all x, and $E(Y|X) = E(Y)$.

Proof. Part(2) is easy to show, at least when X and Y are either both discrete or both continuous. Hence we prove only part 1, and this only in the case where X and Y are both discrete. We have

$$\begin{aligned} E(\phi(X, Y)|X = x) &= \sum_{x^* \in R_X} \sum_{y \in R_Y} \phi(x^*, y) \Pr\{X = x^*, Y = y|X = x\} \\ &= \sum_{x^* \in R_X} \sum_{y \in R_Y} \phi(x^*, y) \frac{\Pr\{X = x^*, X = x, Y = y\}}{\Pr\{X = x\}} \\ &= \sum_{y \in R_Y} \phi(x, y) \frac{\Pr\{X = x, Y = y\}}{\Pr\{X = x\}} \\ &= \sum_{y \in R_Y} \phi(x, y) \Pr\{Y = y|X = x\} \\ &= E(\phi(x, Y)|X = x), \end{aligned}$$

as required. In the above argument, x is fixed, and x^* was used to denote an arbitrary value of X. We also used the fact that

$$\Pr\{X = x^*, X = x, Y = y\} = \begin{cases} \Pr\{X = x, Y = y\} & \text{if } x^* = x, \\ 0 & \text{if } x^* \neq x. \end{cases}$$

We will use the results of this proposition in the next example.

Example 1.4. Suppose we have a sequence $X_1, X_2, \ldots$ of identically distributed random variables. Let N be another random variable, independent of the random variables $X_1, X_2, \ldots$, and taking values in the set $\{0, 1, 2, \ldots\}$. For any integer $n \geq 0$, define

$$S_n = \begin{cases} 0 & \text{if } n = 0, \\ X_1 + X_2 + \ldots + X_n & \text{if } n \geq 1. \end{cases}$$

Show that $E(S_N) = E(X)E(N)$, where $E(X)$ denotes the expectation of any one of the random variables $X_1, X_2, \ldots$.

Before giving the solution, let us make clear what is meant by S_N. S_N is a sum of a random number N of random variables. Suppose ω is an outcome in the sample space. Then S_N is defined by $S_N(\omega) = S_{N(w)}(w)$. To give a numerical illustration, suppose that

$$N(\omega) = 3, \quad X_1(\omega) = 2, \quad X_2(\omega) = -0.5, \quad X_3(\omega) = -1, \quad X_4(\omega) = 17.$$

Then

$$S_N(\omega) = S_{N(\omega)}(\omega) = S_3(\omega) = X_1(\omega) + X_2(\omega) + X_3(\omega) = 0.5.$$

Solution. The solution would be very simple if N were constant and not a random variable. The way to achieve this is to use the law of total expectation and write $E(S_N) = E\big[E(S_N|N)\big]$. This is called "conditioning on N". In order to compute $E(S_N|N)$, we first calculate $E(S_N|N = n)$. We have, by parts 1 and 2 of the last proposition, and the fact that N is independent of S_n for every n,

$$E(S_N|N = n) = E(S_n|N = n) = E(S_n) = E(X_1) + \ldots + E(X_n) = nE(X), \quad n \geq 0.$$

Then $E(S_N|N) = NE(X)$, and so by the law of total expectation, we have

$$E(S_N) = E\big[E(S_N|N)\big] = E\big[NE(X)\big] = E(N)E(X)$$

as required.

Problem 1.2. In example 1.4, suppose that $X_1, X_2, \ldots$ are independent of each other as well. Show that

$$Var(S_N) = E(N)Var(X) + \big(E(X)\big)^2 Var(N).$$

Example 1.5. Tent caterpillars live in colonies, and there may be several colonies on a tree. Suppose that the mean colony size is 1000 caterpillars, and the mean number of colonies per tree is 5. Then assuming colony sizes are independent of each other and of the number of colonies per tree, find the mean and variance of the number of caterpillars per tree.

Solution. Let N = number of colonies on the tree, and let X_i be the size of the ith colony. Then S_N is the number of caterpillars on the tree, and

$$E(S_N) = E(N)E(X) = 5 \times 1000 = 5000.$$

$Var(S_N)$ can be similarily computed.

Problem 1.3. Define the conditional variance of Y given X to be

$$Var(Y|X) = E(Y^2|X) - \big(E(Y|X)\big)^2.$$

Show that

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)].$$

Example 1.6. Theseus is at the centre of a labyrinth. There are three paths leading away from the centre. One of the paths will lead Theseus out of the labyrinth after one hour's walk. Another will lead Theseus out after two hour's walk, and the remaining path leads back to the centre after one hour's walk. If, from the centre of the labyrinth, the paths look indistinguishable to Theseus, find the expected amount of time in hours until Theseus exits from the labyrinth.

Solution. Let us number the paths 1,2, and 3 in the order described above. There is a positive probability that Theseus will find himself at the centre of the labyrinth several times. Let T denote the time in hours until Theseus exits from the labyrinth, and let X denote the number of the path which Theseus initially chooses. By the law of total expectation,

$$E(T) = \sum_{i=1}^{3} E(T|X = i) \Pr\{X = i\}.$$

Now $\Pr\{X = i\} = 1/3$ for each $i = 1, 2, 3$. Also, we know that $E(T|X = 1) = 1$ and $E(T|X = 2) = 2$. It only remains to calculate $E(T|X = 3)$. This involves the path that leads back to the centre, whereupon all conditions are the same as when he originally set out from the centre. Again he chooses one of the three paths (each with probability 1/3) and off he goes again. Hence

$$E(T|X = 3) = 1 + E(T).$$

We therefore have

$$E(T) = \sum_{i=1}^{3} E(T|X = i) \Pr\{X = i\} = \frac{1}{3}\big(1 + 2 + 1 + E(T)\big).$$

Solving for $E(T)$, we find $E(T) = 2$.

In the above problem, every point in time at which Theseus returns to the centre of the labyrinth is called a *renewal epoch*. At each renewal epoch, the problem begins again as if we were back at time zero, and everything that happens is independent of what went on before. You cannot completely understand the solution above until you have used the same method to successfully solve the following problem.

Problem 1.4. Find Var(T). (Hint: use the above method to find $E(T^2)$.)

Conditional probabilities such as $\Pr(A|X = x)$ and $P(A|X)$ where A is an event, can be defined in the same way as $E(Y|X = x)$ and $E(Y|X)$. and have much the same properties. For example, if X is discrete, then

$$\Pr(A|X = x) = \frac{\Pr(A \cap \{X = x\})}{\Pr(X = x)}$$

by the usual definition of conditional probability. Corresponding to the law of total expectation, we have the *Law of Total Probability*:

$$\Pr(A) = E\big[\Pr(A|X)\big],$$

or equivalently

$$\Pr(A) = \begin{cases} \sum_x \Pr(A|X = x) f_X(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{+\infty} \Pr(A|X = x) f_X(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

Example 1.7. In example 1.4, suppose that $X_1, X_2, \ldots$ are independent of each other, and that each X_i has the distribution

$$\Pr(X_i = 1) = p, \qquad \Pr(X_i = 0) = q,$$

where $q = 1 - p$ and $0 \leq p \leq 1$. Furthermore, suppose that N has the Poisson distribution with parameter λ. Find the distribution of S_N.

Solution. For each n, $S_n = X_1 + X_2 + \ldots + X_n$ has a binomial distribution with parameters n and p. Therefore $\Pr(S_N = k|N = n) = \Pr(S_n = k|N = n) = \Pr(S_n = k) = \binom{n}{k} p^k q^{n-k}$, and so by the law of total probability,

$$\begin{aligned}\Pr(S_N = k) &= \sum_{n=k}^{\infty} \Pr(S_N = k|N = n) \Pr(N = n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \cdot \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(\lambda q)^{n-k}}{(n-k)!} = \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda q} \\ &= \frac{(\lambda p)^k e^{-\lambda p}}{k!}.\end{aligned}$$

Thus S_N has the Poisson distribution with parameter λp.

Example 1.8. Nicholas is in a long line at a bank with a single teller. After he arrives at the front of the line, it takes what Nicholas thinks is an inordinately long time to be served. After he is served, Nicholas retires to a good vantage point and observes the service times of the customers following him in the line. Assuming that all service times are independent, identically distributed continuous random variables, find the probability distribution and expected value of the number of people served until one experiences worse luck than Nicholas.

Solution. Let X_0 be Nicholas's service time, and let $X_1, X_2, X_3, \ldots$ be the service times of the customers following Nicholas. We are assuming that $X_0, X_1, X_2, \ldots$ are i.i.d. continuous random variables. Let N be the first value of n such that $X_n > X_0$. We are to find the probability function and expected value of N.

First notice that $\{N > n\} = \{X_1 \leq X_0, X_2 \leq X_0, \ldots, X_n \leq X_0\}$, and so

$$\begin{aligned}\Pr\{N > n|X_0 = x\} &= \Pr\{X_1 \leq X_0, X_2 \leq X_0, \ldots, X_n \leq X_0|X_0 = x\} \\ &= \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x|X_0 = x\} \\ &= \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} \\ &= \Pr\{X_1 \leq x\} \Pr\{X_2 \leq x\} \ldots \Pr\{X_n \leq x\} \\ &= \big[F(x)\big]^n.\end{aligned}$$

where $F(x)$ is the distribution function of each of the random variables $X_0, X_1, \ldots$. Hence by the law of total probability, we have

$$\Pr\{N > n\} = E\big[\Pr(N > n|X_0)\big] = E\big[\big(F(X_0)\big)^n\big].$$

Now recall that since X_0 has distribution function $F(x)$, then $Y = F(X_0)$ is uniformly distributed on the interval $[0, 1]$. It follows that

$$\Pr\{N > n\} = E(Y^n) = \int_0^1 y^n \, dy = \frac{1}{n+1}.$$

Now we can easily find the probability function of N. We have

$$\Pr\{N = n\} = \Pr\{N > n-1\} - \Pr\{N > n\} = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}$$

for each $n \geq 1$, and so

$$E(N) = \sum_{n=1}^{\infty} n \Pr\{N = n\} = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty.$$

This is an interesting, even surprising result. Since $\Pr\{N = 1\} = 1/2$, the very next customer after Nicholas will experience worse luck with probability 1/2. However, Nicholas will have to wait, on the average, infinitely long until a customer experiences worse luck than he did. Perhaps Nicholas might feel singled out for poor treatment, but it is only the laws of chance at work.

Example 1.9 Let X and Y be independent continuous random variables with density functions $f_X(x)$ and $f_Y(y)$, and distribution functions $F_X(x)$ and $F_Y(y)$. Let $Z = X + Y$. Show that

$$F_Z(z) = \int_{-\infty}^{+\infty} F_Y(z-x) f_X(x)\, dx, \qquad f_Z(z) = \int_{-\infty}^{+\infty} f_Y(z-x) f_X(x)\, dx.$$

Solution. The second formula is obtained by differentiating the first one with respect to z. For the first formula, we use the Law of Total Probability to get

$$\begin{aligned} F_Z(z) = \Pr\{X + Y \leq z\} &= \int_{-\infty}^{+\infty} \Pr\{X + Y \leq z | X = x\} f_X(x)\, dx \\ &= \int_{-\infty}^{+\infty} \Pr\{x + Y \leq z | X = x\} f_X(x)\, dx = \int_{-\infty}^{+\infty} \Pr\{Y \leq z - x | X = x\} f_X(x)\, dx \\ &= \int_{-\infty}^{+\infty} F_Y(z-x) f_X(x)\, dx. \end{aligned}$$

Example 1.10 Jill, Janet, and Jean are playing a game in which Jill and Janet each try to guess Jean's weight. Jill guesses first and announces her guess, and then Janet guesses and announces her guess. Whoever's guess is closer to Jean's true weight is the winner.

Note that Janet has the advantage of knowing Jill's estimate before making her own. Can a strategy be devised for Janet so that her probability of winning is higher than Jill's? The answer is yes.

Let θ = Jean's true weight, and define

$$X_1 = \text{Jill's (reported) guess}, \quad X_2 = \text{Janet's (unreported) guess}.$$

We assume that X_1 and X_2 are i.i.d. continuous random variables each with median θ (i.e $\Pr\{X_1 \leq \theta\} = \Pr\{X_2 \leq \theta\} = 1/2$). Note that Janet's guess X_2 is independently arrived at. She does not report this guess, but uses her knowledge of X_1 and X_2 to arrive at her reported guess G by employing the following strategy:

$$G = \begin{cases} X_1 + \epsilon & \text{if } X_2 > X_1, \\ X_1 - \epsilon & \text{if } X_2 < X_1, \end{cases}$$

where ϵ is chosen to be at least as small as the units of measurement for weight used by the three women, so that θ cannot be closer to X_1 than to G.

Proposition 1.1.3 *If Janet adopts the above strategy, then* $\Pr\{\textit{Janet wins}\} = 3/4$.

Proof. Define $\tilde{X}_1 = X_1 - \theta$, $\tilde{X}_2 = X_2 - \theta$, and note that $\tilde{X}_1$ and $\tilde{X}_2$ are i.i.d. continuous with median zero. Then

$$\begin{aligned} \Pr\{\text{Janet loses}\} &= \Pr\{\text{loses}, X_1 < \theta\} + \Pr\{\text{loses}, X_1 > \theta\} \\ &= \Pr\{X_2 < X_1 < \theta\} + \Pr\{X_2 > X_1 > \theta\} \\ &= \Pr\{\tilde{X}_2 < \tilde{X}_1 < 0\} + \Pr\{\tilde{X}_2 > \tilde{X}_1 > 0\}. \end{aligned}$$

Let $f(x)$ and $F(x)$ denote the density function and distribution function of $\tilde{X}_1$. Then by the law of total probability,

$$\Pr\{\tilde{X}_2 < \tilde{X}_1 < 0\} = \int_{-\infty}^{+\infty} \Pr\{\tilde{X}_2 < \tilde{X}_1 < 0 | \tilde{X}_1 = x\} f(x)\, dx = \int_{-\infty}^{0} \Pr\{\tilde{X}_2 < x | \tilde{X}_1 = x\} f(x)\, dx$$
$$= \int_{-\infty}^{0} F(x) f(x)\, dx = \frac{\big(F(x)\big)^2}{2}\Big|_{-\infty}^{0} = \frac{1}{8},$$

since $F(0) = 1/2$, and similarly

$$\Pr\{\tilde{X}_2 > \tilde{X}_1 > 0\} = \int_{-\infty}^{+\infty} \Pr\{\tilde{X}_2 > \tilde{X}_1 > 0 | \tilde{X}_1 = x\} f(x)\, dx = \int_{0}^{+\infty} \Pr\{\tilde{X}_2 > x | \tilde{X}_1 = x\} f(x)\, dx$$
$$= \int_{0}^{+\infty} (1 - F(x)) f(x)\, dx = -\frac{\big(1 - F(x)\big)^2}{2}\Big|_{0}^{+\infty} = \frac{1}{8}.$$

∎

In conclusion, conditional expectation and conditional probability are important tools of the applied probabilist. Problems in applied probability and in particular stochastic processes usually involve several random variables simultaneously. Conditioning allows us to effectively fix one or more of these random variables as constants, and then concentrate on a simplified version of the problem. Once the simplified version is solved, the Law of Total Expectation (or Probability) enables us to complete the solution of the problem.

1.2 Probability Generating Functions

Definition 2.1. Suppose $g_0, g_1, g_2, \ldots$ is a (possibly infinite) sequence of real numbers. If there is a positive number ρ such that the series

$$G(s) = \sum_{k=0}^{\infty} g_k s^k = g_0 + g_1 s + g_2 s^2 + g_3 s^3 + \ldots \tag{2.1}$$

converges (to a finite number) for all s with $-\rho < s < \rho$, then $G(s)$ is called the generating function of the sequence, and we say $G(s)$ is *defined* for all s with $-\rho < s < \rho$.

Now it may be possible, as many of the examples below will show, to "sum up"the series on the right hand side of (2.1), and therefore obtain $G(s)$ in "closed form"(see example 2.2 below, for instance). In that case, the series on the right in (2.2) is the MacLaurin series expansion of $G(s)$, with radius of convergence ρ.

Example 2.1. Consider the sequence $g_0 = 1, g_1 = -1.5, g_2 = 3.2$. Then the generating function of this sequence is $G(s) = 1 - 1.5s + 3.2s^2$.

Example 2.2 Consider the sequence $g_k = (-1)^k,\ k = 0, 1, 2, \ldots$. The generating function for this sequence is

$$G(s) = \sum_{k=0}^{\infty} (-1)^k s^k = \sum_{k=0}^{\infty} (-s)^k. \tag{2.2}$$

If $-1 < s < +1$, this series converges to $1/(s+1)$. Hence $G(s)$ is defined for $-1 < s < +1$ and has the "closed form"

$$G(s) = \frac{1}{s+1}.$$

Example 2.3. Consider the sequence $g_k = 3/k!,\ k = 0, 1, 2, \ldots$. Then $G(s) = 3\sum_{k=0}^{\infty} s^k/k!$. We know that this series converges for all s, and in fact $G(s) = 3e^s$.

Sometimes we are given a generating function in closed form, and we want to "go backwards"to find the sequence $g_0, g_1, g_2, \ldots$

Example 2.4 For the generating function given in closed form as

$$G(s) = \frac{5s}{3-2s}, \quad -\frac{3}{2} < s < +\frac{3}{2}$$

find the sequence $g_0, g_1, g_2, \ldots$.

Solution. We use the identity

$$\frac{1}{1-t} = \sum_{k=0}^{\infty} t^k,$$

which is valid for all t with $-1 < t < +1$, to find that

$$G(s) = \frac{5s}{3-2s} = \frac{5}{3}s\frac{1}{1-\frac{2}{3}s} = \frac{5}{3}s\sum_{k=0}^{\infty}(\frac{2}{3}s)^k = \sum_{k=0}^{\infty}\frac{5}{3}(\frac{2}{3})^k s^{k+1} = \sum_{k=1}^{\infty}\frac{5}{3}(\frac{2}{3})^{k-1}s^k. \tag{2.3}$$

Now the MacLaurin series expansion of a function is unique. Therefore, equating coefficients of the two power series on the right hand sides of (2.1) and (2.3), we see that

$$g_k = \begin{cases} 0 & \text{if } k = 0, \\ \frac{5}{3}(\frac{2}{3})^{k-1}, & \text{if } k = 1, 2, \ldots \end{cases} \tag{2.4}$$

Alternate Solution. We can also obtain the sequence $g_0, g_1, g_2, \ldots$ as follows. Recall from calculus that if a function $G(s)$ has a MacLaurin series expansion, then that expansion is

$$G(s) = \sum_{k=0}^{\infty}\frac{G^{(k)}(0)}{k!}s^k,$$

where $G^{(0)}(s) = G(s)$ and $G^{(k)}(s)$ for $k \geq 1$ is the kth derivative of the function $G(s)$. Comparing coefficients with the right hand side of (2.1), we see that

$$g_k = \frac{G^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \ldots \tag{2.5}$$

Now the first few three derivatives of $G(s)$ are

$$G^{(1)}(s) = \frac{15}{(3-2s)^2}, \quad G^{(2)}(z) = \frac{60}{(3-2s)^3}, \quad G^{(3)}(s) = \frac{360}{(3-2s)^4},$$

and in general, we see that

$$G^{(k)}(s) = \frac{15k!2^{k-1}}{(3-2s)^{k+1}}, \quad k = 1, 2, 3, \ldots$$

Setting $s = 0$, and using (2.5), we obtain g_k again as in (2.4).

The method we used in the alternate solution of this last example is important, so we write it as a proposition.

Proposition 1.2.1 *Let $G(s)$ be a generating function. Then the original sequence $g_0, g_1, g_2, \ldots$ may be obtained from $G(s)$ from the equation*

$$g_k = \frac{G^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \ldots$$

Before going on to *probability* generating functions, we note that if the sequence $g_0, g_1, g_2, \dots$ is *bounded*; that is, if for some M we have $|g_k| \le M$ for all k, then the generating function $G(s)$ of the sequence is defined for all s with $-1 < s < +1$. This is because

$$\sum_{k=0}^{\infty} |g_k s^k| = \sum_{k=0}^{\infty} |g_k||s|^k \le M \sum_{k=0}^{\infty} |s|^k = \frac{M}{1-|s|} < +\infty$$

(provided $|s| < 1$) and so the series on the right hand side of (2.1) converges absolutely, and therefore converges, if $-1 < s < +1$.

Definition. Suppose in definition 2.1 that the numbers $g_0, g_1, g_2, \dots$ are all non-negative, and furthermore sum to 1. Then the generating function $G(s)$ of this sequence is called a *probability* generating function. If X is a random variable with values in the set $\{0, 1, 2, 3, \dots\}$, then the generating function

$$P_X(s) = E(s^X) = \sum_{k=0}^{\infty} \Pr\{X = k\} s^k \tag{2.6}$$

(i.e. here $g_k = \Pr\{X = k\}$) is called the probability generating function (pgf) of the random variable X.

Remarks.

(1) Since all the coefficients g_k in a probability generating function $G(s)$ are bounded by $M = 1$, then $G(s)$ is defined for all s with $-1 < s < +1$. Since $G(1) = 1$, then $G(s)$ is also defined at $s = 1$.

(2) One determines the probability generating function of X from the probability function of X via equation (2.6). Conversely, if one knows the probability generating function of a random variable X in closed form, then the probability function of X can be calculated via proposition 2.1; namely, through the formula

$$\Pr\{X = k\} = \frac{P_X^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \dots \tag{2.7}$$

Thus there is a one-to-one relationship between probability functions of non-negative, integer-valued random variables and probability generating functions.

