8 Jan 2009 5:13 p.m.

University of Sussex

Department of Mathematics

G1101 RANDOM PROCESSES

Spring Term 2008–09

Lecture Notes by Charles M. Goldie, adapted from originals by Dr Mark Broom

Contents

Chapter 0: Introduction

A huge number of real-life situations can be modelled mathematically, and many of these are probabilistic (as opposed to deterministic). The course *Probability Models*, as well as giving a rigorous introduction to probability, introduced some examples of such models (mainly in discrete time), e.g. discrete-time Markov chains.

Random Processes follows on from *Probability Models*, introducing continuous-time probabilistic models. We deal with several different types of process in continuous time, giving a basic introduction to the processes that we are trying to model, the assumptions of the model, an analysis of the model and mathematical examples of how to use the model in each case. Work is divided into six chapters (after this one) which are summarised as follows:

Chapter 1, *Poisson processes.* This is the simplest continuous-time process: points occur at random in time at a constant rate (e.g. radioactive emissions). We ask questions such as:

- what is the distribution of the number of points to have occurred before time t?

Chapter 2, *Birth processes.* We model a population of organisms each of which gives birth at a constant rate, so that the average rate of population increase is proportional to population size. This is not a realistic population model, since the population size always increases, but is a good place to start to develop more complex models. We also deal with *death processes* where the population decreases. These form building blocks for the models of Chapter 3.

Chapter 3, *Birth and death processes.* We combine birth or immigration with death so that, unlike Chapter 2, the population is not monotone increasing or monotone decreasing. We consider important features such as the probability of the population becoming extinct (birth-death model), and consider equilibrium distributions (immigration-death model). We discover an interesting relationship between the birth and death process and the random walk.

Chapter 4, *Queues.* We meet queues in many aspects of our lives: banks, telephone systems, etc. This chapter introduces a way of modelling queues mathematically using three key features: the method of arrival (usually a Poisson process), the distribution of the time it takes to serve a customer (e.g. exponential, constant), and the number of servers. Several different types of queue are considered. We ask questions such as the following.

- What is the average queue size?
- How long does a customer have to wait?
- What proportion of the time will a server be idle?

Chapter 5, *Renewal processes.* We model a sequence of time points separated by independent identically distributed waiting times. For example, a machine has a vital component which must be replaced every time it burns out, a point of the process being the failure of this component.

- How many such events do we expect in a given length of time?

We model discrete and continuous processes. We consider the equilibrium renewal process:

- what is the expected total lifetime of the component currently in use if we arrive at the machine at a random time?

- How much longer do we expect this component to last?

Chapter 6, *Epidemics*. Epidemics have had catastrophic effects throughout history. We consider two models of epidemics, the simple epidemic and the general epidemic. For the general epidemic,

- what is the expected number of people to catch the disease?
- What is the probability that everyone catches it?

Chapter 1: Poisson processes

1.1 Assumptions of the Poisson process

We model a situation where the time instants that we're interested in occur spontaneously, at random. Examples are

- the emissions of α-particles in a radioactive experiment,
- arrivals of customers at a post office,
- the passing of cars on a quiet road.

The important feature of these time instants or time points is that they are unpredictable (they occur 'at random').

How do we model such a process? Modelling Bernoulli trials, we assume that there is a constant probability of success and that events are independent. We make these assumptions in this case, too. Here we are modelling a situation in continuous time, where the time instants or *points* of the process can occur at any time. Thus these assumptions become:

(a) the average rate λ at which points occur is constant over time (not true for post office customers over a whole day, but good enough for a half hour period);
(b) the occurrence of points after time t is independent of what happened up to time t;
(c) we also assume that points can only occur singly (never ≥ 2 simultaneously).

Consider what happens in a short time interval of length δt. Suppose that the probability of one point in $(t, t+\delta t]$ is $p(\delta t)$.

The rate of occurrence of points is λ, and so

$$\frac{p(\delta t)}{\delta t} \to \lambda \qquad (\delta t \downarrow 0).$$

This says $p(\delta t) = \lambda\delta t + \mathrm{o}(\delta t)$, where

$$\frac{\mathrm{o}(\delta t)}{\delta t} \to 0 \qquad (\delta t \downarrow 0).$$

In general we shall use $\mathrm{o}(\delta t)$ for any function for which this is true. For instance $2(\delta t)^2$ is $\mathrm{o}(\delta t)$. This useful concept occurs throughout the course.

Definition 1.1.1. *A Poisson process of rate λ is a process $X = (X(t))_{t\geq 0}$ taking values in $S = \{0, 1, 2, \ldots\}$ such that*

(a) $X(0) = 0$; if $s < t$ then $X(s) \leq X(t)$;
(b) as $\delta t \downarrow 0$,

$$P(X(t+\delta t) = x+m | X(t) = x) = \begin{cases} \lambda\delta t + \mathrm{o}(\delta t) & \text{if } m = 1, \\ \mathrm{o}(\delta t) & \text{if } m > 1, \\ 1 - \lambda\delta t + \mathrm{o}(\delta t) & \text{if } m = 0. \end{cases}$$

(c) if $s < t$ then $X(t) - X(s)$, the number of points in the time interval $(s, t]$, is independent of the process prior to s.