(3) There is a close relationship beween the pgf $P_X(s) = E(s^X)$ of X and the moment generating function (mgf) $M_X(t) = E(e^{tX})$ of X. Obviously, we can get the mgf from the pgf by replacing s by e^t, and conversely the pgf from the mgf by replacing t by $\log s$. Note that whereas the mgf "generates"moments through the formula $E(X^k) = M^{(k)}(0)$, the pgf generates terms of the probability function of X through the formula in (2.7).

Example 2.5. Suppose X is a random variable having the distribution

$$\Pr\{X = 1\} = .2, \quad \Pr\{X = 2\} = .4, \quad \Pr\{X = 3\} = .4.$$

Find the pgf of X.

Solution. Since $\Pr\{X = k\} = 0$ when $k = 0$ or $k = 4, 5, 6, \dots$, we have

$$P_X(s) = .2s + .4s^2 + .4s^3.$$

Example 2.6. Suppose X has the Poisson distribution with parameter λ. Find the pgf and mgf of X.

Solution. We have

$$P_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}.$$

Replacing s by e^t, the mgf is therefore

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Example 2.7. Suppose X has the geometric distribution, with probability function given by

$$\Pr\{X = k\} = pq^{k-1}, \quad k = 1, 2, 3, \ldots$$

where $0 < p < 1$ and $q = 1 - p$. Find the pgf of X.

Solution. We have

$$P_X(s) = \sum_{k=1}^{\infty} pq^{k-1} s^k = ps \sum_{k=1}^{\infty} (qs)^{k-1} = \frac{ps}{1 - qs},$$

provided $|qs| < 1$. Since $q < 1$, this means that $P_X(s)$ is defined over an interval *larger* than $-1 < s \leq 1$.

Proposition 1.2.2 (Properties of Probability Generating Functions) *(1)* $P_X(0) = \Pr\{X = 0\}, \quad P_X(1) = 1$.

(2) For $0 \leq s \leq 1$, $P_X(s)$ *is an continuous, increasing, convex function of* s.

Proof. (1) is obvious. For (2), note that $P_X(s)$, being a MacLaurin series, has derivatives of all orders within its interval of definition. Furthermore, these derivatives are obtained termwise. That is,

$$P'_X(s) = \sum_{k=1}^{\infty} \Pr\{X = k\} k s^{k-1} \tag{2.8}$$

$$P''_X(s) = \sum_{k=2}^{\infty} k(k-1) \Pr\{X = k\} s^{k-2}, \tag{2.9}$$

and so on, for $-1 < s \leq 1$. Note that we will have $P_X^{(k)}(s) \geq 0$ for all s with $0 \leq s \leq 1$, and all k.

Proposition 1.2.3 (More Properties) *(1)* $E(X) = P'_X(1)$.

(2) $Var(X) = P''_X(1) + P'_X(1) - (P'_X(1))^2$.

Proof. For (1), just put $s = 1$ in (2.8). For (2), note that putting $s = 1$ in (2.9) gives $P''_X(1) = E\big(X(X-1)\big)$. Then

$$Var(X) = E(X^2) - \big(E(X)\big)^2 = E\big(X(X-1)\big) + E(X) - \big(E(X)\big)^2,$$

and is therefore as given.

Proposition 1.2.4

(1) If X *and* Y *are independent, then*

$$P_{X+Y}(s) = P_X(s) P_Y(s).$$

More generally, if $X_1, X_2, X_3, \ldots, X_n$ *are* $n \geq 1$ *independent random variables, and if* $S_n = X_1 + X_2 + \ldots + X_n$*, then*

$$P_{S_n}(s) = P_{X_1}(s) P_{X_2}(s) \ldots P_{X_n}(s).$$

(Of course, all random variables here have values in the non-negative integers.)

(2) Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables, and N a random variable independent of the X_i's (again, all these r.v.'s take their values in the non-negative integers). Let the random variable S_N be as defined in example 1.4. Then

$$P_{S_N}(s) = P_N\big(P_X(s)\big),$$

where $P_X(s)$ stands for the pgf of each of the X_i's.

Proof. For (1), we have $P_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y) = P_X(s)P_Y(s)$. (2) is proved in the same way. Now for (3). Since N is independent of the X_i's, it is also independent of S_n for any n. Hence

$$E\big(s^{S_N}|N=n\big) = E\big(s^{S_n}|N=n\big) = E\big(s^{S_n}\big) = P_{S_n}(s) = \big(P_X(s)\big)^n.$$

Using the law of total expectation, we then have

$$P_{S_N}(s) = E\big(s^{S_N}\big) = E\big[E\big(s^{S_N}|N\big)\big] = E\big[\big(P_X(s)\big)^N\big] = P_N\big(P_X(s)\big),$$

as required.

Example 2.8. Give an alternative solution to example 1.7, this time using generating functions.
Proof. We have $P_X(s) = q + ps = 1 + p(s-1)$, and from example 2.6, $P_N(s) = \exp\big[\lambda(s-1)\big]$. Hence by the above proposition,

$$P_{S_N}(s) = e^{\lambda(P_X(s)\,-\,1)} = e^{\lambda p(s\,-\,1)}.$$

By the uniqueness of pgf's, we therefore recognize S_N as being Poisson with parameter λp.

1.3 References

A good reference for conditional probability and conditional expectation is *Probability Models* by S.A. Ross. A good reference for generating functions, and one of the greatest books on mathematics ever is *An Introduction to Probability Theory and its Applications: Volume I*, by William Feller (one of the greatest mathematicians of this century).

Chapter 2

Two Examples of Stochastic Processes

2.1 The Galton-Watson-Bienaymé Branching Process

This will be our first example of a stochastic process. The theory of branching processes began about 1850, and originated with Galton, Watson, and Bienaymé. It has grown into a large area of applied probability, and much research is still being done on generalizations of what we are going to study.

The Historical Problem. Originally, interest in branching processes was fuelled by the following problem: Let $p_0, p_1, p_2, \ldots$ be the probabilities that a man has $0, 1, 2, \ldots$ sons, respectively, and let each son have the same probabilities for sons of his own, and so on. What is the probability that the male line is extinct after r generations, and more generally, what is the probability for any given number of descendants in the male line in any given generation? This is the problem of extinction of family names.

Abstract Formulation of the Problem. We consider particles which are able, by splitting or otherwise, to give birth to new particles of a like kind. An original set of particles forms the original, or zeroth generation. The direct descendants of the nth generation form the $(n+1)$st generation. In any generation, each particle has probability p_k of producing k new particles, where $k = 0, 1, 2, \ldots$, and $p_0 + p_1 + p_2 + \ldots = 1$. The particles of the nth generation act independently of one another and of the generation size. We are interested in the sizes of the successive generations, and in the probability of extinction.

Notation Let us define

$$Z_n = \text{ number of particles in the } n\text{th generation}, \quad n = 0, 1, 2, \ldots$$

$$P_n(s) = \text{the pgf of } Z_n, \quad n = 0, 1, 2, \ldots$$

$$P(s) = \sum_{k=0}^{\infty} p_k s^k,$$

Let also X be a generic random variable standing for the number of particles produced by a given particle. Then $\Pr\{X = k\} = p_k$ for every $k = 0, 1, 2, \ldots$, and $P_X(s) = P(s)$. We also define

$$m = E(X) = P'(1) = \text{the mean of the number of particles produced by a given particle}$$

and

$$\begin{aligned} \sigma^2 = Var(X) &= P''(1) + P'(1) - \left(P'(1)\right)^2 \\ &= \text{the variance of the number of particles produced by a given particle.} \end{aligned}$$

The stochastic process $\{Z_n, n \geq 0\}$ is called the Galton-Watson-Bienaymé branching process.

Assumption. Until the end of this section, we will assume that $Z_0 \equiv 1$; that is, there is *one* particle in the zeroth generation. Note then that $P_0(s) = s$ and $P_1(s) = P(s)$, $m = E(Z_1)$, $\sigma^2 = Var(Z_1)$.

Proposition 2.1.1 (The Fundamental Equations) *(1)* $P_{n+1}(s) = P_n\big(P(s)\big), \quad n = 0, 1, 2, \ldots.$

(2) $P_{n+1}(s) = P\big(P_n(s)\big), \quad n = 0, 1, 2, \ldots.$

Proof.

(1) Let us look carefully at how generation number n produces descendants for generation $n+1$. Let us number the particles in the nth generation as $1, 2, 3, \ldots, Z_n$. Let X_i denote the number of particles produced for the $(n+1)$st generation by the ith particle of the nth generation. Then

$$Z_{n+1} = X_1 + X_2 + \ldots + X_{Z_n}.$$

By assumption, the X_i's are independent of each other and of Z_n. Hence by proposition 1.2.3, we have

$$P_{Z_{n+1}}(s) = P_{Z_n}\big(P_X(s)\big).$$

Using the notation adopted above, and the fact that $P_X(s) = P(s)$, we obtain $P_{n+1}(s) = P_n(P(s))$, as required.

(2) This may be proved directly from (1), or as follows. Let Y_i denote the number of particles in the $(n+1)$st generation that can trace their ancestry to the ith particle in the first generation, $i = 1, 2, \ldots, Z_1$. Then

$$Z_{n+1} = Y_1 + Y_2 + \ldots + Y_{Z_1}.$$

Now the Y_i's are independent of one another, and of Z_1, and since Y_i has the same distribution as Z_n (because of our assumption that $Z_0 = 1$), then $P_{Y_i}(s) = P_n(s)$. Then again by proposition 1.2.3, we have the required result.

∎

Example 1.1. Suppose the p_k's are given by

$$p_k = pq^k, \quad k = 0, 1, 2, \ldots$$

Then

$$P(s) = \sum_{k=0}^{\infty} pq^k s^k = \frac{p}{1-qs}.$$

Using the above proposition, the pgf of Z_2 is

$$P_2(s) = P_1(P(s)) = \frac{p}{1-qP(s)} = \frac{p(1-qs)}{1-qs-pq}.$$

Carrying on, we can use either of the fundamental recursion formulas to find $P_n(s)$ explicitly for any n. Then the pgf $P_n(s)$ can be "inverted"to find the terms $\Pr\{Z_n = k\}$ of the probability function of Z_n.

Thus, in theory at least, one can use the fundamental recursion formulas of proposition 1.1 to find the distribution of Z_n for any n.

Proposition 2.1.2

(1) If $m < +\infty$, *then* $E(Z_n) = m^n$ *for all* $n \geq 0$.

(2) If $\sigma^2 < +\infty$, *then*

$$Var(Z_n) = \begin{cases} \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m} & \text{if } m \neq 1, \\ n\sigma^2 & \text{if } m = 1. \end{cases}$$

Proof. To prove (1), we differentiate either of the recursion formulas in proposition 1.1, say the first, and obtain

$$P'_{n+1}(s) = P'_n\big(P(s)\big)P'(s)$$

from the chain rule. Now put $s = 1$, and recall that $P(1) = 1$. We get

$$E(Z_{n+1}) = E(Z_n)m, \tag{1.1}$$

valid for any $n \geq 0$. We can iterate equation (1.1) repeatedly, getting

$$E(Z_{n+1}) = E(Z_n)m = E(Z_{n-1})m^2 = E(Z_{n-2})m^3 = \ldots E(Z_1)m^n = m^{n+1}.$$

This proves the first part of the proposition. The proof of the second part is similar (you have to differentiate the recursion formula twice) and is left as a problem. ∎

We now are able to calculate the distribution of any of the random variables $Z_n, n \geq 0$ using the recursion formulas, as in example 1.1 (at least in theory—the calculations may be quite difficult), and we can use proposition 1.2 to easily find the mean and variance of any of the Z_n's. We now turn to something possibly more interesting—the probability of eventual extinction of the process.

If $p_0 = 0$, each particle produces at least one new particle for the next generation, and the branching process can never die out. If moreover, $p_1 = 1$, there will be exactly one particle in each generation, while if $p_1 < 1$, we will have $Z_n \to +\infty$ as $n \to \infty$. But if $p_0 > 0$, we may have $Z_n = 0$ for some n. In this case, there would be no particles left to produce particles for the next generation, and so the process would have died out (i.e. become extinct). Hence we will now assume $p_0 > 0$, and we will determine the probability ζ of eventual extinction of our branching process.

First, we note that the event "extinction" can be written

$$\{\text{extinction}\} = \{Z_n = 0 \text{for some } n \geq 1\} = \cup_{n=1}^{\infty}\{Z_n = 0\} = \lim_{n\to\infty} \uparrow \{Z_n = 0\},$$

because $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$ for all n. By what is called the "continuity" property for probabilities, we then have

$$\zeta = \Pr\{\text{extinction}\} = \lim_{n\to\infty} \uparrow \Pr\{Z_n = 0\}. \tag{1.2}$$

Note that the probabilities $\Pr\{Z_n = 0\}$ are getting larger as n increases. That is what the $\uparrow$ means.

Theorem 2.1.3 *ζ is the smallest non-negative root of the equation $P(s) = s$.*

Proof. First, note that $P(1) = 1$, so that 1 is a root, and therefore a candidate for ζ (but there may be a smaller non-negative root than 1). Let us write $x_n = P_n(0) = \Pr\{Z_n = 0\}$ for $n \geq 0$, so that by the second of the two recursion equations, we have

$$x_{n+1} = P_{n+1}(0) = P(P_n(0)) = P(x_n)$$

for all $n \geq 0$. Using this fact, equation (1.2), and the continuity of the pgf $P(s)$, we get

$$\zeta = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} P(x_n) = P(\lim_{n\to\infty} x_n) = P(\zeta).$$

This shows that ζ is a root of the equation $P(s) = s$. We now have to show it is the smallest non-negative root. Suppose that ρ is any non-negative root of $P(s) = s$. We have $x_0 = 0$, and so $x_0 \leq \rho$. Since $P(s)$ is an increasing function of s for $s \geq 0$, we find

$$x_1 = P(x_0) \leq P(\rho) = \rho.$$

Repeating, we get

$$x_2 = P(x_1) \leq P(\rho) = \rho,$$

and so on, finding that $x_n \leq \rho$ for every n. But then $\zeta = \lim_{n\to\infty} x_n \leq \rho$, and so ζ really is the *smallest* such root.

Proposition 2.1.4 *If $m \leq 1$, then $\zeta = 1$. If $m > 1$, then $\zeta < 1$.*

Proof. The proof is really graphical. Recall that the pgf $P(s)$ is convex, increasing, and continuous for $0 \leq s \leq 1$, that $P(0) = p_0 > 0$, and that $P(1) = 1$. Also, the slope of the curve $P(s)$ at $s = 1$ is $P'(1) = m$. Thus the graph of $P(s)$ can look only like one of the two cases shown in figure 1.1 below. The top curve is typical of the case $m \leq 1$, and one can see that it cannot intersect the diagonal straight line except at $s = 1$, so that ζ in this case is 1. The bottom curve is typical of the case $m > 1$, and we see that it must intersect the diagonal exactly once at a point strictly between 0 and 1. By theorem 1.3, this point must be ζ,so $\zeta < 1$ in this case. ∎

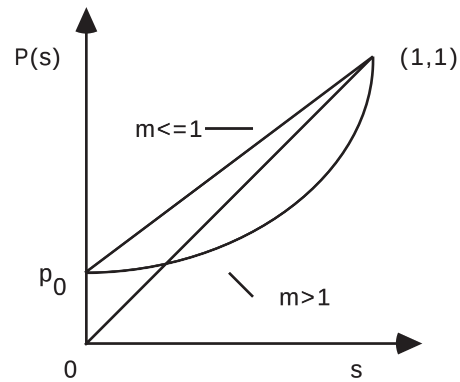

Figure 1.1

Note that we really needed $p_0 > 0$ in this proposition, to rule out the case where $p_1 = 1$ (and as a result, $m = 1$ and $\zeta = 0$, contradicting the proposition).

Example 1.2. Let us return to the problem of extinction (or survival) of family names, posed at the very beginning of this section. In 1939, Lotka used the methods of this section to calculate the extinction probability for American male lines of descent. From the 1920 U.S. census, he found that the probabilities $p_k, k \geq 0$ were well represented by the distribution

$$p_k = bc^{k-1}, \quad k \geq 1, \quad p_0 = 1 - \sum_{k=1}^{\infty} p_k,$$

where $b = 0.2126$ and $c = 0.5893$.

The basic pgf $P(s)$ is

$$\begin{aligned} P(s) &= p_0 + \sum_{k=1}^{\infty} p_k s^k = p_0 + \sum_{k=1}^{\infty} bc^{k-1} s^k \\ &= p_0 + bs \sum_{k=1}^{\infty} (cs)^{k-1} = p_0 + \frac{bs}{1 - cs}. \end{aligned} \tag{1.3}$$

Putting $s = 1$ and using the fact that $P(1) = 1$, we can solve for p_0, and we find

$$p_0 = \frac{1 - b - c}{1 - c} = 0.4825.$$

Differentiating (1.3), and setting $s = 1$ gives

$$m = P'(1) = \frac{b}{(1 - c)^2} = 1.261,$$

so that by proposition 1.4, we know that $\zeta < 1$. To find the exact value of ζ, we have to find the smallest non-negative root of $P(s) = s$, that is, of

$$p_0 + \frac{bs}{1 - cs} = s.$$

Taking a common denominator, and substituting the numerical values of b, c, and p_0, we have to solve a simple little quadratic equation. The roots are $s = 1$ (note: 1 is *always* a root of $P(s) = s$) and $s = 0.816$. We therefore conclude that $\zeta = 0.816$.

To conclude this section, we will now lift the assumption that $Z_0 = 1$. We will assume that $Z_0 = i \geq 1$. Now $P_1(s)$ is no longer $P(s)$ (unless $i = 1$). Then each of the i particles in the zeroth generation produces its own private branching process, and these processes are independent of one another. We therefore obtain, among others, the following facts:

(1) $E(Z_n) = im^n$ for all $n \geq 0$

(2) $\Pr\{\text{extinction}\} = \zeta^i$, where ζ is the smallest non-negative root of $P(s) = s$.

2.2 The Simple Random Walk on the Integers

Definition. Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables, each having the distribution

$$\Pr\{X_i = x\} = \begin{cases} p & \text{if } x = +1, \\ q & \text{if } x = -1, \end{cases}$$

where $0 \leq p \leq 1$ and $q = 1 - p$. Thus each X_i is a Bernoulli random variable with just the two possible values -1 and $+1$. Now define

$$S_n = \begin{cases} 0 & \text{if } n = 0, \\ X_1 + X_2 + \ldots + X_n & \text{if } n \geq 1, \end{cases}$$

where n is any integer ≥ 0. The stochastic process $\{S_n, n \geq 0\}$ is called a *simple random walk on the integers starting from* 0 (the last because $S_0 = 0$).

The simple random walk on the integers gets its name as follows. Think of a particle which is able to bounce from integer to integer on the infinitely long horizontal line in figure 2.1 below. We assume that initially the particle is at position 0. Also, we have a coin which, when tossed, will give heads with probability p, and tails with probability q. We toss the coin. If we get a head, we move the particle one unit to the right. If we get tails, we move the particle one unit to the left. Thus the particle is now at position $S_1 = X_1$. We toss the coin again. If the result is heads, the particle is moved one step to the right—if not, then one step to the left. The particle is now at position $S_2 = X_1 + X_2$. Repeating this process, we see that for any n, S_n is the position of this particle after n jumps (i.e. n tosses of the coin). We say that the particle is executing a simple random walk on the integers, where the term "simple" refers to the fact that the particle can jump from where it is at a given time only to one of the nearest neighbour integers. As a last point, we note that the stochastic process made up of the random variables $i + S_n, n \geq 0$, where i is an integer, would describe a random walk particle, but this time starting from position i. This would just be a translation of the motion starting from 0, so in the remainder of this section, we will assume $i = 0$.

The next example shows that the distribution of S_n is almost binomial.

Example 2.1. What is the probability that

(1) after four steps, the particle is at position 6?

(2) after four steps, the particle is at position -4?

(3) after three steps, the particle is at position 2?

(4) after five steps, the particle is in position 3?

Solution Obviously, the solutions to (1),(2), and (3) are 0,q^4, and 0 respectively. So let us turn to (4) Let us find in general the probability that after n steps, the particle is at position k; that is, we will find $\Pr\{S_n = k\}$. To do this, let r denote the number of steps the particle took to the right among the first n steps, so that $n-r$ is the number of steps to the left. If after n steps, the particle is at k, then $r-(n-r)=k$, so $r=(k+n)/2$. Hence

$$\Pr\{S_n = k\} = \Pr\{\frac{k+n}{2}\text{successes in } n \text{ trials}\} = \binom{n}{\frac{k+n}{2}} p^{\frac{k+n}{2}} q^{\frac{n-k}{2}}, \tag{2.1}$$

where of course $(k+n)/2$ must be an integer. Hence the answer to 4. is $\Pr\{S_5 = 3\} = \binom{5}{4}p^4 q = 5p^4 q$.

Now we turn to something more interesting. We are going to answer the following three questions.

(1) What is the probability that the particle, once it starts the random walk, will *ever* return to zero?

(2) On the average, how long (i.e. how many steps) does it take before the particle returns to zero?

(3) On the average, how many times will the particle return to zero?

Before getting under way, we pause to develop some notation. Let T be a random variable representing the number of steps required for the particle to return (for the first time) to zero, and let Y be the number of times the particle returns to zero. Define

$$f_n = \begin{cases} 0 & \text{if } n = 0, \\ \Pr\{T = n\} = \Pr\{S_1 \neq 0, S_2 \neq 0, \dots S_{n-1} \neq 0, S_n = 0\} & \text{if } n \geq 1, \end{cases}$$

$$u_n = \begin{cases} 1 & \text{if } n = 0, \\ \Pr\{S_n = 0\} & \text{if } n \geq 1, \end{cases}$$

and

$$f = \sum_{n=1}^{\infty} f_n.$$

Then questions 1,2, and 3 are to calculate f, $E(T)$, and $E(Y)$ respectively. We begin with the following lemma.

Lemma 2.2.1

$$u_n = \sum_{i=0}^{n} f_i u_{n-i} \quad \textit{for all } n \geq 1. \tag{2.2}$$

Proof. We have

$$
\begin{aligned}
u_n = \Pr\{S_n = 0\} &= \sum_{i=1}^{n} \Pr\{T = i, S_n = 0\} \\
&= \sum_{i=1}^{n} \Pr\{S_1 \neq 0, S_2 \neq 0, \ldots S_{i-1} \neq 0, S_i = 0, X_{i+1} + \ldots + X_n = 0\} \\
&= \sum_{i=1}^{n} \Pr\{S_1 \neq 0, S_2 \neq 0, \ldots S_{i-1} \neq 0, S_i = 0\} \Pr\{X_{i+1} + \ldots + X_n = 0\} \\
&= \sum_{i=1}^{n} f_i u_{n-i},
\end{aligned}
$$

where we used the fact that $S_1, S_2, \ldots S_i$ are independent of $X_{i+1}, \ldots X_n$. Since $u_0 = 0$, we can start the summation at $i = 0$ rather than $i = 1$. ∎

Equation (2.2) is called the *renewal equation*, and occurs often in the study of stochastic processes. We will encounter it again in the next chapter. Note that we already know the u_n's (from equation (2.1) with $k = 0$), so we are going to use the renewal equation to find the f_n's and f. We will "solve"the renewal equation by the method of generating functions. Define

$$
U(s) = \sum_{n=0}^{\infty} u_n s^n, \qquad F(s) = \sum_{n=0}^{\infty} f_n s^n.
$$

Lemma 2.2.2

$$
F(s) = 1 - \frac{1}{U(s)}.
$$

Proof. We have

$$
\begin{aligned}
F(s)U(s) &= (f_0 + f_1 s + f_2 s^2 + \ldots) \cdot (u_0 + u_1 s + u_2 s^2 + \ldots) \\
&= f_0 u_0 + (f_0 u_1 + f_1 u_0)s + (f_0 u_2 + f_1 u_1 + f_2 u_0)s^2 + \ldots \\
&= 0 + u_1 s + u_2 s^2 + \ldots \\
&= U(s) - 1,
\end{aligned}
$$

where we used the renewal equation. Now just solve for $F(s)$. ∎

We now have all the tools we need. Certainly $u_n = 0$ if n is odd, and from equation (2.1),

$$
u_{2n} = \binom{2n}{n} p^n q^n, \quad n \geq 0.
$$

Then

$$
U(s) = \sum_{n=0}^{\infty} u_{2n} s^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} (pqs^2)^n = \frac{1}{\sqrt{1 - 4pqs^2}}. \tag{2.3}
$$

(For the last step, you will have to check that the MacLaurin series expansion of the function $1/\sqrt{1-4x}$ is $\sum_{n=0}^{\infty} \binom{2n}{n} x^n$.) It now follows from lemma 2.2 that

$$
F(s) = 1 - \sqrt{1 - 4pqs^2}.
$$

The probabilities f_n may now be computed by inverting this generating function. This will be left as a problem. Let us now proceed to answer the three questions posed at the beginning of this section.

(1) The probability f of eventually returning to 0 is given by

$$f = F(1) = 1 - \sqrt{1-4pq} = 1 - \sqrt{1-4p(1-p)} = 1 - |p-q|.$$

Note that if $p = q = 1/2$, we have $f = 1$, so that the random walk particle returns to 0 with probability 1. But if $p \neq q$, then $f < 1$, so that $\Pr\{T = +\infty\} = 1 - f > 0$.

(2) If $p \neq q$, we saw in (1) that $\Pr\{T = +\infty\} > 0$, and this implies that $E(T) = +\infty$. Next assume that $p = q = 1/2$. Then $F(s) = 1 - \sqrt{1-s^2}$ is the pgf of T, and since $F'(s) = 1/\sqrt{1-s^2}$, then $E(T) = F'(1) = +\infty$. Thus either way, for *all* p, we have $E(T) = +\infty$.

(3) Define the random variables $Y_n,\ n \geq 1$ by

$$Y_n = \begin{cases} 1 & \text{if } S_n = 0, \\ 0 & \text{if } S_n \neq 0. \end{cases}$$

Then the number of times that the particle will return to 0 is the random variable $Y = \sum_{n=1}^{\infty} Y_n$, so that $E(Y) = \sum_{n=1}^{\infty} E(Y_n)$. But $E(Y_n) = 1 \cdot \Pr\{Y_n = 1\} + 0 \cdot \Pr\{Y_n = 0\} = \Pr\{Y_n = 1\} = \Pr\{S_n = 0\} = u_n$, so $E(Y) = \sum_{n=1}^{\infty} u_n = U(1) - 1$. We have already determined $U(s)$ in (2.3), so

$$E(Y) = \frac{1}{\sqrt{1-4pq}} - 1 = \frac{1}{|p-q|} - 1.$$

Notice that $E(Y) = +\infty$ if and only if $p = q$.

In summary, the case $p = q$ presents an interesting "paradox", since the particle returns to 0 with probability 1, but takes on the average infinitely long (i.e. infinitely many steps) in order to do so. Furthermore, on the average, it returns infinitely many times.

It is interesting to carry over the idea of a simple random walk to more than one dimension. To illustrate, let us consider a simple random walk on the two-dimensional integer lattice. The two-dimensional integer lattice is the set of all pairs (m, n) of integers, as shown in figure 2.2. Given a certain "point"(m, n) in the lattice as shown, the *nearest neighbours* of (m, n) are the four points $(m+1, n)$, $(m, n-1)$, $(m-1, n)$, and $(m, n+1)$ as shown by the x's in the figure. We define a simple random walk on this lattice as follows: for convenience, the particle starts out from the point $(0, 0)$. If at any time the particle is at the point (m, n), then on the next step it will go to one of the nearest neighbours of (m, n). It will go to $(m+1, n)$ (i.e. right) with probability p_1, to $(m, n-1)$ (i.e. down) with probability p_2, to $(m-1, n)$ (i.e. left) with probability p_3, or to $(m, n+1)$ (i.e. up) with probability p_4, where $p_1 + p_2 + p_3 + p_4 = 1$. A three dimensional simple random walk is similarily defined, except that now there are six nearest neighbours to every point. In fact, we can conceive of a simple random walk on an integer lattice of any dimension.

A simple random walk is called *symmetrical* if from any given position the particle goes at the next step to any of its nearest neighbours with the same probability. That is, if $p = q = 1/2$ in one dimension, or if $p_1 = p_2 = p_3 = p_4 = 1/4$ in two dimensions, and so on. In that case, we have the following result.

Theorem 2.2.3 *For a symmetrical simple random walk on the d-dimensional integer lattice, we have*

$$f = \begin{cases} 1 & \text{if } d = 1 \text{ or } 2, \\ < 1 & \text{if } d \geq 3. \end{cases}$$

2.3 The Gambler's Ruin Problem

The random walk studied in the last section could more properly be called a simple *unrestricted* random walk on the integers. In this section, we examine the same motion, but we place boundaries (or barriers) beyond which the particle cannot pass. We will actually study this problem in a more interesting scenario, called the Gambler's Ruin Problem. Suppose we have two gamblers, A and B, who play a series of "hands". On each hand, gambler A will win (and receive one dollar from gambler B) with probability p, or lose (and pay one dollar to gambler B) with probability q, where $0 < p < 1$ and $p + q = 1$ (the case $p = 0$ or $p = 1$ is trivial). Of course, gambler A winning is the same as gambler B losing. Gambler A starts with a fortune of k dollars, and gambler B starts with ℓ dollars, so the total "pot"is $a = k + \ell$ dollars. The game (i.e. the series of hands) ends when one of the gamblers is "ruined"(has his fortune reduced to zero).

We will look at the game through the eyes of gambler A. Let S_n =the fortune of gambler A after n hands. Then $S_0 = k$, and we easily see that S_n is the position after n steps of a particle undergoing a simple random walk on the integers as in the previous section, except that now the particle starts at k, and the walk ends when the particle hits either of the barriers 0 and a. It will be more interesting, however, to pursue the gambling analogy. Therefore define

$$q_k = \Pr\{\text{gambler A is eventually ruined}|\text{gambler A starts with } k \text{ dollars}\}, \quad k = 0, 1, 2, \ldots, a$$

and observe that $q_0 = 1$, $q_a = 0$. Then if $1 \leq k \leq a - 1$, we have

$$\begin{aligned} q_k &= \frac{\Pr\{\text{A ruined}, S_0 = k\}}{\Pr\{S_0 = k\}} = \frac{\Pr\{\text{A ruined}, S_0 = k, S_1 = k-1\}}{\Pr\{S_0 = k\}} + \frac{\Pr\{\text{A ruined}, S_0 = k, S_1 = k+1\}}{\Pr\{S_0 = k\}} \\ &= \Pr\{\text{A ruined}|S_1 = k-1, S_0 = k\} \Pr\{S_1 = k-1|S_0 = k\} \\ &\qquad + \Pr\{\text{A ruined}|S_1 = k+1, S_0 = k\} \Pr\{S_1 = k+1|S_0 = k\} \\ &= \Pr\{\text{A ruined}|\text{starts at } k-1\}q + \Pr\{\text{A ruined}|\text{starts at } k+1\}p \\ &= q_{k-1}q + q_{k+1}p. \end{aligned}$$

Thus we have to solve the difference equation

$$q_k = qq_{k-1} + pq_{k+1}, \quad k = 1, 2, \ldots, a, \quad \text{with boundary conditions } q_0 = 1 \text{ and } q_a = 0. \tag{3.1}$$

Let us first examine the difference equation

$$x_k = qx_{k-1} + px_{k+1}, \quad k = 1, 2, \ldots, a-1 \tag{3.2}$$

without any boundary conditions. Since $x_k = px_k + qx_k$, the equation can be rearranged as

$$x_{k+1} - x_k = \frac{q}{p} \cdot \big(x_k - x_{k-1}\big).$$

Iterating, we find

$$\begin{aligned} x_{k+1} - x_k &= \frac{q}{p} \cdot \big(x_k - x_{k-1}\big) \\ &= \big(\frac{q}{p}\big)^2 \cdot \big(x_{k-1} - x_{k-2}\big) \\ &= \ldots \\ &= \big(\frac{q}{p}\big)^k \cdot \big(x_1 - x_0\big). \end{aligned}$$

Now we sum both sides of this equation from $k = 0$ to $k = n-1$, and get

$$x_n - x_0 = \sum_{k=0}^{n-1} (x_{k+1} - x_k) = (x_1 - x_0) \sum_{k=0}^{n-1} \big(\frac{q}{p}\big)^k = \begin{cases} \frac{1-(q/p)^n}{1-q/p} \cdot (x_1 - x_0) & \text{if } p \neq q, \\ n(x_1 - x_0) & \text{if } p = q. \end{cases} \tag{3.3}$$

Returning now to (3.1), we have from (3.3) and the fact that $q_0 = 1$,

$$q_n - 1 = \begin{cases} \frac{1-(q/p)^n}{1-q/p} \cdot (q_1 - 1) & \text{if } p \neq q \\ n(q_1 - 1) & \text{if } p = q. \end{cases} \tag{3.4}$$

Taking $n = a$ in (3.4), we find that

$$q_1 - 1 = \begin{cases} -\frac{1-q/p}{1-(q/p)^a} & \text{if } p \neq q \\ -\frac{1}{a} & \text{if } p = q. \end{cases}$$

Substituting this back into (3.4), and rearranging, we get

$$q_n = \begin{cases} \frac{(q/p)^n - (q/p)^a}{1-(q/p)^a} & \text{if } p \neq q, \\ 1 - \frac{n}{a} & \text{if } p = q \end{cases} \tag{3.5}$$

for $n = 0, 1, 2, \ldots, a$.

Now that the ruin probabilities q_n for gambler A have been computed, we turn to finding the expected duration of the game. This is the expected number of hands until either gambler A or B is ruined. Define

$$D_k = E(\text{duration of game } | \text{gambler A starts with } k \text{ dollars}), \quad k = 0, 1, \ldots, a.$$

Then

$$D_k = q(D_{k-1} + 1) + p(D_{k+1} + 1),$$

and we see that the D_k's satisfy the difference equation

$$D_k = qD_{k-1} + pD_{k+1} + 1, \quad k = 0, 1, \ldots, a \quad \text{with boundary conditions } D_0 = D_a = 0.$$

The way we solve this equation is as follows. It is straightforward to check that

$$\Delta_k = \begin{cases} \frac{k}{q-p} & \text{if } p \neq q, \\ -k^2 & \text{if } p = q, \end{cases}$$

is a particular solution of the equation $D_k = qD_{k-1} + pD_{k+1} + 1$ (without the boundary conditions). Let D_k be any other solution. Then $x_k = D_k - \Delta_k$ is a solution of the difference equation in (3.2) and so by (3.3), we get

$$D_n = \Delta_n + x_0 + \begin{cases} \frac{1-(q/p)^n}{1-q/p} \cdot (x_1 - x_0) & \text{if } p \neq q, \\ n(x_1 - x_0) & \text{if } p = q. \end{cases}$$

Now we use the boundary conditions $D_0 = D_a = 0$ to determine the unknown terms x_0 and $x_1 - x_0$. Leaving the details as an exercise, we finally find

$$D_n = \begin{cases} \frac{n}{q-p} - \frac{a}{q-p} \cdot \frac{1-(q/p)^n}{1-(q/p)^a} & \text{if } p \neq q, \\ n(a-n) & \text{if } p = q \end{cases} \tag{3.6}$$

for $n = 0, 1, 2, \ldots, a$.

Example 3.1 Suppose $p = q = 1/2$. If $k = 500$ and $\ell = 500$, then $a = 1000$, $q_k = 1/2$, and $D_k = 250,000$. If $k = 1$, $\ell = 1000$, then $q_k = .999$ and $D_k = 1000$. Note that in this second case, it is a virtual certainty that gambler A will be ruined, but he will have a lot of fun for his dollar.

Example 3.2 It is easy to see from (3.5) and (3.6) that as ℓ and therefore a tends to $+\infty$, we have

$$q_n \longrightarrow \begin{cases} 1 & \text{if } p \leq q, \\ \left(\frac{q}{p}\right)^n & \text{if } p > q, \end{cases}$$

and that

$$D_n \longrightarrow \begin{cases} \frac{n}{q-p} & \text{if } p < q, \\ +\infty & \text{if } p \geq q. \end{cases}$$

Therefore if we think of gambler B as an infinitely rich adversary, say (to all intents and purposes) a casino, and if $p < q$, as it would be in a casino, then $q_n = 1$ and $D_n = n/(q-p)$. Thus if you went with one dollar to a casino where $p = .43$ and $q = .57$ (typical odds), you would certainly lose your dollar, but on the average, you would get to play about six hands.

Chapter 3

Discrete-Time Markov Chains

3.1 Definitions and Examples

Definition. Let $\{X_n, n \geq 0\}$ be a discrete-time stochastic process with countable state space E. $\{X_n, n \geq 0\}$ is a *Markov chain* if for every integer $n \geq 0$ and set $i, j; i_0, i_1, \ldots, i_{n-1}$ of states in E, we have

$$\Pr\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \Pr\{X_{n+1} = j | X_n = i\}. \tag{1.1}$$

(1.1) is called the *Markov property*.

Theorem 3.1.1 *Let $\{X_n, n \geq 0\}$ be a Markov chain. Then for any $m \geq 1$, any set $0 \leq n_1 < n_2 < \ldots < n_k < \ldots < n_m$ of "times", and any set $i_1, i_2, \ldots, i_k, \ldots, i_m$ of states in E, we have*

$$\begin{aligned} &\Pr\{X_{n_m} = i_m, \ldots, X_{n_{k+1}} = i_{k+1} | X_{n_k} = i_k, \ldots, X_{n_1} = i_1\} \\ &\qquad = \Pr\{X_{n_m} = i_m, \ldots, X_{n_{k+1}} = i_{k+1} | X_{n_k} = i_k\}. \end{aligned} \tag{1.2}$$

(1.2) is also called the Markov property, and includes (1.1) as a special case. However, the above theorem says that (1.1) and (1.2) are actually equivalent statements. The proof that (1.1) implies (1.2) is not easy, though.

Definition. The conditional probability $\Pr\{X_{n+1} = j | X_n = i\}$ on the right hand side of (1.1) is called a *one-step transition probability*. If $\Pr\{X_{n+1} = j | X_n = i\}$ does not depend on n, for all states $i, j \in E$, we say the Markov chain has *stationary transition probabilities*, and we write

$$P_{ij} = \Pr\{X_{n+1} = j | X_n = i\} = \Pr\{X_1 = j | X_0 = i\}.$$

Example 1.1. A GWB branching process $\{Z_n, n \geq 0\}$ is a Markov chain with state space $E = \{0, 1, 2, 3, \ldots\}$ and stationary transition probabilities. For fix a value of $n \geq 0$, and let X_k denote the number of particles in the $n+1$ st generation produced by the kth particle in the nth generation. Then

$$\Pr\{Z_{n+1} = j | Z_n = i, Z_{n-1} = i_{n-1}, \ldots, Z_0 = i_0\} = \Pr\{X_1 + X_2 + \ldots + X_i = j\} = \Pr\{Z_{n+1} = j | Z_n = i\}.$$

Notice that the one-step transition probabilities are given by $P_{ij} = \Pr\{X_1 + X_2 + \ldots + X_i = j\}$.

Example 1.2. A simple random walk $\{S_n, n \geq 0\}$ on the integers is a Markov chain with stationary transition probabilities. For suppose our random walk is given, as in chapter 2, by

$$S_n = \begin{cases} 0 & \text{if } n = 0, \\ X_1 + X_2 + \ldots + X_n & \text{if } n \geq 1, \end{cases}$$

where $X_1, X_2, \dots$ are i.i.d. integer-valued random variables. Then

$$\begin{aligned}\Pr\{S_{n+1}=j|S_n=i,S_{n-1}=i_{n-1},\dots,S_0=0\} &= \Pr\{X_{n+1}=j-i|S_n=i,S_{n-1}=i_{n-1},\dots,S_0=0\}\\ &= \Pr\{X_{n+1}=j-i\} = \Pr\{S_{n+1}=j|S_n=i\}.\end{aligned}$$

The one-step transition probabilities are $P_{ij} = \Pr\{X_1 = j-i\}$.

Assumption. All Markov chains in this chapter will be assumed to have stationary transition probabilities.

Proposition 3.1.2 *For any states i and j, $P\{X_{n+m}=j|X_n=i\}$ does not depend on n.*

Proof. For any integers $m, n, r \geq 0$, we have

$$\begin{aligned}P\{X_{m+n+r}=j|X_n=i\} &= \frac{P\{X_{m+n+r}=j, X_n=i\}}{P\{X_n=i\}} = \frac{\sum_{k\in E} P\{X_{m+n+r}=j, X_{m+n}=k, X_n=i\}}{P\{X_n=i\}}\\ &= \sum_{k\in E} P\{X_{m+n+r}=j|X_{m+n}=k, X_n=i\}\frac{P\{X_{m+n}=k, X_n=i\}}{P\{X_n=i\}}\\ &= \sum_{k\in E} P\{X_{m+n+r}=j|X_{m+n}=k\}P\{X_{m+n}=k|X_n=i\}. \qquad (1.3)\end{aligned}$$

In particular, with $r=1$ this reads as

$$P\{X_{m+n+1}=j|X_n=i\} = \sum_{k\in E} P_{kj} P\{X_{m+n}=k|X_n=i\}.$$

This equation shows that if the m-step transition probabilities do not depend on n, then neither do the $m+1$-step transition probabilities. Since by assumption the 1-step transition probabilities do not depend on n, the rest of the proof follows by induction. ∎

Notation. Because of the previous proposition, it makes sense to define

$$P_{ij}^{(m)} = \Pr\{X_{n+m}=j|X_n=i\} = \Pr\{X_m=j|X_0=i\}, \quad i,j\in E, m\geq 0.$$

$P_{ij}^{(m)}$ is called the *m-step transition probability* of going from i to j. In particular, $P_{ij}^{(1)} = P_{ij}$ and

$$P_{ij}^{(0)} = \begin{cases}1 & \text{if } i=j,\\ 0 & \text{if } i\neq j.\end{cases}$$

Proposition 3.1.3 *For any integers $r, s \geq 0$, we have*

$$P_{ij}^{(r+s)} = \sum_{k\in E} P_{ik}^{(r)} P_{kj}^{(s)}, \quad i,j\in E. \qquad (1.4)$$

This is called the Chapman-Kolmogorov equation.

Proof. This follows by simply rewriting (1.3) with our new notation. ∎

Matrix Notation Since E is countable, it can be put into a 1-1 correspondence with a subset of the set $\{0,1,2,3,\dots\}$. We will therefore assume that E is either $\{0,1,2,\dots\}$ or a subset of $\{0,1,2,\dots\}$. We can then arrange the P_{ij}'s in a matrix

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots\\ P_{10} & P_{11} & P_{12} & \dots\\ P_{20} & P_{21} & P_{22} & \dots\\ \vdots & \vdots & & \end{pmatrix}$$

called the *one-step transition matrix*, or just *transition matrix* of the chain. In addition, the row vector

$$p(n) = \big(\Pr\{X_n = 0\}, \Pr\{X_n = 1\}, \Pr\{X_n = 2\}, \dots\big)$$

is called the *probability distribution vector at time* n, and in particular,

$$p(0) = \big(\Pr\{X_0 = 0\}, \Pr\{X_0 = 1\}, \Pr\{X_0 = 2\}, \dots\big)$$

is called the *initial probability distribution vector*.

Definition. A square matrix A (possibly infinite-dimensional) with non-negative components and all row sums equal to 1 is called a *stochastic matrix*. A row vector, all of whose components are non-negative and which sum to 1 is called a *probability vector*.

Proposition 3.1.4

(1) $p(n)$ is a probability vector.

(2) P is a stochastic matrix.

Proof. We have

$$\sum_{j=0}^{\infty} P_{ij} = \sum_{j=0}^{\infty} \frac{\Pr\{X_{n+1} = j, X_n = i\}}{\Pr\{X_n = i\}} = \frac{\Pr\{X_n = i\}}{\Pr\{X_n = i\}} = 1.$$

∎

Temporarily, let $P^{(m)} = (P_{ij}^{(m)})$ denote the matrix of m-step transition probabilities.

Proposition 3.1.5 $P^{(m)} = P^m$ *for any* $m \geq 0$.

Proof. From the Chapman-Kolmogorov equation (1.4), we have $P^{(r+s)} = P^{(r)} P^{(s)}$ for any integers $r, s \geq 0$. In particular, $P^{(r+1)} = P^{(r)} P$ for all $r \geq 0$. Hence $P^{(m)} = P^{(m-1)} P = P^{(m-2)} P^2 = \dots = P^m$. ∎

Proposition 3.1.6 $p(n+m) = p(n) P^m$ *for any* $n \geq 0$, $m \geq 0$.

Proof.

$$\begin{aligned} k\text{th element of } p(n+m) = \Pr\{X_{n+m} = k\} &= \sum_{i=0}^{\infty} \Pr\{X_{n+m} = k, X_n = i\} \\ &= \sum_{i=0}^{\infty} \Pr\{X_n = i\} \Pr\{X_{n+m} = k | X_n = i\} = \sum_{i=0}^{\infty} \Pr\{X_n = i\} P_{ik}^{(m)} \\ &= k\text{th component of } p(n) P^{(m)} = k\text{th component of } p(n) P^m. \end{aligned}$$

∎

The above two propositions can be used to compute probabilities like

$$\Pr\{X_8 = 4 | X_3 = 5\}, \qquad \Pr\{X_5 = 7\}.$$

For example, $\Pr\{X_8 = 4 | X_3 = 5\}$ is the 5,4th component of P^5, and $\Pr\{X_5 = 7\}$ is the 7th component of the row vector $p(0) P^5$.

We will conclude §1 of this chapter will some more examples of Markov chains, but first we state an important fact which is a result of the famous Kolmogorov-Daniell theorem.

Theorem 3.1.7 *Given a stochastic matrix A and a probability row vector p (of the same dimension), there exists a unique Markov chain with stationary transition probabilities having A as its transition matrix and p as its initial probability distribution vector.*

Thus, in order to specify a Markov chain, we need only specify P and $p(0)$.

Example 1.3. Simple Random Walk on $\{1, 2, 3, 4\}$ with Absorbing Boundaries at 1 and 4.

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$E = \{1, 2, 3, 4\}$ and $p(0)$ can be any probability vector. If for instance the particle starts at 2, then $p(0) = (0, 1, 0, 0)$.

Example 1.4. Simple Random Walk with Reflecting Boundaries at 0 and 5.

$$P = \begin{pmatrix} q & p & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & q & p \end{pmatrix}$$

$E = \{1, 2, 3, 4\}$ and $p(0)$ can be any probability row vector. Note that from state 1, the particle tries to jump to 0 (with probability q), but is immediately reflected back to 1.

Example 1.5. Same but with Elastic Boundary at 0.

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ (1-\delta)q & \delta q & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & q & p \end{pmatrix}$$

Example 1.6. Cyclic Random Walk on $\{1, 2, 3, 4\}$

$$P = \begin{pmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{pmatrix}$$

3.2 Classification of States, Closed Sets, and Irreducibility

Let $\{X_n, n \geq 0\}$ be a Markov chain with transition matrix P and state space E.

Definition. $d(i) = \gcd\{n \geq 1 : P_{ii}^{(n)} > 0\}$ is called the *period* of state i. Equivalently, state i has period d if $P_{ii}^{(n)} = 0$ unless $n = \nu d$ is a multiple of d, and d is the largest integer with this property. If $d(i) = 1$, then i is called aperiodic.

Example 2.1. In the simple random walk with absorbing boundaries at 1 and 4, states 1 and 4 are aperiodic and states 2 and 3 have period 2. In the walk with reflecting boundaries, all states are aperiodic. In the cyclic random walk, all states have period two.

Notation. Let $i, j \in E$ and let

$$T_j = \min\{n \geq 1 : X_n = j\}$$

be the *first hitting time of state* j. Define

$$f_{ij}^{(n)} = \Pr\{T_j = n | X_0 = i\} = \begin{cases} 0 & \text{if } n = 0 \\ \Pr\{X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = i\} & \text{if } n \geq 1, \end{cases}$$

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^{(n)} = \Pr\{T_j < +\infty\} = \Pr\{\text{ eventually hit } j | \text{ start at } i\}$$

$$\mu_i = \sum_{n=0}^{\infty} n f_{ii}^{(n)} = E(T_i | X_0 = i) = \text{ mean recurrence time of state } i.$$

Remark. If $X_0 = i$, then T_i is called the recurrence time of state i.

Definition. State i is *recurrent* (or persistent) if $f_{ii} = 1$, and *transient* if $f_{ii} < 1$. A recurrent state i is called *positive* if $\mu_i < +\infty$, and *null* if $\mu_i = +\infty$. An aperiodic recurrent positive state is called *ergodic*.

Definition. Two states i and j in E are said to be of the same type if they have the same classification. Namely, if

(1) i and j have the same period, and

(2) either

  (a) both i and j are transient, or

  (b) both i and j are recurrent positive, or

  (c) both i and j are recurrent null.

Next, we show that the renewal equation enters in a natural way into the theory of Markov chains.

Proposition 3.2.1 (First Entrance Decomposition of $P_{ij}^{(n)}$) *For any $n \geq 0$, we have*

$$P_{ij}^{(n)} = b_n + \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)}, \tag{2.1}$$

where $b_n = \delta_{0n}\delta_{ij}$ and δ_{ab} denotes the Kronecker delta function

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases}$$

Proof. If $n = 0$, then (2.1) follows because $f_{ij}^{(0)} = 0$. Hence assume $n \geq 1$. We have to show that

$$P_{ij}^{(n)} = \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)}.$$

We have

$$P_{ij}^{(n)} = \Pr\{X_n = j | X_0 = i\} = \sum_{m=0}^{n} \Pr\{X_n = j, T_j = m | X_0 = i\} = \sum_{m=0}^{n} \frac{\Pr\{X_n = j, X_m = j, T_j = m, X_0 = i\}}{\Pr\{X_0 = i\}}$$

$$= \sum_{m=0}^{n} \Pr\{X_n = j | X_m = j, T_j = m, X_0 = i\} \Pr\{T_m = j | X_0 = i\}$$

$$= \sum_{m=0}^{n} \Pr\{X_n = j | X_m = j\} \Pr\{T_j = m | X_0 = i\} = \sum_{m=0}^{n} P_{jj}^{(n-m)} f_{ij}^{(m)}.$$

∎

Remarks.

(1) The fact that $\Pr\{X_n = j | X_m = j, T_j = m, X_0 = i\} = \Pr\{X_n = j | X_m = j\}$, which was used in the above proof, can be established using the Markov property.

(2) Suppose that we define

$$F_{ii}(s) = \sum_{n=0}^{\infty} f_{ii}^{(n)} s^n, \qquad P_{ii}(s) = \sum_{n=0}^{\infty} P_{ii}^{(n)} s^n.$$

Then just as in lemma 2.2 of §2.2, we can show that

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)}.$$

The following proposition gives an alternative definition of a recurrent (or transient) state.

Proposition 3.2.2 *Define $N_{ij} = \sum_{n=0}^{\infty} P_{ij}^{(n)}$. Then*

(1) N_{ij} = expected number of visits to j given the chain starts at i.

(2) $N_{ij} = \delta_{ij} + f_{ij} N_{jj}$. ($\delta_{ij}$ is the Kronecker delta.)

(3) j is recurrent if and only if $N_{jj} = +\infty$.

(4) j is transient if and only if $N_{ij} < +\infty$ for all $i \in E$. In particular, if j is transient, then $P_{ij}^{(n)} \to 0$ as $n \to \infty$.

Proof.

(1) Let

$$Y_n = \begin{cases} 1 & \text{if } X_n = j, \\ 0 & \text{if } X_n \neq j, \end{cases}$$

and put $Y = \sum_{n=0}^{\infty} Y_n$. Then Y is the number of visits to state j, and we have

$$E(Y|X_0 = i) = \sum_{n=0}^{\infty} E(Y_n | X_0 = i) = \sum_{n=0}^{\infty} \Pr\{Y_n = 1 | X_0 = i\} = \sum_{n=0}^{\infty} \Pr\{X_n = j | X_0 = i\} = N_{ij}.$$

(2) If $i \neq j$, then $P_{ij}^{(n)} = \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)}$ (from (2.1)). Hence

$$\begin{aligned} N_{ij} &= \sum_{n=0}^{\infty} \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)} = \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} f_{ij}^{(m)} P_{jj}^{(n-m)} \\ &= \sum_{m=0}^{\infty} f_{ij}^{(m)} \sum_{n=m}^{\infty} P_{jj}^{(n-m)} = \sum_{m=0}^{\infty} f_{ij}^{(m)} N_{jj} \\ &= f_{ij} N_{jj}. \end{aligned}$$

(3) From part (2) with $i = j$, we have $N_{jj} = 1 + f_{jj} N_{jj}$ and so $N_{jj} = +\infty$ if and only if $f_{jj} = 1$.

(4) follows from (2) and (3).

∎

We shall now digress to briefly study the renewal equation and the renewal theorem.

Definition Suppose that $\{f_n, n \geq 0\}$, $\{b_n, n \geq 0\}$, and $\{u_n, n \geq 0\}$ are sequences of numbers. The equation

$$u_n = b_n + \sum_{m=0}^{n} f_m u_{n-m}, \qquad n \geq 0$$

is called the *renewal equation*.

Usually, we want to solve for the u_n's in terms of the f_n's and the b_n's. If $f_0 \neq 1$, there is always a unique solution, derived as follows:

$$\begin{array}{lll} n=0: & u_0 = b_0 + f_0 u_0 & \Rightarrow u_0 = \frac{b_0}{1-f_0} \\ n=1: & u_1 = b_1 + f_0 u_1 + f_1 u_0 & \Rightarrow u_1 = \frac{b_1 + u_0 f_1}{1-f_0} \\ n=2: & u_2 = b_2 + f_0 u_2 + f_1 u_1 + f_2 u_0 & \Rightarrow u_2 = \frac{b_2 + u_0 f_2 + u_1 f_1}{1-f_0} \end{array}$$

and so on. As we have already seen, the renewal equation may be solved by the method of generating functions. Indeed, with the obvious notation, we have

$$U(s) = \frac{B(s)}{1 - F(s)}.$$

Theorem 3.2.3 (The Renewal Theorem) *Suppose that*

(1) $f_0 = 0$, $f_n \geq 0$ *for all* $n \geq 1$, $\sum_{n=0}^{\infty} f_n = 1$.

(2) $\sum_{n=0}^{\infty} |b_n| < +\infty$

Let $d = \gcd\{n \geq 1 : f_n > 0\}$ *denote the "period" of the sequence* $\{f_n, n \geq 0\}$. *Then*

$$\lim_{n\to\infty} u_{nd} = \begin{cases} \frac{d}{\mu} \sum_{n=0}^{\infty} b_{nd} & \text{if } \mu < +\infty, \\ 0 & \text{if } \mu = +\infty, \end{cases}$$

where $\mu = \sum_{n=0}^{\infty} n f_n$.

Remark. It can be shown that if $b_0 = 1$, and $b_n = 0$ for all $n \geq 1$, then

$$\gcd\{n : f_n > 0\} = \gcd\{n : u_n > 0\}.$$

The following corollary follows directly from proposition 2.1, the preceding remark, and the renewal theorem (with $u_n = P_{jj}^{(n)}$, $f_n = f_{jj}^{(n)}$, $b_0 = 1$, and $b_n = 0$ for $n \geq 1$).

Corollary 3.2.4 *Let* j *be a recurrent state with period* d. *Then*

$$P_{jj}^{(nd)} \longrightarrow \begin{cases} \frac{d}{\mu_j} & \text{if } \mu_j < +\infty, \\ 0 & \text{if } \mu_j = +\infty. \end{cases}$$

Proposition 3.2.5 *Suppose* j *is a recurrent state. Then*

$$j \text{ is null if and only if } P_{ij}^{(n)} \to 0 \text{ as } n \to \infty \text{ for all states } i.$$

Proof. Let j be recurrent. By the definition of "null" and by corollary 2.4, we know that

$$j \text{ is null } \iff P_{jj}^{(nd)} \to 0 \text{ as } n \to \infty$$

where d is the period of state j. But $P_{jj}^{(n)} = 0$ unless n is a multiple of d, so

$$j \text{ is null } \iff P_{jj}^{(n)} \to 0 \text{ as } n \to \infty.$$

We now need only show that $P_{jj}^{(n)} \to 0 \implies P_{ij}^{(n)} \to 0$ for $i \neq j$. But we have

$$\lim_{n\to\infty} P_{ij}^{(n)} = \lim_{n\to\infty} \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)} = \sum_{m=0}^{\infty} f_{ij}^{(m)} \lim_{n\to\infty} P_{jj}^{(n-m)} = 0.$$

∎

Problems

(1) Show that if j is not recurrent positive, then $P_{ij}^{(n)} \to 0$ as $n \to \infty$.

(2) Show that if j is ergodic, then

$$P_{ij}^{(n)} \to \frac{f_{ij}}{\mu_j} \quad \text{as } n \to \infty$$

for all states i.

Solution: (1) is obvious. For (2), we have

$$\lim_{n\to\infty} P_{ij}^{(n)} = \lim_{n\to\infty} \sum_{m=0}^{n} f_{ij}^{(m)} P_{jj}^{(n-m)} = \sum_{m=0}^{\infty} f_{ij}^{(m)} \lim_{n\to\infty} P_{jj}^{(n-m)} = \sum_{m=0}^{\infty} f_{ij}^{(m)} \cdot \frac{1}{\mu_j} = \frac{f_{ij}}{\mu_j}.$$

Lemma 3.2.6

(1) $P_{ij}^{(r+s)} \geq P_{ik}^{(r)} P_{kj}^{(s)}$ *for any* $k \in E$.

(2) $P_{ij}^{(n+r+s)} \geq P_{ik}^{(n)} P_{k\ell}^{(r)} P_{\ell j}^{(s)}$ *for any* $k, \ell \in E$.

Proof. Because of the Chapman-Kolmogorov equation, we have

(1) $P_{ij}^{(r+s)} = \sum_{k\in E} P_{ik}^{(r)} P_{kj}^{(s)} \geq P_{ik}^{(r)} P_{kj}^{(s)}$.

(2) $P_{ij}^{(n+r+s)} \geq P_{i\ell}^{(n+r)} P_{\ell j}^{(s)} \geq P_{ik}^{(n)} P_{k\ell}^{(r)} P_{\ell j}^{(s)}$.

∎

Definition. We say that state j *can be reached* from state i, and write $i \hookrightarrow j$, if there is $n \geq 0$ such that $P_{ij}^{(n)} > 0$. If i and j can be reached from each other, we say that i and j *communicate*, and write $i \leftrightarrow j$.

Remarks. Since $P_{ii}^{(0)} = 1$, we certainly have $i \hookrightarrow i$. Part (1) of lemma 2.6 shows that if $i \hookrightarrow k$ and $k \hookrightarrow j$, then $i \hookrightarrow j$. The proof of the following proposition is then immediate.

Proposition 3.2.7 $\leftrightarrow$ *is an equivalence relation on* E*, that is,* $\leftrightarrow$ *is*

(1) reflexive ($i \leftrightarrow i$)

(2) symmetric ($i \leftrightarrow j \implies j \leftrightarrow i$)

(3) transitive (*if* $i \leftrightarrow j$ *and* $j \leftrightarrow k$*, then* $i \leftrightarrow k$).

Consequently, $\leftrightarrow$ *partitions* E *into pairwise disjoint equivalence classes* $C_1, C_2, C_3, \ldots$. *These are called communicating classes.*

Theorem 3.2.8 *Let* i *and* j *be two states which communicate. Then* i *and* j *are of the same type. (In particular, all states in a given communicating class are of the same type.)*

Proof. There are integers r and s so that $\alpha = P_{ij}^{(r)} > 0$ and $\beta = P_{ji}^{(s)} > 0$. By lemma 2.6, we have

$$P_{ii}^{(n+r+s)} \geq P_{ij}^{(r)} P_{jj}^{(n)} P_{ji}^{(s)} = \alpha\beta P_{jj}^{(n)}.$$

We will deduce the theorem from this formula.

(1) *i transient if and only if j transient.* For suppose i is transient. Then

$$N_{jj} = \sum_{n=0}^{\infty} P_{jj}^{(n)} \leq \frac{1}{\alpha\beta} \sum_{n=0}^{\infty} P_{ii}^{(n+r+s)} \leq \frac{N_{ii}}{\alpha\beta} < +\infty,$$

so j is transient. By symmetry, we also have j transient $\Longrightarrow i$ transient.

(2) *i recurrent if and only if j recurrent.* Direct from (1).

(3) *i null if and only if j null.* For suppose i is recurrent null. Then j is recurrent (from 2.) and

$$P_{jj}^{(n)} \leq \frac{1}{\alpha\beta} P_{ii}^{(n+r+s)} \to 0 \text{ as } n \to \infty,$$

so j is null. Again by symmetry, we also have j null $\Longrightarrow i$ null.

(4) *i positive if and only if j positive.* Direct from 3.

(5) *i and j have the same period.* Since $P_{ii}^{(r+s)} > \alpha\beta > 0$, then $r+s$ is a multiple of $d(i)$. Suppose that $n \in \{n | P_{jj}^{(n)} > 0\}$. Then from the above formula, $P_{ii}^{(n+r+s)} > 0$, so that $n+r+s$, and therefore n, is a multiple of $d(i)$. Hence $d(i) \leq d(j)$. Exchanging the roles of i and j shows that $d(j) \leq d(i)$.

∎

Definition. Let $i, j \in E$ be such that $i \neq j$. A finite set $\{i_0, i_1, \ldots, i_n\}$ of states such that $i_0 = i$ and $i_n = j$ is called a *path from i to j* if $P_{i_m i_{m+1}} > 0$ for $m = 0, 1, \ldots, n-1$. That is, a path from i to j must begin at i, end at j, and be such that each state along the path can be reached from the preceding state in one transition.

Proposition 3.2.9 *Let $i, j \in E$ with $i \neq j$. Then j can be reached from i if and only if there is a path leading from i to j.*

Proof. By repeated use of the Chapman-Kolmogorov equation, we have

$$P_{ij}^{(n)} = \sum_{k_1 \in E} \sum_{k_2 \in E} \cdots \sum_{k_{n-1} \in E} P_{ik_1} P_{k_1 k_2} \cdots P_{k_{n-1} j}.$$

If there is a path leading from i to j, say of length n, then one of the terms in the sum here is strictly positive, and then so must be $P_{ij}^{(n)}$. Conversely, if j can be reached from i, then for some $n \geq 1$, we have $P_{ij}^{(n)} > 0$, and then one of the terms in the above sum must be strictly positive, so there must exist a path from i to j. ∎

Definition. A set C of states is *closed* if no state outside of C can be reached from a state in C (i.e. if $P_{ij}^{(n)} = 0$ for all $n \geq 0$ whenever $i \in C$ and $j \in E \setminus C$). If a single state i forms a closed set, it is called *absorbing*. A closed set C is called *irreducible* if it contains no smaller closed set. If the whole state space E is irreducible, the Markov chain is said to be irreducible.

Proposition 3.2.10 *Let $C \subset E$. Then*

$$C \text{ is closed } \Longleftrightarrow P_{ij} = 0 \text{ for all } i \in C, j \notin C.$$

Proof. $\Longrightarrow$ is trivial from the definition. For $\Longleftarrow$, we have, if $i \in C$ and $j \notin C$,

$$P_{ij}^{(2)} = \sum_{k \in C} P_{ik} \underbrace{P_{kj}}_{=0} + \sum_{k \notin C} \underbrace{P_{ik}}_{=0} P_{kj} = 0,$$

$$P_{ij}^{(3)} = \sum_{k \in C} P_{ik}^{(2)} \underbrace{P_{kj}}_{=0} + \sum_{k \notin C} \underbrace{P_{ik}^{(2)}}_{=0} P_{kj} = 0,$$

and so on. Hence $P_{ij}^{(n)} = 0$ for all $n \geq 0$. ∎

Problem. Let $C(i) = \{j \in E | j \text{ can be reached from } i\}$. Show that $C(i)$ is a closed set.

Solution. Hint: suppose $C(i)$ is not closed.

Proposition 3.2.11 *(Alternative definition of "irreducible") Suppose C is a closed set. Then*

$$C \text{ is irreducible } \iff \text{ every state in } C \text{ can be reached from every other state in } C.$$

Proof.
$\Longrightarrow$ Let $i, j \in C$. Then $C(i) \subset C, C(j) \subset C$, and so $C(i) = C(j) = C$ (since $C(i)$ and $C(j)$ are closed). Hence $i \hookrightarrow j$ and $j \hookrightarrow i$.
$\Longleftarrow$ Suppose C is *not* irreducible. Then there is a smaller closed set $C' \subset C$. Let $j \in C', k \in C \setminus C'$. Then k cannot be reached from j. Hence not every state in C can be reached from every other state in C. ∎

Remarks. The following statements are immediate consequences of proposition 2.10.

(1) A Markov chain is irreducible if and only if the state space E forms a single communicating class.

(2) A closed class is the same as an irreducible closed set.

Proposition 3.2.12 *Let i be a recurrent state and suppose that $i \hookrightarrow j$. Then $f_{ji} = 1$ and so $i \leftrightarrow j$.*

Proof. Since we can reach j from i, we can certainly reach j from i without returning to i in any intermediate step. Let α be the probability of this latter event. Once j is reached, the probability of never returning to i is $1 - f_{ji}$. Hence

$$\begin{aligned}
0 = 1 - f_{ii} &= \Pr\{\text{chain never returns to } i | X_0 = i\} \\
&\geq \Pr\{\overbrace{\text{chain reaches } j \text{ without returning to } i \text{ in an intermediate step,}}^{A} \\
&\overbrace{\text{then never returns to } i}^{B} \,|\, \overbrace{X_0 = i}^{C}\} \\
&= \Pr\{A \cap B | C\} = \Pr\{B | A \cap C\} \Pr\{A | C\} = (1 - f_{ji})\alpha \geq 0,
\end{aligned}$$

so that $(1 - f_{ji})\alpha = 0$. Since $\alpha > 0$, then $f_{ji} = 1$. ∎

Corollary 3.2.13 *Let C be a recurrent communicating class. Then*

(1) $C = C(i)$ for any $i \in C$.

(2) C is closed.

Proof.

(1) Let $i \in C$. We always have $C \subset C(i)$. Conversely, if $j \in C(i)$, then $i \leftrightarrow j$ by the previous result and so $j \in C$.

(2) We showed above that $C(i)$ is always closed. Hence this part follows from part (1).

∎

Proposition 3.2.14 *In a finite Markov chain,*

(1) not all states can be transient, and

(2) no states are null.

Proof.

(1) Suppose all states are transient. Then for all j, k we have $P_{jk}^{(n)} \to 0$ as $n \to \infty$, so $\sum_{k \in E} P_{jk}^{(n)} \to 0$ as $n \to \infty$. This contradicts the fact that $\sum_{k \in E} P_{jk}^{(n)} = 1$ (the jth row sum) for all n.

(2) Suppose at least one of the recurrent states, say i, is null, and let C be the class containing i. Then every state k in C is null, so $P_{ik}^{(n)} \to 0$ as $n \to \infty$ for all $k \in C$. Hence $\sum_{k \in C} P_{ik}^{(n)} \to 0$ as $n \to \infty$. This contradicts the fact that $\sum_{k \in C} P_{ik}^{(n)} = 1$ for all n, since by the previous corollary, no escape is possible from C.

∎

Summary. Given a Markov chain with state space E, we can partition E into communicating classes C_1, C_2, $C_3, \ldots$. Moreover,

(1) all states in C_1 are of the same type, all states in C_2 are of the same type, etc.

(2) classes which are not closed consist of transient states.

(3) if a closed class is finite, then every state in it is recurrent positive.

Example. Suppose we have a Markov chain with state space $E = \{1, 2, 3, \ldots, 9\}$, whose 9×9 transition matrix is

$$P = \begin{pmatrix} 0 & 0 & 0 & .3 & 0 & 0 & 0 & 0 & .7 \\ 0 & .1 & .3 & 0 & .4 & 0 & 0 & 0 & .2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .6 & 0 & 0 & 0 & .2 & .2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Find all communicating classes and classify all states.

Solution The state diagram is

so the communicating classes are $C_1 = \{1, 4, 9\}$, $C_2 = \{3, 8\}$, $C_3 = \{5\}$, $C_4 = \{2\}$, $C_5 = \{6\}$, and $C_6 = \{7\}$. The first three classes are closed and the last three are not. Thus states $1, 4, 5, 9$ are recurrent positive and aperiodic (i.e. ergodic), and in particular 5 is absorbing; states $3, 8$ are recurrent positive with period 2; and states $2, 6, 7$ are transient.

3.3 Invariant Distributions and Ergodic Behaviour of Markov Chains

Let $X = \{X_n, n \geq 0\}$ be a Markov chain with state space E and transition matrix P.

Definition. A vector $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ with non-negative components such that $\pi P = \pi$ is called an *invariant measure* for the chain. If furthermore $\sum_{i \in E} \pi_i = 1$ (i.e. π is a probability vector), then π is called an invariant *distribution* (or stationary distribution).

Definition. A probability row vector $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ such that $p(0)P^n \to \pi$ componentwise for all initial distributions $p(0)$ is called a *long run distribution* (or sometimes a *steady state* distribution for the chain.

Definition. A Markov chain all of whose states are ergodic is called an ergodic chain.

Theorem 3.3.1 (Ergodic Theorem for Markov Chains) *Suppose $X = \{X_n, n \geq 0\}$ is irreducible.*

(1) Assume X is recurrent positive. Then

(a) X has a unique invariant distribution $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ and $\pi_k = 1/\mu_k > 0$ for every $k \in E$.

(b) if X is aperiodic, if $p(0)$ is any initial probability distribution vector, and if $p(n) = p(0)P^n$ denotes the corresponding probability distribution vector at time n, then $p(n) \to \pi$ (componentwise) as $n \to \infty$. That is, π is a long run distribution for the chain.

(2) Assume there exists an invariant distribution π. Then

(a) the chain is recurrent positive.

(b) the results of part (1) hold.

Proof. We shall assume throughout the proof that the chain X is aperiodic. The extension of our results to the periodic case can be found in the book by Feller. All the exchanges of $\lim_{n\to\infty}$ and $\sum_{k \in E}$ below can be justified.

(1) Define $\pi_k = 1/\mu_k,\ k \in E$. Note that $\pi_k > 0$ for all $k \in E$, and $\lim_{n\to\infty} P_{jk}^{(n)} = \pi_k$ for all $j, k \in E$. Now $P_{ik}^{(n+1)} = \sum_{j \in E} P_{ij}^{(n)} P_{jk}$, so taking limits on each side as $n \to \infty$ gives

$$\pi_k = \lim_{n\to\infty} \sum_{j \in E} P_{ij}^{(n)} P_{jk} = \sum_{j \in E} \pi_j P_{jk}$$

for all k. This means $\pi = \pi P$, so π is an invariant measure. Since

$$\sum_{k \in E} \pi_k = \sum_{k \in E} \lim_{n\to\infty} P_{jk}^{(n)} = \lim_{n\to\infty} \sum_{k \in E} P_{jk}^{(n)} = 1,$$

then π is an invariant distribution. Now let $p(0)$ be the initial probability distribution vector, and let $p(n) = p(0)P^n$. In terms of components, this reads $p_k^{(n)} = \sum_{j \in E} p_j^{(0)} P_{jk}^{(n)}$. Taking limits on each side gives

$$\lim_{n\to\infty} p_k^{(n)} = \sum_{j \in E} p_j^{(0)} \lim_{n\to\infty} P_{jk}^{(n)} = \sum_{j \in E} p_j^{(0)} \pi_k = \pi_k \sum_{j \in E} p_j^{(0)} = \pi_k.$$

Hence $p(n) \to \pi$ componentwise. To show uniqueness, let v be another invariant distribution. Then

$$v = vP = vP^2 = vP^3 = \dots = vP^n \to \pi \quad \text{as } n \to \infty,$$

so $v = \pi$.

(2) Let π be the invariant distribution. Then $\pi = \pi P^n$ for all n, which we write as $\pi_k = \sum_{i \in E} \pi_i P_{ik}^{(n)}$. Taking limits on each side gives $\pi_k = \sum_{i \in E} \pi_i \lim_{n\to\infty} P_{ik}^{(n)}$. Suppose the chain is *not* recurrent positive. Then $\lim_{n\to\infty} P_{ik}^{(n)} = 0$ for all $k, i \in E$. So $\pi_k = 0$ for all k, contradicting the fact that $\sum_{k \in E} \pi_k = 1$.

Remarks.

(1) The above proof shows that an invariant distribution will *always* exist for a finite Markov chain (because there are no null states, and not all states can be transient, so we cannot have $\lim_{n\to\infty} P_{jk}^{(n)} = 0$ for all $j, k \in E$). This result can be proved easily without using the renewal theorem. For let P be an $n \times n$ transition matrix. Since the row sums of $P - I$ are 0, then the columns of $P - I$ are dependent, so column rank$(P - I) < n$. Since row rank=column rank, then

$$\text{column rank}(P - I)^T = \text{row rank}(P - I) = \text{column rank}(P - I) < n,$$

where ^{T} denotes transpose. Hence there is a row vector $\pi \neq 0$ such that $(P-I)^T \pi^T = 0$, or equivalently $\pi(P-I) = 0$; that is, $\pi P = \pi$. We leave it to the reader to show that π can be taken to be a probability vector.

(2) An invariant distribution need not exist for an infinite Markov chain. (This will happen if all states are transient or recurrent null.)

(3) Even if an invariant distribution exists, it need not be unique. For consider $P = I$ (the identity matrix) in which *every* probability row vector is an invariant distribution. Note that I is recurrent positive and aperiodic, but not irreducible. Reducibility of E into two or more recurrent irreducible closed sets *always* destroys uniqueness.

(4) Even if a unique invariant distribution exists, it need not be a long run distribution. This happens when the chain is periodic. For example, take

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Here, the invariant distribution is $\pi = (.5, .5)$, but if $p(0) = (\alpha, \beta)$, then $p(1) = (\beta, \alpha)$, $p(2) = (\alpha, \beta)$, $p(3) = (\beta, \alpha)$, and so on.

3.4 Examples

3.4.1 Random Walk with Reflecting Boundaries at 0 and a

Let us consider the case $a = 5$. Then the actual state space is $E = \{1, 2, 3, 4\}$ and the transition matrix is

$$P = \begin{pmatrix} q & p & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & q & p \end{pmatrix}.$$

By drawing the state diagram, we see easily that the chain is irreducible and ergodic. By the ergodic theorem, there exists a unique invariant distribution $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$, and π is also the long run distribution. Let us find π. We have

$$(\pi_1, \pi_2, \pi_3, \pi_4) \begin{pmatrix} q & p & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & q & p \end{pmatrix} = (\pi_1, \pi_2, \pi_3, \pi_4), \tag{4.1}$$

which gives

$$\pi_2 = \frac{p}{q} \cdot \pi_1, \quad \pi_3 = \left(\frac{p}{q}\right)^2 \cdot \pi_1, \quad \pi_4 = \left(\frac{p}{q}\right)^3 \cdot \pi_1. \tag{4.2}$$

Note that (4.1) gives four equations, but one is redundant. This will always happen. What allows us to finish our computation is that $(\pi_1, \pi_2, \pi_3, \pi_4)$ is to be an invariant *distribution*. Thus

$$1 = \pi_1 + \pi_2 + \pi_3 + \pi_4 = \pi_1 + \frac{p}{q} \cdot \pi_1 + \left(\frac{p}{q}\right)^2 \cdot \pi_1 + \left(\frac{p}{q}\right)^3 \cdot \pi_1,$$

so that

$$\pi_1 = \frac{1}{1 + (p/q) + (p/q)^2 + (p/q)^3}. \tag{4.3}$$

Hence (4.2) and (4.3) give

$$\pi = \frac{1}{1 + (p/q) + (p/q)^2 + (p/q)^3} \cdot \left(1, \frac{p}{q}, \left(\frac{p}{q}\right)^2, \left(\frac{p}{q}\right)^3\right).$$

For example,

(1) if $p = q$, then $\pi = (1/4, 1/4, 1/4, 1/4)$.

(2) if $p/q = 2$, then $\pi = (1/15, 2/15, 4/15, 8/15)$.

The ergodic theorem tells us that no matter how the chain starts (i.e. no matter what $p(0)$ is), the probability distribution vector $p(n)$ at time n settles down to π as $n \to \infty$. That is, π is the long run distribution. Thus, for instance, if $p/q = 2$, and the chain has been running for a long time, the random walk particle spends about 1/15 of its time in state 1, 2/15 of its time in state 2, 4/15 in 3, and 8/15 in 4.

Also, since the mean recurrence time of a state i is

$$\mu_i = \frac{1}{\pi_i},$$

then, for example, we note that if $p/q = 2$, then $\mu_2 = 7.5$. This means that if the chain is in state 2 at some time, it will require on the average 7.5 more transitions until it returns to state 2.

3.4.2 A Non-irreducible Chain

Suppose we have a Markov chain with state space $E = \{1, 2, 3, \dots, 9\}$, whose 9×9 transition matrix is as given in the example at the end of §3.2, namely

$$P = \begin{pmatrix} 0 & 0 & 0 & .3 & 0 & 0 & 0 & 0 & .7 \\ 0 & .1 & .3 & 0 & .4 & 0 & 0 & 0 & .2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .6 & 0 & 0 & 0 & .2 & .2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We will discuss the long-run behaviour of this chain. Note that this is the same chain that was dealt with at the end of section 3.2. States 2, 6, and 7 are transient. $C_1 = \{1, 4, 9\}$ is a closed ergodic class and $C_2 = \{3, 8\}$ is a closed class with period 2. 5 is an absorbing state. The chain restricted to C_1 has transition matrix

$$\begin{pmatrix} 0 & .3 & .7 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and therefore invariant distribution (.303, .303, .394) (correct to 3 decimal places). The chain restricted to C_2 has transition matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and therefore invariant distribution (.5, .5). If the chain starts out, or at some point in time enters, C_1, then after a long time, it will be in states $1, 4, 9$ with probabilities .303, .303, .394 respectively. If the chain starts out, or at some point in time enters, C_2, it will alternate forever between states 3 and 8. If the chain starts out, or at some point enters state 5, it will stay there forever. If the chain starts out in states $2, 6$, or 7, it will eventually be absorbed into one of C_1, C_2, or state 5, after a finite number of transitions.

3.4.3 The Ehrenfest Model of Gaseous Diffusion

Consider two containers connected by a valve, and containing ρ molecules of a gas between them. The valve is opened at time zero, and the molecules begin to diffuse from one container to the other. What will be the state of the system after a large interval of time has elapsed?

We shall attempt to model the situation as follows: we have two urns A and B containing ρ molecules between them. At each instant $n = 1, 2, 3, \ldots$, a molecule is chosen at random, and moved from its urn into the other (note that the *molecule* is chosen at random, not the urn). Let

$$X_n = \text{ number of molecules in urn A at time } n, n = 0, 1, 2, \ldots$$

where X_0 is the initial number of molecules in urn A. It is easy to see that $\{X_n, n \geq 0\}$ is a Markov chain with state space $E = \{0, 1, 2, \ldots, \rho\}$, and

$$\Pr\{X_{n+1} = k + 1 | X_n = k\} = 1 - k/\rho \quad \text{if } k = 0, 1, 2, \ldots, \rho - 1,$$

$$\Pr\{X_{n+1} = k - 1 | X_n = k\} = k/\rho \quad \text{if } k = 1, 2, \ldots, \rho.$$

All other transitions have probability zero. The $(\rho + 1) \times (\rho + 1)$ transition matrix is

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1/\rho & 0 & 1 - 1/\rho & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2/\rho & 0 & 1 - 2/\rho & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & & & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 - 1/\rho & 0 & 1/\rho \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

By drawing the state diagram, you will see that the chain is irreducible, that all states are recurrent positive, but that the chain is periodic with period 2. By the ergodic theorem, there will be a unique invariant distribution π with $\pi_k = 1/\mu_k$ (the reciprocal of the mean recurrence time of state k), but π will not be the long run distribution. Let us find π. From $\pi = \pi P$, we find

$$\begin{aligned} \pi_0 &= \frac{\pi_1}{\rho}, \\ \pi_1 &= \pi_0 + \frac{2}{\rho}\pi_2, \\ &\vdots \\ \pi_k &= \big(1 - \frac{k-1}{\rho}\big)\pi_{k-1} + \frac{k+1}{\rho}\pi_{k+1}, \quad k = 1, 2, \ldots, \rho - 1 \\ &\vdots \\ \pi_{\rho-1} &= \frac{2}{\rho}\pi_{\rho-2} + \pi_\rho \\ \pi_\rho &= \frac{\pi_{\rho-1}}{\rho}. \end{aligned}$$

Solving recursively, we find that

$$\pi_1 = \rho u_0, \quad \pi_2 = \frac{\rho(\rho-1)}{2},$$

and in general

$$\pi_k = \binom{\rho}{k}\pi_0.$$

We find π_0 from the fact that $\sum_{k=0}^{\rho} \pi_k = 1$. We have

$$1 = \sum_{k=0}^{\rho} \pi_k = \pi_0 \sum_{k=0}^{\rho} \binom{\rho}{k} = \pi_0 2^\rho,$$

so that $\pi_0 = 2^{-\rho}$, and

$$\boxed{\pi_k = \binom{\rho}{k}(\frac{1}{2})^\rho, \quad k = 0, 1, 2, \ldots, \rho.}$$

The mean recurrence times are given by

$$\mu_k = \frac{2^\rho}{\binom{\rho}{k}}, \quad k = 0, 1, 2, \ldots, \rho.$$

Even though a single overall long run distribution does not exist, we can still investigate the long run behaviour of the chain. From the state diagram, it is apparent that if we consider the random variables $X_0, X_2, X_4, \ldots$, we have an aperiodic Markov chain, for which all states are recurrent positive. However, this chain is not irreducible, but consists of two irreducible closed sets. For example, suppose ρ is odd. Then the two irreducible closed sets are $C_1 = \{0, 2, 4, \ldots, \rho - 1\}$ and $C_2 = \{1, 3, \ldots, \rho\}$. If we look at the chain $X_0, X_2, X_4, \ldots$ only on C_1 (or only on C_2), we have an irreducible ergodic chain, for which the invariant distribution will also be the long run distribution. Similar remarks apply to the remaining random variables $X_1, X_3, X_5, \ldots$. In this way we can determine the long-run behaviour of the chain, but the situation is not so clear cut.

Problem. Show that

(1)

$$E(X_n) = \frac{\rho}{2}\big[1 - (1 - \frac{2}{\rho})^n\big] + (1 - \frac{2}{\rho})^n E(X_0)$$

(2)

$$E(X_n) \to \frac{\rho}{2} \text{ as } n \to \infty$$

.

Solution. Part (ii) follows directly from (i). To prove (i), we write

$$E(X_{n+1}|X_n = i) = \begin{cases} 1 & \text{if } i = 0, \\ (i+1)(1 - \frac{i}{\rho}) + (i-1)\frac{i}{\rho} = 1 + (1 - \frac{2}{\rho})i & \text{if } 1 \le i \le \rho - 1, \\ \rho - 1 & \text{if } i = \rho. \end{cases}$$

By the law of total expectation, we get, for $n \ge 0$,

$$\begin{aligned} E(X_{n+1}) &= \sum_{i=0}^{\rho} E(X_{n+1}|X_n = i)\Pr\{X_n = i\} \\ &= \sum_{i=0}^{\rho-1}[1 + (1 - \frac{2}{\rho})i]\Pr\{X_n = i\} + (\rho - 1)\Pr\{X_n = \rho\} = 1 + (1 - \frac{2}{\rho})E(X_n). \end{aligned}$$

This is a difference equation of the form $a_{n+1} = 1 + ca_n$, $n \ge 0$ which can be recursively solved to find

$$a_n = \frac{1 - c^n}{1 - c} + c^n a_0, \quad n \ge 0.$$

3.5 Additional Topics

3.5.1 Foster's Theorem and the Classification of Infinite Chains

We have seen in §3.2 that it is a very simple matter to classify the states of a finite Markov chain. When the state space is infinite, though, the problem is a little harder.

Let us suppose that $X = \{X_n, n \geq 0\}$ is an irreducible Markov chain with state space E and transition matrix P. By the Ergodic Theorem (theorem 3.1), X is recurrent positive if and only if there is an invariant distribution for P. Thus we have an applicable criterion to decide whether X is recurrent positive or not. But, if it is not, how do we decide whether X is recurrent null or transient? The answer is via Foster's Theorem.

Theorem 3.5.1 (Foster's Theorem) *Let i_0 be any state in E, and let Q be the matrix derived from P by deleting the row and column corresponding to the state i_0. Then the chain X is recurrent if and only if the only column vector h which is bounded and non-negative (i.e. $0 \leq h_i \leq M$ for all i and some finite number M) satisfying the equation $Qh = h$ is the zero vector.*

A proof of Foster's Theorem may be found in Karlin and Taylor (1975). More important to us is how to apply it. Given an irreducible chain X, we should first test P using Foster's Theorem to see whether X is recurrent or not. If X is not recurrent (i.e is transient), we are finished. If X is recurrent, we test P using the Ergodic Theorem to see if X is recurrent positive or not.

Example. Consider the Markov chain with state space $\{0, 1, 2, \dots\}$ and irreducible transition matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & \cdots \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Taking the state i_0 to be 0, the equation $Qh = h$ becomes

$$\begin{pmatrix} 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 & \cdots \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \end{pmatrix}.$$

Carrying out the matrix operation, we obtain

$$\begin{aligned} \frac{2}{3}h_2 &= h_1 \\ \frac{1}{3}h_1 + \frac{2}{3}h_3 &= h_2 \\ \frac{1}{3}h_2 + \frac{2}{3}h_4 &= h_3 \\ &\vdots \end{aligned}$$

or more briefly

$$\frac{1}{3}h_n + \frac{2}{3}h_{n+2} = h_{n+1}, \quad n \geq 0,$$

provided we interpret $h_0 = 0$. Multiply this equation by 3/2 and subtract h_{n+1} from both sides, getting

$$h_{n+2} - h_{n+1} = \frac{1}{2}[h_{n+1} - h_n], \quad n \geq 0.$$

Thus with $n = 0$, we have

$$h_2 - h_1 = \frac{1}{2}[h_1 - h_0],$$

and for $n = 1$, we get

$$h_3 - h_2 = \frac{1}{2}[h_2 - h_1] = (\frac{1}{2})^2[h_1 - h_0].$$

Repeating, we find

$$h_{n+1} - h_n = (\frac{1}{2})^n[h_1 - h_0] = (\frac{1}{2})^n h_1, \quad n \geq 0.$$

Summing both sides of this equality from $n = 0$ to $n = N$, we get

$$h_{N+1} = h_1 \sum_{n=0}^{N} (\frac{1}{2})^n = 2h_1[1 - (\frac{1}{2})^{N+1}].$$

We now take h_1 to be any strictly positive number, say $h_1 = 1$. Then this formula defines a bounded non-zero, non-negative solution of the equation $Qh = h$, and so by the above theorem, the chain must be transient.

3.5.2 Another Limit Theorem

First, we need the following lemma.

Lemma 3.5.2 *Let $\{a_n, n \geq 1\}$ be a sequence of numbers.*

(1) If $a_n \to a$ as $n \to \infty$, then also

$$\frac{1}{n} \sum_{m=1}^{n} a_m \to a \qquad \text{as } n \to \infty.$$

(2) Suppose there are integers d and r with $0 \leq r < d$ such that

(a) $a_n = 0$ unless $n = kd + r$ for some $k \geq 1$, and

(b) $a_{kd+r} \to a$ as $k \to \infty$.

Then

$$\frac{1}{n} \sum_{m=1}^{n} a_m \to \frac{a}{d} \qquad \text{as } \to \infty.$$

Proof. Part (1) is standard. Let us turn to part (2), which is only a slight generalization of part (1). Let $k^*(n) = \max\{k : kd + r \leq n\}$. Then

$$\frac{1}{n} \sum_{m=1}^{n} a_m = \frac{k^*(n)}{n} \cdot \frac{1}{k^*(n)} \sum_{k=1}^{k^*(n)} a_{kd+r} \to \frac{a}{d}$$

since

$$\frac{k^*(n)}{n} \to \frac{1}{d} \quad \text{and} \quad \frac{1}{k^*(n)} \sum_{k=1}^{k^*(n)} a_{kd+r} \to a$$

as $n \to \infty$. ∎

Theorem 3.5.3 *Suppose $i, j \in E$ are such that i and j communicate. Then*

$$\frac{1}{n} \sum_{m=1}^{n} P_{ij}^{(m)} \to \frac{1}{\mu_j} \text{ as } n \to \infty.$$

Proof. If the class is transient or recurrent null, then $P_{ij}^{(n)} \to 0$ as $n \to \infty$, so the result follows from part (1). There remains the case where the class is recurrent positive with period d. Then there is an integer r with $0 \leq r < d$ such that $P_{ij}^{(kd+r)} \to d/\mu_j$ as $k \to \infty$, and so this case follows from part (2) of the lemma. ∎

Remark. Since

$$\sum_{m=1}^{n} P_{ij}^{(m)} = E[\sum_{m=1}^{n} I_{\{X_m=j\}}|X_0 = i],$$

then $(1/n)\sum_{m=1}^{n} P_{ij}^{(m)}$ is the proportion of time during the first n transitions that the chain spends in state j.

3.5.3 Markov Chain Arising from a Simple Dynamical System

Let E be a finite set and f a function from E into E. Define a transition matrix P on the state space E by

$$P_{ij} = \begin{cases} 1 & \text{if } j = f(i), \\ 0 & \text{otherwise.} \end{cases}$$

Note that the random variables of the corresponding Markov chain satisfy $X_{n+1}(\omega) = f(X_n(\omega))$ for all $n \geq 0$.

(1) Under what conditions on f is the resulting Markov chain irreducible?

(2) Assuming the chain is irreducible, find the invariant distribution. Can it be also the long-run distribution?

Solution

(1) Since $f(E) \subset E$, then $f(f(E)) \subset f(E) \subset E$, and so $f \circ \cdots \circ f(E) \subset E$, where $\circ$ denotes composition of functions. It follows that if f is not onto, then states in $E \setminus f(E)$ cannot be reached from those in $f(E)$. Hence we must assume f is onto, which implies that f must also be one-to-one (since E is finite). A one-to-one onto function such as f is called a permutation on E. Moreover, for irreducibility, we demand that the permutation f be a cycle (since otherwise it could be resolved into sub-cycles, implying reducibility). Hence the answer is that the chain will be irreducible if and only if f is a cyclic permutation.

(2) The resulting chain will be periodic with period equal to the cardinality N of E, and the transition matrix P will be doubly stochastic (since f is one-to-one). The ergodic theorem implies that there must exist a unique invariant distribution, which because of double stochasticity, is the uniform distribution. There is no nice long-run behaviour because of periodicity. However, we do have

$$\frac{1}{n+1}\sum_{m=0}^{n} P_{ij}^{(m)} \to \frac{1}{N} \text{ as } n \to \infty.$$

That is, given that the chain starts in state i, the proportion of time spent in state j is $1/N$.

Chapter 4

Birth-and-Death Processes

In this chapter, we shall examine stochastic processes $\{X(t), t \geq 0\}$ with state space $E = \{0, 1, 2, 3, \dots\}$, for which t is a *continuous* time parameter.

4.1 Some General Definitions

Definition. $\{X(t), t \geq 0\}$ has *independent increments* if given any times $0 \leq t_0 < t_1 < t_2 < \dots < t_n$, the random variables

$$X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

are independent. Furthermore, if for any $t \geq 0$ and $h \geq 0$, the distribution of the increment $X_{t+h} - X_t$ depends only on h, we say $\{X(t), t \geq 0\}$ has *stationary* independent increments.

Definition. $\{X(t), t \geq 0\}$ is a *continuous time parameter Markov chain* if given any times $0 \leq t_0 < t_1 < t_2 < \dots < t_n$, the random variables $X_{t_0}, X_{t_1}, , \dots, X_{t_n}$ form a discrete-time Markov chain (as studied in chapter 3). Or equivalently if given times $0 \leq t_0 < t_1 < t_2 < \dots < t_n$ and states $i_0, i_1, \dots, i_{n-2}, i, j$, we have

$$\Pr\{X_{t_n} = j | X_{t_{n-1}} = i, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1} = i_1, X_{t_0} = i_0\} = \Pr\{X_{t_n} = j | X_{t_{n-1}} = i\}.$$

The conditional probability on the right-hand side is of the form $\Pr\{X_{t+h} = j | X_t = i\}$, and is called the transition function. We say $\{X(t), t \geq 0\}$ has stationary transition probabilities if $\Pr\{X_{t+h} = j | X_t = i\}$ does not depend on t.

Problem 1.1. Show that if $\{X(t), t \geq 0\}$ has independent increments, it is a Markov chain. If furthermore, $\{X(t), t \geq 0\}$ has *stationary* independent increments, then $\{X(t), t \geq 0\}$ is a Markov chain with stationary transition probabilities.

Solution. Suppose $\{X(t), t \geq 0\}$ has independent increments. Then

$$\begin{aligned}
&\Pr\{X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1} = i_1, X_{t_0} = i_0\} \\
&\quad = \Pr\{X_{t_{n-1}} + \left(X_{t_n} - X_{t_{n-1}}\right) = i_n | X_{t_{n-1}} = i_{n-1}, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1} = i_1, X_{t_0} = i_0\} \\
&\quad = \Pr\{X_{t_n} - X_{t_{n-1}} = i_n - i_{n-1} | X_{t_{n-1}} = i_{n-1}, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1} = i_1, X_{t_0} = i_0\} \\
&\quad = \Pr\{X_{t_n} - X_{t_{n-1}} = i_n - i_{n-1} | X_{t_{n-1}} - X_{t_{n-2}} = i_{n-1} - i_{n-2}, \dots, X_{t_1} - X_{t_0} = i_1 - i_0, X_{t_0} = i_0\} \\
&\quad = \Pr\{X_{t_n} - X_{t_{n-1}} = i_n - i_{n-1}\}.
\end{aligned}$$

Similarily,

$$\Pr\{X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}\} = \Pr\{X_{t_n} - X_{t_{n-1}} = i_n - i_{n-1}\}. \tag{1.1}$$

Hence $\{X(t), t \geq 0\}$ is a Markov chain, and by (1.1) obviously has stationary transition probabilities if $\{X(t), t \geq 0\}$ has stationary increments.

Definition. A stochastic process $\{X(t), t \geq 0\}$ satisfying

(1) $X_0 = 0$

(2) $X_s \leq X_t$ if $s \leq t$

is called a *counting process.*

Example. Suppose you stand at a certain point along a highway, and define X_t = the number of cars which pass during the interval $(0, t]$. The process $\{X(t), t \geq 0\}$ is a counting process counting occurences of the event "car passes".

4.2 The Poisson Process

Definition 1. A counting process $\{X(t), t \geq 0\}$ is called a *Poisson process with rate* $\lambda > 0$ if

(1) $\{X(t), t \geq 0\}$ has independent increments

(2) for all $s, t \geq 0$,

$$\Pr\{X_{s+t} - X_s = i\} = \frac{(\lambda t)^i e^{-\lambda t}}{i!}, \quad i = 0, 1, 2, \ldots$$

(Note that $\{X(t), t \geq 0\}$ therefore has *stationary* independent increments)

We will give an equivalent definition below. First we need the following notation.

Notation. A function $f(h)$ is said to be $o(h)$ if $\lim_{h\to 0} \frac{f(h)}{h} = 0$.

Example.

$$f(h) = 5h^2 \text{ is o(h)}$$
$$f(h) = h^{1/2} \text{ is not o(h)}$$
$$f(h) = \sin h \text{ is not o(h)}$$

Definition 2. A continuous time Markov chain $\{X(t), t \geq 0\}$ with stationary transition probabilities is called a *Poisson process with rate* $\lambda > 0$ if

(1) $X_0 = 0$ and transitions from a state i are allowed only to $i + 1$

(2)

$$\Pr\{X_{t+h} = n + 1 | X_t = n\} = \lambda h + o(h)$$
$$\Pr\{X_{t+h} \geq n + 2 | X_t = n\} = o(h).$$

We are now going to show that these two definitions of a Poisson process are in fact equivalent.

Definition 1. $\Longrightarrow$ **Definition 2.** The proof that only transitions from i to $i + 1$ are possible is quite difficult and will not be done. Because of a previous problem result, there remains only to show that part (2) of definition 2 holds. Put

$$\begin{aligned} f(h) &= \Pr\{X_{t+h} = n + 1 | X_t = n\} - \lambda h = \Pr\{X_{t+h} - X_t = 1 | X_t = n\} - \lambda h \\ &= \Pr\{X_{t+h} - X_t = 1\} - \lambda h = \lambda h e^{-\lambda h} - \lambda h, \\ g(h) &= \Pr\{X_{t+h} \geq n + 2 | X_t = n\} = \Pr\{X_{t+h} - X_t \geq 2 | X_t = n\} \\ &= \Pr\{X_{t+h} - X_t \geq 2\} = 1 - e^{-\lambda h} - \lambda h e^{-\lambda h}. \end{aligned}$$

Then

$$\lim_{h\to 0}\frac{f(h)}{h} = \lambda \lim_{h\to 0}\left(e^{-\lambda h}-1\right) = 0$$
$$\lim_{h\to 0}\frac{g(h)}{h} = \lim_{h\to 0}\frac{1-e^{-\lambda h}-\lambda h e^{-\lambda h}}{h} = 0$$

where we used L'Hopital's rule for $g(h)$.

Definition 2. $\Longrightarrow$ Definition 1. The proof that $\{X(t), t\geq 0\}$ has stationary independent increments is quite difficult. We shall assume this has been done. Hence we need only show that

$$\Pr\{X_t = n\} = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, 2, \dots$$

Define

$$p_n(t) = \Pr\{X_t = n\}.$$

Then if $n \geq 1$,

$$\begin{aligned} p_n(t+h) &= \Pr\{X_{t+h} = n\} = \Pr\{X_{t+h} = n|X_t = n-1\}p_{n-1}(t) + \Pr\{X_{t+h} = n|X_t = n\}p_n(t) \\ &+ \sum_{i=0}^{n-2}\Pr\{X_{t+h} = n|X_t = i\}p_i(t) \\ &= \big(\lambda h + o(h)\big)p_{n-1}(t) + \big(1-\lambda h + o(h)\big)p_n(t) + o(h) \\ &= \lambda h p_{n-1}(t) + p_n(t) - \lambda h p_n(t) + o(h). \end{aligned}$$

Hence

$$\frac{p_n(t+h)-p_n(t)}{h} = \lambda p_{n-1}(t) - \lambda p_n(t) + \frac{o(h)}{h}.$$

Letting $h\to 0$ gives

$$p_n'(t) = \lambda p_{n-1}(t) - \lambda p_n(t).$$

If $n = 0$, then

$$p_0(t+h) = \Pr\{X_{t+h} = 0|X_t = 0\}p_o(t) = \big(1-\lambda h + o(h)\big)p_0(t),$$

which leads to

$$p_0'(t) = -\lambda p_0(t).$$

$$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n\geq 1 \tag{2.1}$$
$$p_0'(t) = -\lambda p_0(t), \tag{2.2}$$
$$p_0(0) = 1, \quad p_n(0) = 0 \quad \text{if } n \geq 1.$$

are a special case of the Kolmogorov Forward Equations, to be discussed in the next section. We give two methods of solution.

Method A. (Recursive Method) Write (2.2) as

$$p_0'(t) + \lambda p_0(t) = 0.$$

Multiply both sides by the integrating factor $e^{\lambda t}$. We get $\big(p_0(t)e^{\lambda t}\big)' = 0$. Integrating both sides, we find $p_0(t)e^{\lambda t} = k$ (a constant), so $p_0(t) = ke^{-\lambda t}$. We identify k by $1 = p_0(0) = k$. Hence

$$p_0(t) = e^{-\lambda t}.$$

Now put $n = 1$ in (2.1). We get

$$p_1'(t) = -\lambda p_1(t) + \lambda p_0(t) = -\lambda p_1(t) + \lambda e^{-\lambda t},$$

so that $p_1'(t)+\lambda p_1(t) = \lambda e^{-\lambda t}$. Multiplying by the integrating factor $e^{\lambda t}$, we find $\big(p_1(t)e^{\lambda t}\big)' = \lambda$. Integrating both sides gives $p_1(t)e^{\lambda t} = \lambda t + k$, so that $p_1(t) = \lambda t e^{-\lambda t} + ke^{-\lambda t}$. We find k by $0 = p_1(0) = k$. Hence

$$p_1(t) = \lambda t e^{-\lambda t}.$$

Repetition (i.e. an induction argument) gives

$$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, 2, \ldots$$

Method B. (Solution by Generating Functions) For each $t \geq 0$, let

$$P(t,s) = \sum_{n=0}^{\infty} p_n(t)s^n \quad \Big(= E(s^{X_t})\Big)$$

be the p.g.f. of the random variable X_t. Multiply (2.1) by s^n and sum over $n \geq 1$, getting

$$\sum_{n=1}^{\infty} p_n'(t)s^n = -\lambda \sum_{n=1}^{\infty} p_n(t)s^n + \lambda \sum_{n=1}^{\infty} p_{n-1}(t)s^n.$$

Add in (2.2) and get

$$\begin{aligned}\sum_{n=0}^{\infty} p_n'(t)s^n &= -\lambda \sum_{n=0}^{\infty} p_n(t)s^n + \lambda \sum_{n=1}^{\infty} p_{n-1}(t)s^n \\ &= -\lambda \sum_{n=0}^{\infty} p_n(t)s^n + \lambda s \sum_{n=1}^{\infty} p_{n-1}(t)s^{n-1}.\end{aligned}$$

Thus

$$\frac{dP}{dt} = -\lambda P + \lambda s P = \lambda(s-1)P \tag{2.3}$$

with condition

$$P(0,s) = 1.$$

Now we solve (2.3). We have

$$\frac{dP}{P} = \lambda(s-1),$$

so integration gives $\log P(t,s) = \lambda t(s-1) + K$. Hence

$$P(t,s) = Ce^{\lambda t(s-1)},$$

where $C = e^K$. Since $1 = P(0,s) = C$, then

$$P(t,s) = e^{\lambda t(s-1)},$$

which we recognize as the p.g.f. of a Poisson random variable with parameter λt. This finishes the proof that definitions 1 and 2 are equivalent.

Problem 2.1 (To show the structure of a Poisson process) Suppose A is a certain random event which occurs (possibly repeatedly) in continuous time $t \geq 0$. Let

$$\begin{aligned}
T_1 &= \text{time from 0 to 1st occurence of } A\\
T_2 &= \text{time from 1st to 2nd occurences of } A\\
T_3 &= \text{time from 2nd to 3rd occurences of } A\\
&\vdots\\
T_n &= \text{time from } n-1\text{st to } n\text{th occurences of } A\\
&\vdots
\end{aligned}$$

A is called a *pure renewal event* if $T_1, T_2, T_3, \ldots$ are i.i.d. Define

$$\begin{aligned}
S_n &= \begin{cases} 0 & \text{if } n = 0,\\ T_1 + T_2 + \ldots + T_n & \text{if } n \geq 1, \end{cases}\\
N_t &= \text{ number of occurences of } A \text{ in the interval } (0, t]\\
N_0 &= 0.
\end{aligned}$$

The process $\{S_n, n \geq 0\}$ is called a *pure renewal process*, and $\{N_t, t \geq 0\}$ is called the associated *renewal counting process*. Observe that $\{S_n \leq t\} = \{N_t \geq n\}$. Suppose that $T_1, T_2, T_3, \ldots$ are i.i.d. with exponential density

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0,\\ 0 & \text{if } t < 0. \end{cases}$$

(1) Use moment generating functions to find the distribution of S_n.

(2) Use the fact that $\{N_t = n\} = \{N_t \geq n\} \setminus \{N_t \geq n+1\}$ to find the probability function of N_t.(Hint: use part (a). You will find that $\Pr\{N_t = n\}$ is the difference of two complicated integrals. Use integration by parts on one of these integrals and the rest is easy.)

The point of this problem is that it displays the basic structure of a Poisson process as a certain renewal counting process in which the times between events are i.i.d. and exponential.

Problem 2.2 Let $\{X(t), t \geq 0\}$ be a Poisson process with rate λ. Show that if $0 < s < t$, then

$$\Pr\{X_s = m | X_t = n\} = \binom{n}{m} \big(\frac{s}{t}\big)^m \big(1 - \frac{s}{t}\big)^{n-m} \qquad \text{for } n \geq m.$$

Solution

$$\begin{aligned}
\Pr\{X_t = m | X_s = n\} &= \frac{\Pr\{X_s = m, X_t = n\}}{\Pr\{X_t = n\}} = \frac{\Pr\{X_s = m, X_t - X_s = n - m\}}{\Pr\{X_t = n\}}\\
&= \frac{\Pr\{X_s = m\} \Pr\{X_t - X_s = n - m\}}{\Pr\{X_t = n\}} = \binom{n}{m} \big(\frac{s}{t}\big)^m \big(1 - \frac{s}{t}\big)^{n-m},
\end{aligned}$$

after some cancellation and rearrangement.

4.3 Birth-and-Death Processes

In this section, we are going to generalize definition 2 of the Poisson process, by allowing transitions from a state i to the state $i-1$ as well as the state $i+1$, and by allowing the probability of making such transitions to depend on the current state of the process. The resulting stochastic process is called a birth-and-death process, and is popularly used as a model for the population growth of a system whose members can give birth or die (or drop out), such as a bacterial colony or a line of customers at a service installation.

Definition. A continuous time Markov chain $\{X(t), t \geq 0\}$ with stationary transition probabilities and state space $E = \{0, 1, 2, \dots\}$ is called a *birth-and-death (B&D) process* if

(1) $X_0 = i \geq 0$ and transitions from a state n are allowed only to the states $n+1$ and $n-1$ (the latter only if $n \geq 1$)

(2)

$$\begin{aligned}
&\Pr\{X_{t+h} = n+1 | X_t = n\} = \lambda_n h + o(h)\\
&\Pr\{X_{t+h} = n-1 | X_t = n\} = \mu_n h + o(h) \quad \text{ if } n \geq 1\\
&\Pr\{\text{more than one transition in the interval } t \text{ to } t+h | X_t = n\} = o(h)\\
&\text{and (as a result)}\\
&\Pr\{X_{t+h} = n | X_t = n\} = 1 - \lambda_n h - \mu_n h + o(h),
\end{aligned}$$

where $\lambda_n, n \geq 0$ and $\mu_n, n \geq 0$ are non-negative constants with $\mu_0 = 0$, called the birth coefficients and death coefficients, respectively.

Example. A Poisson process is a birth and death process with $\lambda_n = \lambda$ for $n \geq 0$ and $\mu_n = 0$ for $n \geq 1$.

Now we attempt to obtain a family of differential equations for $p_n(t) = \Pr\{X_t = n\}$, as was done in the previous section for the Poisson process. We have

$$\begin{aligned}
\{X_{t+h} = n\} = \{X_{t+h} = n, X_t = n\} &\cup \{X_{t+h} = n, X_t = n-1\} \cup \{X_{t+h} = n, X_t = n+1\}\\
&\cup \{X_{t+h} = n, |X_{t+h} - X_t| > 1\},
\end{aligned}$$

where, if $n = 0$, the set with $n-1$ in it is taken to be empty. Then

$$\begin{aligned}
p_n(t+h) &= \Pr\{X_{t+h} = n, X_t = n\} + \Pr\{X_{t+h} = n, X_t = n-1\}\\
&\quad + \Pr\{X_{t+h} = n, X_t = n+1\} + \Pr\{X_{t+h} = n, |X_{t+h} - X_t| > 1\}\\
&= \Pr\{X_{t+h} = n | X_t = n\} p_n(t) + \Pr\{X_{t+h} = n | X_t = n-1\} p_{n-1}(t)\\
&\quad + \Pr\{X_{t+h} = n | X_t = n+1\} p_{n+1}(t) + \Pr\{\text{more than one transition between } t \text{ and } t+h\}\\
&= \big(1 - \lambda_n h - \mu_n h + o(h)\big) p_n(t) + \big(\lambda_{n-1} h + o(h)\big) p_{n-1}(t) + \big(\mu_{n+1} h + o(h)\big) p_{n+1}(t) + o(h).
\end{aligned}$$

Hence

$$\frac{p_n(t+h) - p_n(t)}{h} = -\big(\lambda_n + \mu_n\big) p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t) + \frac{o(h)}{h},$$

so that letting $h \to 0$, we find

$$p_n'(t) = -(\lambda_n + \mu_n) p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t),$$

where, if $n = 0$, the term with $n-1$ in it is taken to be zero. We have therefore obtained a family of differential equations (i.e. one equation for each n), called the *Kolmogorov Forward Equations*. We summarize them as

$$\begin{aligned}
&p_n'(t) = -(\lambda_n + \mu_n) p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t), \quad n \geq 1 && (3.1)\\
&p_0'(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t) && (3.2)\\
&p_n(0) = \begin{cases} 1 & \text{if } n = i\\ 0 & \text{if } n \neq i. \end{cases}
\end{aligned}$$

To find $p_n(t)$, we must solve this system of differential equations. In general, this is extremely difficult.

However, the following theorem assures us that a solution always exists.

Theorem 4.3.1 *For any given sequences $\lambda_n \geq 0, \mu_n \geq 0$ of birth and death coefficients, there always exists a non-negative solution $p_n(t)$ of the above system such that $\sum_{n=0}^{\infty} p_n(t) \leq 1$ for all $t \geq 0$. This solution $p_n(t)$ is called the minimal solution because if $p_n^*(t)$ is another non-negative solution, then $p_n(t) \leq p_n^*(t)$ for all $n \geq 0$ and $t \geq 0$. If the coefficients are bounded (or increase sufficiently slowly with n), this minimal solution is unique and satisfies $\sum_{n=0}^{\infty} p_n(t) = 1$ for all $t \geq 0$. However, it is possible to choose the λ_n and μ_n in such a way that $\sum_{n=0}^{\infty} p_n(t) < 1$ for all $t \geq 0$, and in this case there will be infinitely many solutions of the forward equations.*

Remark. Note that

$$\sum_{n=0}^{\infty} p_n(t) = \sum_{n=0}^{\infty} \Pr\{X_t = n\} = \Pr\{X_t < +\infty\}.$$

Thus the possibility that $\sum_{n=0}^{\infty} p_n(t) < 1$ corresponds to a positive probability of X_t, the "population", being infinite by time t. When this happens, we say an "explosion"has taken place.

The Simple Birth-and-Death Process. This simply amounts to taking $\lambda_n = n\lambda$ and $\mu_n = n\mu$ for the birth and death coefficients, where $\lambda \geq 0$ and $\mu \geq 0$. This is a model for the following situation: consider a population of members which can (by splitting or otherwise) give birth to new members, and also at some point die. Assume that during any short time interval of length h, each member has probability $\lambda h + o(h)$ of giving birth to a new member, and $\mu h + o(h)$ of dying. There is no interaction between members. Then if at time t, the population size is n, the probability of a birth in the time interval $(t, t+h]$ is $n\lambda h + o(h)$, and the probability of a death is $n\mu h + o(h)$.

Rationale: For simplicity, suppose $\mu_n = 0 \forall n$, so we are dealing with a pure birth process. We can see the actions of the n members as n independent trials, each with probability $\lambda h + o(h)$ of success. Then $P[X_{t+h} - X_t = j | X_t = n] = \binom{n}{j}[\lambda h + o(h)]^j [1 - \lambda h + o(h)]^{n-j}$. Now $[\lambda h + o(h)]^j = \sum_{r=0}^{j} \binom{j}{r} (\lambda h)^r o(h)^{j-r} = (\lambda h)^j + o(h)$, and similarly $[1 - \lambda h + o(h)]^k = \sum_{r=0}^{k} \binom{k}{r} (-\lambda h + o(h))^r = \sum_{r=0}^{k} \binom{k}{r} (-1)^r (\lambda h + o(h))^r = \sum_{r=0}^{k} \binom{k}{r} (-1)^r [(\lambda h)^r + o(h)] = 1 - k\lambda h + o(h)$. Hence

$$P[X_{t+h} - X_t = j | X_t = n] = \binom{n}{j}[(\lambda h)^j + o(h)][1 - (n-j)\lambda h + o(h)] = \binom{n}{j}[(\lambda h)^j - (n-j)(\lambda h)^{j+1} + o(h)]$$
$$= \begin{cases} o(h) & \text{if } j \geq 2, \\ n\lambda h + o(h) & \text{if } j = 1, \\ 1 - n\lambda h + o(h) & \text{if } j = 0. \end{cases}$$

In the case of the simple birth and death process, the forward equations are

$$p_n'(t) = -(\lambda + \mu) n p_n(t) + \lambda(n-1) p_{n-1}(t) + \mu(n+1) p_{n+1}(t), \quad n \geq 1 \tag{3.3}$$
$$p_0'(t) = \mu p_1(t) \tag{3.4}$$
$$p_n(0) = \begin{cases} 1 & \text{if } n = i \\ 0 & \text{if } n \neq i. \end{cases}$$

It is not too difficult to calculate the mean $m(t) = E(X_t)$ and variance $\sigma^2(t) = Var(X_t)$ for the simple birth and death process. We do this in the next proposition.

Proposition 4.3.2

$$m(t) = ie^{(\lambda-\mu)t}$$
$$\sigma^2(t) = \begin{cases} \frac{e^{2(\lambda-\mu)t}\left(1-e^{(\mu-\lambda)t}\right)\left(\lambda+\mu\right)}{\lambda-\mu} & \text{if } \lambda \neq \mu, \\ 2\lambda t & \text{if } \lambda = \mu, \end{cases}$$

Proof. Multiply (3.3) by n to get

$$\begin{aligned} np_n'(t) &= -(\lambda+\mu)n^2 p_n(t) + \lambda n(n-1)p_{n-1}(t) + \mu n(n+1)p_{n+1}(t) \\ &= -(\lambda+\mu)n^2 p_n(t) + \lambda(n-1)^2 p_{n-1}(t) + \lambda(n-1)p_{n-1}(t) \\ &\quad + \mu(n+1)^2 p_{n+1}(t) - \mu(n+1)p_{n+1}(t), \quad n \geq 1, \end{aligned}$$

and then sum over $n \geq 1$ to find

$$\begin{aligned} \sum_{n=1}^{\infty} np_n'(t) &= -(\lambda+\mu)\sum_{n=1}^{\infty} n^2 p_n(t) + \lambda \sum_{n=1}^{\infty}(n-1)^2 p_{n-1}(t) \\ &\quad + \lambda \sum_{n=1}^{\infty}(n-1)p_{n-1}(t) + \mu \sum_{n=1}^{\infty}(n+1)^2 p_{n+1}(t) - \mu \sum_{n=1}^{\infty}(n+1)p_{n+1}(t) \\ &= -(\lambda+\mu)\sum_{n=1}^{\infty} n^2 p_n(t) + \lambda \sum_{n=0}^{\infty} n^2 p_n(t) \\ &\quad + \lambda \sum_{n=0}^{\infty} np_n(t) + \mu \sum_{n=2}^{\infty} n^2 p_n(t) - \mu \sum_{n=2}^{\infty} np_n(t) \\ &= (\lambda-\mu)\sum_{n=1}^{\infty} np_n(t). \end{aligned}$$

Since $m(t) = \sum_{n=1}^{\infty} np_n(t)$, we have shown that

$$m'(t) = (\lambda-\mu)m(t).$$

Solving, and using the fact that $m(0) = i$, we find $m(t) = ie^{(\lambda-\mu)t}$, as required. A similar procedure gives the expression for $\sigma^2(t)$.

Remark. The deterministic analog of the birth and death process is the following: let $x(t)$ denote the population at time t, and assume that

$$\frac{dx(t)}{dt} = (\lambda-\mu)x(t), \qquad t \geq 0, \qquad x(0) = i.$$

Then $x(t) = ie^{(\lambda-\mu)t}$. Thus the mean $m(t)$ for the birth and death process coincides with the solution for the deterministic case.

Now we turn to the actual solution of the simple birth and death equations (3.3) and (3.4). Note that the recursive method of the last section will not work here. We will use the second method of solution by generating functions. (But note that there *are* other methods of solution, such as the Laplace transform method).

We multiply (3.3) by s^n and sum over $n \geq 1$, getting

$$\sum_{n=1}^{\infty} p_n'(t)s^n = -(\lambda+\mu)\sum_{n=1}^{\infty} np_n(t)s^n + \lambda \sum_{n=1}^{\infty}(n-1)p_{n-1}(t)s^n + \mu \sum_{n=1}^{\infty}(n+1)p_{n+1}(t)s^n$$

We now add in (3.4) and find

$$\begin{aligned}
\sum_{n=0}^{\infty} p_n'(t)s^n &= -(\lambda+\mu)\sum_{n=1}^{\infty} np_n(t)s^n + \lambda\sum_{n=1}^{\infty}(n-1)p_{n-1}(t)s^n + \mu\sum_{n=0}^{\infty}(n+1)p_{n+1}(t)s^n \\
&= -(\lambda+\mu)\sum_{n=1}^{\infty} np_n(t)s^n + \lambda s\sum_{n=1}^{\infty}(n-1)p_{n-1}(t)s^{n-1} + \frac{\mu}{s}\sum_{n=0}^{\infty}(n+1)p_{n+1}(t)s^{n+1} \\
&= -(\lambda+\mu)\sum_{n=1}^{\infty} np_n(t)s^n + \lambda s\sum_{n=0}^{\infty} np_n(t)s^n + \frac{\mu}{s}\sum_{n=1}^{\infty} np_n(t)s^n \\
&= \lambda s(s-1)\sum_{n=0}^{\infty} np_n(t)s^{n-1} + \mu(1-s)\sum_{n=0}^{\infty} np_n(t)s^{n-1} \\
&= \big(\lambda s(s-1)+\mu(1-s)\big)\sum_{n=0}^{\infty} np_n(t)s^{n-1}
\end{aligned}$$

Now let

$$P(t,s) = E(s^{X_t}) = \sum_{n=0}^{\infty} p_n(t)s^n$$

denote the p.g.f. of X_t. Then we have shown that $P(t,s)$ satisfies

$$\frac{\partial P}{\partial t} = (\mu - \lambda s)(1-s)\frac{\partial P}{\partial s} \tag{3.5}$$

with initial condition $P(0,s) = s^i$.

Equation (3.5) is an example of a *first order linear partial differential equation*. Such equations are usually solved by the *method of characteristics*, but we will adopt a simpler makeshift method. To simplify things, we will assume that $\mu = 0$ (although our method applies to the case $\mu > 0$ as well). The process is then known as the Yule-Furry Process, and is a model for a population whose members can give birth but not die. Equation (3.5) becomes

$$\frac{\partial P}{\partial t} = -\lambda s(1-s)\frac{\partial P}{\partial s} \tag{3.6}$$

with initial condition $P(0,s) = s^i$.

(Naturally, we assume $\lambda > 0$, since if both λ and μ are zero, then the problem is trivial because $X_t = i$ for all t). Our method of solving (3.6) is to obtain a "preliminary"solution by a separation of variables approach. Suppose $P(t,s) = U(t)V(s)$. Substitution into (3.6) gives

$$U'(t)V(s) = -\lambda s(1-s)U(t)V'(s)$$

whence

$$\frac{U'(t)}{U(t)} = -\lambda s(1-s)\frac{V'(s)}{V(s)} = c \text{ a constant.}$$

Integrating $U'(t)/U(t) = c$ gives

$$\log U(t) = ct + k_1.$$

On the other hand, we have

$$\frac{V'(s)}{V(s)} = -\frac{c}{\lambda s(1-s)} = -\frac{c}{\lambda}\Big[\frac{1}{s} + \frac{1}{1-s}\Big]$$

so that integration gives

$$\log V(s) = -\frac{c}{\lambda}\big[\log s - \log(1-s)\big] + k_2 = \log\big(\frac{1-s}{s}\big)^{c/\lambda} + k_2$$

There results

$$U(t) = k_3 e^{ct} \qquad \text{and} \qquad V(s) = k_4\big(\frac{1}{s} - 1\big)^{c/\lambda}.$$

so that

$$P(t,s) = ke^{ct}\big(\frac{1}{s} - 1\big)^{c/\lambda}. \tag{3.7}$$

While $P(t,s)$ as given in (3.7) will satisfy the PDE in (3.6) for any choice of the constants c and k, there unfortunately is no choice of constants c, k such that this $P(t,s)$ will satisfy the initial condition as well. However, it is a simple matter to show that if $P(t,s)$ is any solution of the partial differential equation in (3.6) (not necessarily the solution obtained in (3.7), and if $\phi(x)$ is any differentiable function, then $\phi[P(t,s)]$ is also a solution of (3.6). It follows from (3.7) that

$$P(t,s) = \phi[e^{\lambda t}(\frac{1}{s} - 1)] \tag{3.8}$$

is also a solution to the PDE in (3.6), and the idea is to choose ϕ in order that the initial condition be satisfied.

Putting $t = 0$ in (3.8), we have

$$s^i = P(0,s) = \phi[\frac{1}{s} - 1].$$

If we now set $u = \frac{1}{s} - 1$, then $s = (1+u)^{-1}$, and so $(1+u)^{-i} = s^i = \phi(u)$. The form of the function $\phi(u)$ has now been determined, and we have from (3.8), finally,

$$P(t,s) = \big[1 + e^{\lambda t}(\frac{1}{s} - 1)\big]^{-i}.$$

We recognize this p.g.f. as belonging to a negative binomial distribution, and in fact, we have

$$p_n(t) = \begin{cases} \binom{n-1}{i-1} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} & \text{if } n \ge i, \\ 0 & \text{if } 0 \le n < i. \end{cases}$$

$p_n(t)$ is the probability that, in a sequence of Bernoulli trials with probability $e^{-\lambda t}$ of success on each trial, the ith success occurs on the nth trial.

Problem. Repeat the above method to find the solution of (3.5) when $\mu > 0$.

Answer.

$$P(t,s) = \begin{cases} \left(\dfrac{r\mu - 1}{r\lambda - 1}\right)^i, \quad \text{where} \quad r = e^{(\lambda-\mu)t}\left(\dfrac{1-s}{\mu - \lambda s}\right) & \text{if } 0 < \lambda \ne \mu, \\ \big(1 - \dfrac{1}{\lambda r}\big)^i, \quad \text{where} \quad r = t + \dfrac{1}{\lambda(1-s)} & \text{if } 0 < \lambda = \mu, \\ [1 - (1-s)e^{-\mu t}]^i & \text{if } 0 = \lambda < \mu. \end{cases}$$

Ergodic Behaviour of the Birth and Death Process By "ergodic behaviour"of the birth and death process $\{X(t), t \ge 0\}$, we mean behaviour of the probabilities $p_n(t)$ as $t \to \infty$. It turns out that the theory here is much simpler than it was for discrete-time Markov chains, in large part because there is no notion of periodicity of states.

Recall the birth and death equations (3.1) and (3.2), namely

$$p_n'(t) = -(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t), \quad n \ge 1,$$
$$p_0'(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t).$$

Then we have the following theorem.

Theorem 4.3.3 *For arbitrarily prescribed coefficients* λ_n *and* μ_n*, the limits*

$$p_n = \lim_{t\to\infty} p_n(t), \qquad n \geq 0 \tag{3.9}$$

exist and are independent of the initial conditions. The numbers $p_n, n \geq 0$ *satisfy the above system of equations with* $p_n'(t) = 0$*, namely*

$$0 = -(\lambda_n + \mu_n)p_n + \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1}, \quad n \geq 1, \tag{3.10}$$

$$0 = -\lambda_0 p_0 + \mu_1 p_1. \tag{3.11}$$

These equations can be "derived" by letting $t \to \infty$ in the birth and death equations. Once we know that the limits $p_n = \lim_{t\to\infty} p_n(t)$ exist, then clearly $\lim_{t\to\infty} p_n'(t) = 0$ for all n.

Remarks. It is possible that all the p_n's in (3.9) are zero (corresponding to the case where all states are transient or recurrent null). Hence, if a steady-state (or long-run) distribution $\{p_n, n \geq 0\}$ exists, it must satisfy (3.10) and (3.11) as well as the condition $\sum_{n=0}^{\infty} p_n = 1$.

Problem. In a hospital ward are m patients with one nurse presiding. From time to time, a patient may become discontent and ask for the nurse's attention. If he is already attending another discontent person, the patient must wait her turn.

If at time t, a patient is content, then the probability that she will become discontent in the short time interval between t and $t+h$ is $\lambda h + o(h)$. If at time t a patient is being attended to by the nurse, then the probability that she will become content in the interval t to $t+h$ is $\mu h + o(h)$. Let $X(t)$ = number of discontent patients at time t.

(1) What are the appropriate Kolmogorov equations for this process?

(2) Find the steady state behaviour; that is, the limiting probabilities $p_n = \lim_{t\to\infty} \Pr\{X(t) = n\}$

Solution. The birth and death coefficients are $\lambda_n = (m-n)\lambda,\ 0 \leq n \leq m$, and $\mu_n = \mu,\ n \geq 1$. The forward equations are

$$\begin{aligned} p_m'(t) &= -\mu p_m(t) + \lambda p_{m-1}(t) \\ p_n'(t) &= -(\lambda(m-n) + \mu)p_n(t) + \lambda(m-n+1)p_{n-1}(t) + \mu p_{n+1}(t), \quad 1 \leq n < m, \\ p_0'(t) &= -\lambda m p_0(t) + \mu p_1(t). \end{aligned}$$

and so the steady-state equations are

$$\begin{aligned} 0 &= -\mu p_m + \lambda p_{m-1} \\ 0 &= -(\lambda(m-n) + \mu)p_n + \lambda(m-n+1)p_{n-1} + \mu p_{n+1}, \quad 1 \leq n < m, \\ 0 &= -\lambda m p_0 + \mu p_1. \end{aligned}$$

From the first equation, we have $p_{m-1} = (\mu/\lambda)p_m$. From the middle equation with $n = m-1$, we find $p_{m-2} = (\mu/\lambda)^2 p_m/2!$. In general, we find

$$p_{m-n} = \frac{(\mu/\lambda)^n}{n!} \cdot p_m, \quad 0 \leq n \leq m.$$

Using the fact that $\sum_{n=0}^{m} p_n = 1$, we then get

$$p_{m-n} = \frac{\frac{(\mu/\lambda)^n}{n!}}{\sum_{k=0}^{m} \frac{(\mu/\lambda)^k}{k!}}, \quad 0 \leq n \leq m.$$

Structure of a Birth and Death Process. Suppose that a birth and death process is in state n at some time. It is possible to show that the time T_n that the process stays in that state (called the holding time in state n) has an exponential distribution with mean $1/(\lambda_n + \mu_n)$. That is, T_n has density function

$$f(t) = \begin{cases} (\lambda_n + \mu_n)e^{-(\lambda_n+\mu_n)t} & \text{if } t \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

At the end of this holding time, the process either makes a transition to state $n+1$ with probability $\lambda_n/(\lambda_n + \mu_n)$, or to state $n-1$ (if $n \geq 0$) with probability $\mu_n/(\lambda_n + \mu_n)$. Suppose it goes to state $n+1$. Then it stays in that state for the holding time T_{n+1}, at the end of which it makes a transition down to state n or up to state $n+2$. The motion goes on and on like this. The individual holding times in the states visited are independent of each other.

Explosions. An *explosion* is when a birth and death process goes through infinitely many transitions in a finite amount of time. Let us see what this means in the case of a pure birth process starting in state i. It stays in state i for the holding time T_i, and then makes a transition to state $i+1$. It stays there for the holding time T_{i+1}, and then goes to state $i+2$, and so on. Let $T = \sum_{n=i}^{\infty} T_n$. Then T is the amount of time required to go through all the states. One might think that $T = \infty$, but this is not necessarily so. In fact, for a pure birth process, we have an explosion iff $T < \infty$. That is, $P\{\text{explosion}\} = P\{T < \infty\}$.

We need the following fact about sequences of exponential random variables.

Fact: Let $X_n, n \geq 1$ be independent exponentially distributed random variables, and let $X = \sum_{n=1}^{\infty} X_n$. Then $P\{X = \infty\} = 1$ iff $\sum_{n=1}^{\infty} E(X_n) = \infty$.

Hence we have $P\{T = \infty\} = 1$ (i.e. no explosion) iff $\sum_{n=i}^{\infty} \frac{1}{\lambda_n} = \infty$.

Example. In the case of a Poisson process, we have $\lambda_n = \lambda$ for all n, and $\sum_{n=i}^{\infty} \frac{1}{\lambda} = \infty$, so there can be no explosion. For a Yule-Furry process, where $\lambda_n = n\lambda$, we have $\sum_{n=i}^{\infty} \frac{1}{n\lambda} = \frac{1}{\lambda} \sum_{n=i}^{\infty} \frac{1}{n} = \infty$, so again there is no explosion. But if $\lambda_n = n^2\lambda$, then there will be an explosion.

If we are dealing with a birth and death process where deaths are possible, there still can be explosions. For example, suppose both $\lambda_n = \mu_n = n^2\lambda$. Then the process is equally likely to go down as up, so it stays at a finite value, but the mean holding times are $\frac{1}{2n^2\lambda}$, so with positive probability, there will at some point be infinitely many transitions in a finite time interval.

Chapter 5

Queueing Theory

5.1 Introduction and Notation

We consider a stream of *customers* arriving at a *service facility*. The nth customer will be denoted by C_n, $n \geq 1$, and arrives at the service facility at time τ_n (a random variable). The service facility contains m *servers*, each capable of serving only one customer at a time. If a customer arrives to find all m servers busy, he enters the *queue* (i.e. waiting line). Then, when one of the servers becomes *idle*, a customer from the waiting line is admitted to that server. On the other hand, if a customer arrives and one (or more) of the servers is idle, that customer is immediately admitted to the server. The time spent by customer C_n in a server is a random variable X_n, called the *service time*. This does not include the time W_n (if any) spent by C_n in the waiting line. The total time spent by C_n in the system, then, (where "system"consists of servers and waiting line) is therefore $S_n = W_n + X_n$. Also, let

$$T_n = \begin{cases} \tau_n - \tau_{n-1} \text{ (the interarrival time between } C_{n-1} \text{ and } C_n), & \text{if } n \geq 2, \\ \tau_1 & \text{if } n = 1. \end{cases}$$

$$N_t = \text{the number of customers in the system at time } t \geq 0.$$

Traditionally, the customer interarrival times $T_n, n \geq 1$ are assumed to be i.i.d. random variables, and the customer service times $X_n, n \geq 1$ are assumed to be i.i.d. random variables. Moreover, the interarrival times T_n are all assumed to be independent of the service times X_n.

Over and above the above conventions, there is a brief notation used to describe a particular queueing system. It is

$$G_1/G_2/m/K$$

where

$$\begin{aligned} &G_1 \text{ stands for the distribution of the } T_n\text{'s} \\ &G_2 \text{ stands for the distribution of the } X_n\text{'s} \\ &m = \text{ the number of servers} \\ &K = \text{the storage capacity of the queue.} \end{aligned}$$

If $K = +\infty$, $G_1/G_2/m/K$ is usually written as just $G_1/G_2/m$.

5.2 The M/M/1 Queueing System

This is the simplest queueing system. There is only one server and there are no bounds on the length of the waiting line. The "M"stands for "memoryless", and means that the interarrival times and service times are exponentially distributed. (Recall that the exponential distribution is the only continuous distribution that has the memoryless property.)

Specifically, we assume that

(1) the interarrival times T_n have exponential density function with mean λ^{-1}. This is equivalent to assuming that if X_t = number of customers to arrive in the interval $(0, t]$, then $\{X(t), t \geq 0\}$ is a Poisson process.

(2) the service times X_n are exponentially distributed with mean μ^{-1}.

The impact of assumptions (1) and (2) is that the process $\{N_t, t \geq 0\}$ is a birth and death process with birth parameters $\lambda_n = \lambda$ for all $n \geq 0$, and death parameters $\mu_n = \mu$ for all $n \geq 1$. Unfortunately, a proof of this fact requires slightly more knowledge of the "inside structure"of a birth and death process than was covered in chapter 4.

Let us define

$$p_n(t) = \Pr\{N_t = n\} = \Pr\{n \text{ customers in the system at time } t\}.$$

Then the $p_n(t)$'s satisfy the Kolmogorov equations

$$p_n'(t) = -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), \quad n \geq 1 \tag{2.1}$$

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t) \tag{2.2}$$

$$p_n(0) = \begin{cases} 1 & \text{if } n = i \\ 0 & \text{if } n \neq i. \end{cases}$$

where i is the number of customers in the system at time $t = 0$. We shall take i to be 0.

Transient behaviour. By "transient"we mean the behaviour for finite t. This means we have to solve the Kolmogorov equations for $p_n(t)$. Define

$$P(t, z) = E\left(z^{N_t}\right) = \sum_{n=0}^{\infty} p_n(t) z^n \quad \text{(the p.g.f. of } N_t).$$

In the usual way (multiplying (2.1) by z^n, summing over $n \geq 1$, then adding in (2.2)), we get

$$z\frac{\partial P}{\partial t} = (1 - z)\left[(\mu - \lambda z)P(t, z) - \mu p_0(t)\right], \tag{2.3}$$

$$P(0, z) = 1. \tag{2.4}$$

To solve this system, we introduce the Laplace transforms

$$P^*(s, z) = \int_0^{\infty} e^{-st} P(t, z)\, dt, \qquad p_0^*(s) = \int_0^{\infty} e^{-st} p_0(t)\, dt$$

of $P(t, z)$ and $p_0(t)$ respectively. Then taking Laplace transforms on both sides of (2.3), we get

$$z[sP^*(s, z) - P(0, z)] = (1 - z)[(\mu - \lambda z)P^*(s, z) - \mu p_0^*(s)].$$

Using (2.4) and solving for $P^*(s, z)$, we find

$$P^*(s, z) = \frac{z - \mu(1 - z)p_0^*(s)}{sz - (1 - z)(\mu - \lambda z)}.$$

Note that $p_0^*(s)$ here is unknown (it is part of the solution). Through a fairly complicated analysis, this double transform can be inverted to give

$$p_n(t) = e^{-(\lambda+\mu)t}\left[\rho^{(n-1)/2} I_n(at) + \rho^{(n-2)/2} I_{n+1}(at) + (1-\rho)\rho^n \sum_{j=n+3}^{\infty} \rho^{-j/2} I_j(at)\right]$$

where

$$\rho = \frac{\lambda}{\mu} \text{ is called the \textit{traffic intensity}},$$

$$a = 2\mu\rho^{1/2},$$

$$I_k(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{k+2m}}{(k+m)!m!} \quad \text{(the modified Bessel function of the first kind of order } k\text{)}.$$

It is surprising that for such a simple choice of birth and death coefficients, the solution of the Kolmogorov equations can be so complicated.

Ergodic Behaviour. The solution obtained above is so complicated, it is difficult to obtain information from it. In addition, we envisage that more complicated queueing systems will be even more difficult to solve for the transient behaviour. Thus in this section, we try another approach. We examine the ergodic behaviour of the queueing system — namely, what happens as $t \to \infty$. We will see that under certain conditions (as with ergodic discrete-time Markov chains), the system settles down to a steady state. Furthermore, the transient solution may settle down very quickly to the steady state, so that the latter may reasonably be taken as an approximation to the actual behaviour for moderately large finite t.

Let us define

$$p_n = \lim_{t\to\infty} p_n(t) \approx \Pr\{n \text{ customers in system after a long time}\}$$

From the very last part of chapter 4, we know that the p_n's satisfy the system of equations

$$0 = -(\lambda+\mu)p_n + \lambda p_{n-1} + \mu p_{n+1}, \quad n \ge 1,$$

$$0 = -\lambda p_0 + \mu p_1.$$

Solving recursively, we have

$$\mu p_1 = \lambda p_0 \quad \Rightarrow \quad p_1 = \frac{\lambda}{\mu} p_0,$$

$$\mu p_2 = (\lambda+\mu)p_1 - \lambda p_0 = \frac{\lambda^2}{\mu} p_0 \quad \Rightarrow p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0,$$

and so on, finding that

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0.$$

Define $\rho = \lambda/\mu$ (the traffic intensity). We can find p_0 from the condition $\sum_{n=0}^{\infty} p_n = 1$. We have

$$1 = \sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} p_0 \rho^n = p_0 \cdot \frac{1}{1-\rho},$$

provided $\rho < 1$. Then $p_0 = 1-\rho$, and we conclude that a steady state distribution exists for the $M/M/1$ queueing system if $\rho < 1$, and that distribution is

$$p_n = (1-\rho)\rho^n, \quad n = 0, 1, 2, \ldots$$

Remarks. Assume that $\rho < 1$, so that a steady-state distribution exists.

(1) The average number of customers in the system at steady state is

$$E(N) = \sum_{n=0}^{\infty} np_n = \frac{\rho}{1-\rho}.$$

Similarly,

$$Var(N) = \frac{\rho}{(1-\rho)^2}.$$

(2) The average waiting time of an arriving customer in the queue at steady state is

$$E(W) = \sum_{n=1}^{\infty} E(W|N=n)p_n = \sum_{n=1}^{\infty} \frac{n}{\mu} \cdot p_n = \frac{1}{\mu}E(N) = \frac{\rho}{\mu(1-\rho)}.$$

Similarly,

$$Var(W) = \frac{\rho(2-\rho)}{\mu^2(1-\rho)^2}.$$

(3) The average time spent in the system at steady state is

$$E(S) = \frac{1}{\mu} + E(W) = \frac{1}{\mu(1-\rho)}.$$

5.3 The M/M/m Queueing System

We have the same setup as in the last section, but now there are $m \geq 1$ servers. The number N_t of customers in the system at time t is a birth and death process with coefficients

$$\lambda_n = \lambda, \quad n \geq 0,$$

$$\mu_n = \begin{cases} n\mu & \text{if } 0 \leq n \leq m, \\ m\mu & \text{if } n \geq m. \end{cases}$$

The transient behaviour is too difficult to solve for, so we will look only at the steady state behaviour. The steady state equations are

$$0 = -(\lambda + m\mu)p_n + \lambda p_{n-1} + m\mu p_{n+1}, \quad n \geq m,$$

$$0 = -(\lambda + n\mu)p_n + \lambda p_{n-1} + (n+1)\mu p_{n+1}, \quad 1 \leq n < m,$$

$$0 = -\lambda p_0 + \mu p_1.$$

Solving recursively, we have

$$p_1 = \frac{\lambda}{\mu}p_0, \quad p_2 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 p_0,$$

and in general, for $0 \leq n \leq m$, that

$$p_n = \frac{(\lambda/\mu)^n p_0}{n!}.$$

For $n \geq m$, the form of the equation changes, and we get

$$p_n = \frac{(\lambda/\mu)^n p_0}{m! m^{n-m}}.$$

Let us define the *traffic intensity* $\rho = \lambda/m\mu$. Then we can write the p_n's as

$$p_n = \begin{cases} \frac{(m\rho)^n}{n!} \cdot p_0 & \text{if } 0 \le n \le m, \\ \frac{\rho^n m^m}{m!} \cdot p_0 & \text{if } n \ge m. \end{cases} \tag{3.1}$$

Notice that

$$\sum_{n=m}^{\infty} p_n = p_0 \cdot \frac{m^m}{m!} \sum_{n=m}^{\infty} \rho^n = p_0 \cdot \frac{(\rho m)^m}{m!} \sum_{n=m}^{\infty} \rho^{n-m} = p_0 \cdot \frac{(\rho m)^m}{m!} \frac{1}{1-\rho} < +\infty \tag{3.2}$$

provided $\rho < 1$. Hence a steady state distribution, given by (3.1), exists if and only if $\rho < 1$. Note that in (3.1), p_0 is determined by the condition $1 = \sum_{n=0}^{\infty} p_n = \sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} p_0 + p_0 \frac{(\rho m)^m}{m!(1-\rho)}$, so

$$p_0 = \left[\sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} + \frac{(\rho m)^m}{m!(1-\rho)} \right]^{-1}. \tag{3.3}$$

Remarks. Assume that $\rho < 1$, so that a steady-state distribution exists.

(1) Using (3.2), we have

$$\Pr\{\text{arriving customer must queue}\} = \sum_{n=m}^{\infty} p_n = p_0 \cdot \frac{(\rho m)^m}{m!} \frac{1}{1-\rho}. \tag{3.4}$$

This is called *Erlang's C Formula*.

(2) We compute the average waiting time in the queue at steady state. Let

$$W = \text{waiting time of an arriving customer in the queue.}$$

If a customer arrives to find $n \ge m$ others in the system, he must wait until exactly $n-m+1$ customers depart before he can enter service. Hence

$$\begin{aligned} E(W) &= \sum_{n=m}^{\infty} E(W|N=n) p_n = \sum_{n=m}^{\infty} \frac{n-m+1}{m\mu} p_n \\ &= p_0 \cdot \frac{(\rho m)^m}{(m\mu) m!} \sum_{n=m}^{\infty} (n-m+1) \rho^{n-m} = p_0 \cdot \frac{(\rho m)^m}{(m\mu) m!} \sum_{k=1}^{\infty} k \rho^{k-1} \\ &= p_0 \cdot \frac{(\rho m)^m}{(m\mu) m!} \cdot \frac{1}{(1-\rho)^2}. \end{aligned}$$

Example. Consider the case where $m = 2$. Then from (3.3) we get

$$p_0 = \frac{1-\rho}{1+\rho}.$$

Hence

$$\Pr\{\text{customer must queue}\} = \frac{2\rho^2}{1+\rho},$$

$$E(W) = \frac{\rho^2}{\mu(1-\rho^2)}.$$

Problem. Suppose we have a Poisson stream of customers entering a service facility with two servers (e.g. a bank with two tellers). Which is better —

(1) a waiting system in which lines form in front of each teller (and no jockeying is allowed)?

(2) a single line served by the two tellers?

In case (a), we may assume two separate input streams, each with rate $\lambda/2$. The traffic intensity for each queue is therefore $\rho = \lambda/2\mu$, the same as in case (b). We may therefore directly compare the expressions for $E(W)$ obtained above and in the previous section. We have

$$E(W)_{(b)} = \frac{\rho^2}{\mu(1-\rho^2)} = \frac{\rho}{1+\rho} \cdot \frac{\rho}{\mu(1-\rho)} = \frac{\rho}{1+\rho} E(W)_{(a)}$$

where $E(W)_{(a)}$ denotes the expected waiting time for case (a), and $E(W)_{(b)}$ is similarly defined. Since $\rho/(1+\rho) < 1/2$, case (b) is clearly superior.

Problem. Repeat the preceding analysis in the case of m servers.

Solution. From (3.3) and remark (2), we have

$$EW_{(b)} = p_0 \frac{(\rho m)^{m-1}}{m!(1-\rho)} \cdot \frac{\rho}{\mu(1-\rho)} = R(\rho) EW_{(a)},$$

where $\rho = \lambda/m\mu$ and

$$R(\rho) = p_0 \frac{(\rho m)^{m-1}}{m!(1-\rho)} = \frac{(\rho m)^{m-1}}{m!(1-\rho)\sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} + (\rho m)^m}.$$

Observe that $\lim_{\rho \uparrow 1} R(\rho) = \frac{1}{m}$, generalizing the result for $m = 2$.

5.4 The M/G/1 Queueing System

The situation here is the same as with the $M/M/1$ queue, except for the fact that the service time density function $b(t)$ is arbitrary, and not necessarily exponential. Indeed, the "G"in $M/G/1$ stands for "general". If $b(t)$ is not exponential, the process $\{N_t, t \geq 0\}$ is no longer a Markov process, and we cannot use the birth and death formulation as we did for the $M/M/1$ and $M/M/m$ queues.

However, even though N_t is no longer Markovian, we will be able to find a Markov chain imbedded in the structure, and this will allow us to solve for the steady state behaviour of the system. For each $n \geq 1$, define

$$v_n = \text{the number of customers arriving during the service of } C_n,$$

$$q_n = \text{the number of customers left behind in the system by the departure of } C_n \text{ from service.}$$

Then

$$q_{n+1} = \begin{cases} q_n - 1 + v_{n+1} & \text{if } q_n > 0, \\ v_{n+1} & \text{if } q_n = 0, \end{cases} \tag{4.1}$$

where the -1 is C_{n+1} herself. Note that

(1) v_n is a random variable with probability function

$$\alpha_k = \Pr\{v_n = k\} = \int_0^\infty \Pr\{v_n = k | X_n = t\} b(t)\, dt = \int_0^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} b(t)\, dt, \quad k \geq 0.$$

(2) v_{n+1} is independent of q_n, q_{n-1}, etc. From (4.1), we see that the random variables $q_1, q_2, q_3, \ldots$ form a Markov chain with state space $E = \{0, 1, 2, \ldots\}$. Let us determine the transition matrix. From (4.1), we have

$$P_{ij} = \Pr\{q_{n+1} = j | q_n = i\} = \begin{cases} 0 & \text{if } 0 \le j < i-1, \\ \Pr\{v_{n+1} = j - i + 1\} = \alpha_{j-i+1} & \text{if } i > 0, j \ge i-1, \\ \Pr\{v_{n+1} = j\} = \alpha_j & \text{if } i = 0. \end{cases}$$

Hence the transition matrix is

$$P = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \ldots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \ldots \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \ldots \\ 0 & 0 & \alpha_0 & \alpha_1 & \alpha_2 & \ldots \\ 0 & 0 & 0 & \alpha_0 & \alpha_1 & \ldots \\ \vdots & \vdots & & \vdots & & \end{pmatrix}.$$

It can be shown that for $\alpha_j > 0$, $j \ge 0$, the above transition matrix is irreducible, aperiodic, and

(1) *transient* if $\sum_{k=0}^{\infty} k\alpha_k > 1$,

(2) *null* if $\sum_{k=0}^{\infty} k\alpha_k = 1$,

(3) *positive* if $\sum_{k=0}^{\infty} k\alpha_k < 1$.

Observe also that

$$\begin{aligned} \sum_{k=0}^{\infty} k\alpha_k &= \sum_{k=1}^{\infty} k \int_0^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} b(t)\, dt = \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{(k-1)!} \right\} e^{-\lambda t} b(t)\, dt \\ &= \int_0^{\infty} \lambda t \left\{ \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} \right\} e^{-\lambda t} b(t)\, dt = \int_0^{\infty} \lambda t e^{\lambda t} e^{-\lambda t} b(t)\, dt \\ &= \lambda \int_0^{\infty} t b(t)\, dt = \lambda E(X) \end{aligned}$$

where $E(X)$ denotes the mean service time per customer. Thus we define

$$\rho = \lambda E(X),$$

and as usual we obtain

$$\text{a steady state distribution exists} \iff \rho < 1.$$

The Pollaczek-Khintchine (P–K) Equations We will assume that $\rho < 1$, so that a steady state exists for the system. Then $q_n \to \tilde{q}$ (in the sense that $\Pr\{q_n = k\} \to \Pr\{\tilde{q} = k\}$ for each k) as $n \to \infty$, where $\tilde{q}$ is the number of customers in the system at steady state. We wish to find the distribution of $\tilde{q}$. From the ergodic theorem for Markov chains, we know this is the unique invariant distribution for the transition matrix P. That is, if $\pi_k = \Pr\{\tilde{q} = k\}$ for $k \ge 0$, then

$$\pi = \pi P.$$

This gives rise to the equations

$$\begin{aligned} \pi_0 &= (\pi_0 + \pi_1)\alpha_0 \\ \pi_1 &= \pi_0\alpha_1 + \pi_1\alpha_1 + \pi_2\alpha_0 \\ \pi_2 &= \pi_0\alpha_2 + \pi_1\alpha_2 + \pi_2\alpha_1 + \pi_3\alpha_0 \\ &\vdots \\ \pi_n &= \pi_0\alpha_n + \sum_{i=1}^{n+1} \pi_i \alpha_{n+1-i}, \quad n \ge 1. \end{aligned} \tag{4.2}$$

Defining

$$Q(s) = \sum_{n=0}^{\infty} \pi_n s^n, \qquad V(s) = \sum_{n=0}^{\infty} \alpha_n s^n,$$

we multiply (4.2) by s^n, sum the result over $n \geq 1$, and add in $\pi_0 = (\pi_0 + \pi_1)\alpha_0$ to get (after some arithmetic)

$$sQ(s) = \pi_0 sV(s) + [Q(s) - \pi_0]V(s),$$

so that

$$Q(s) = \frac{\pi_0(1-s)V(s)}{V(s) - s}.$$

Also, if $b^*(s)$ denotes the Laplace transform of the service time density $b(t)$, then

$$\begin{aligned} V(s) = \sum_{k=0}^{\infty} \alpha_k s^k &= \sum_{k=0}^{\infty} \left\{ \int_0^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} b(t)\,dt \right\} s^k = \int_0^{\infty} \left\{ \sum_{k=0}^{\infty} \frac{(\lambda t s)^k}{k!} \right\} e^{-\lambda t} b(t)\,dt \\ &= \int_0^{\infty} e^{\lambda t s} e^{-\lambda t} b(t)\,dt = \int_0^{\infty} e^{-(\lambda - \lambda s)t} b(t)\,dt \\ &= b^*(\lambda - \lambda s), \end{aligned}$$

so that

$$Q(s) = \frac{\pi_0(1-s)b^*(\lambda - \lambda s)}{b^*(\lambda - \lambda s) - s}. \tag{4.3}$$

We still have to determine π_0. Letting $s \uparrow 1$ in (4.3), we get

$$1 = Q(1) = \lim_{s \uparrow 1} \frac{\pi_0(1-s)b^*(\lambda - \lambda s)}{b^*(\lambda - \lambda s) - s} = \frac{\pi_0}{\lambda b^{*\prime}(0) + 1},$$

where the limit on the right was calculated using L'Hopital's rule. Since $b^{*\prime}(0) = -E(X)$ (mean service time), then

$$\pi_0 = \lambda b^{*\prime}(0) + 1 = 1 - \rho.$$

We have thus derived the famous *Pollaczek-Khintchine* (P–K) formula

$$\boxed{Q(s) = \frac{(1-\rho)(1-s)b^*(\lambda - \lambda s)}{b^*(\lambda - \lambda s) - s}}$$

The formula may be used to determine the waiting time distribution, average waiting time, average queue length, and so on. For example,

$$E(\tilde{q}) = Q'(1) = \rho + \frac{\lambda E(X^2)}{2(1-\rho)},$$

where $Q'(1)$ has to be computed using L'Hopital's rule twice. Using $Var(X) = E(X^2) - (E(X))^2$, we find

$$E(\tilde{q}) = \rho + \frac{\lambda \left(E(X)\right)^2}{2(1-\rho)} + \frac{\lambda Var(X)}{2(1-\rho)},$$

This is a surprising result, since it indicates that over all types of server with the same average service time, the most "consistent"server is best, and the optimal server is one whose service time is constant (not a random variable).