We will consider two random variables:

- $X(t)$, the number of points that occur in $(0, t]$;
- T, the time until the first point occurs.

1.2 The probability function of $X(t)$

For simplicity, we shall write $p_x(t)$ for $P(X(t) = x)$.

Theorem 1.2.1. *$X(t)$ has a Poisson distribution with parameter λt.*

Proof.

$$p_0(t+\delta t)$$

$$= P(\text{no points by time } t + \delta t)$$
$$= P(\text{no points in } (0,t])P(\text{no points in } (t, t+\delta t])$$
$$= p_0(t)\big(1 - \lambda\delta t + \mathrm{o}(\delta t)\big).$$

Therefore

$$\frac{p_0(t+\delta t) - p_0(t)}{\delta t} = -\lambda p_0(t) + \frac{\mathrm{o}(\delta t)}{\delta t}.$$

Letting $\delta t \downarrow 0$, we obtain

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

Thus

$$\frac{p_0'(t)}{p_0(t)} = -\lambda;$$
$$\therefore \qquad \ln p_0(t) = -\lambda t + c.$$

And

$$p_0(0) = 1 \implies c = 0 \implies \ln p_0(t) = -\lambda t;$$

i.e. $p_0(t) = e^{-\lambda t}$. So the probability of $X(t)$ taking the value 0 is that from the required Poisson distribution.

In general,

$$p_x(t+\delta t)$$
$$= P(x \text{ points by time } t + \delta t)$$
$$= P(x \text{ points in } (0,t])P(\text{no points in } (t, t+\delta t])$$
$$+ P(x-1 \text{ points in } (0,t])P(\text{one point in } (t, t+\delta t])$$
$$+ P(x-2 \text{ points in } (0,t])P(\text{two points in } (t, t+\delta t])$$
$$+ \dots$$
$$= p_x(t)\big(1 - \lambda\delta t + \mathrm{o}(\delta t)\big) + p_{x-1}(t)\big(\lambda\delta t + \mathrm{o}(\delta t)\big) + \mathrm{o}(\delta t).$$

Therefore

$$\frac{p_x(t+\delta t) - p_x(t)}{\delta t} = \lambda\big(p_{x-1}(t) - p_x(t)\big) + \frac{\mathrm{o}(\delta t)}{\delta t}.$$

Letting $\delta t \downarrow 0$, we obtain

$$p_x'(t) = \lambda\big(p_{x-1}(t) - p_x(t)\big).$$

We know that $p_x(t)$ is of the right form for $x = 0$. Assume that it is also in the right form for $x - 1$, i.e.

$$p_{x-1}(t) = e^{-\lambda t}\frac{(\lambda t)^{x-1}}{(x-1)!}.$$

Thus

$$p_x'(t) + \lambda p_x(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{x-1}}{(x-1)!}.$$

Multiply by the integrating factor $e^{\lambda t}$:

$$e^{\lambda t}p_x'(t) + \lambda e^{\lambda t}p_x(t) = \lambda^x \frac{t^{x-1}}{(x-1)!}.$$

So

$$p_x(t)e^{\lambda t} = \lambda^x \frac{t^x}{x!} + C;$$
$$p_x(t) = \frac{(\lambda t)^x}{x!}e^{-\lambda t} + Ce^{-\lambda t}.$$

Now $p_x(0) = 0$ for $x > 0$, so $C = 0$. Thus we have shown by induction that

$$p_x(t) = e^{-\lambda t}\frac{(\lambda t)^x}{x!} \qquad (x = 0, 1, 2, \dots).$$

□

EXAMPLE 1.2.2. *Fax messages arrive at an office according to a Poisson process at mean rate three per hour.*

(a) What is the probability that exactly two messages are received between 9.00 and 9.40?
(b) What is the probability that no messages arrive between 10.00 and 10.30?
(c) What is the probability that not more than three messages are received between 10.00 and 12.00?

SOLUTION. Let an hour be the unit of time that we work with. Thus we have a Poisson process of rate $\lambda = 3$. The number of messages in time t is thus $\text{Pois}(3t)$ distributed.

(a) 9.00 to 9.40: the number of messages has distribution $\text{Pois}\big(3 \times \frac{2}{3}\big) = \text{Pois}(2)$, so

$$P(2 \text{ messages}) = e^{-2}\frac{2^2}{2!} \doteq 0{\cdot}271.$$

(b) 10.00 to 10.30: the number of messages has distribution $\text{Pois}(3 \times 0{\cdot}5) = \text{Pois}(1{\cdot}5)$, so

$$P(0 \text{ messages}) = e^{-1{\cdot}5} \doteq 0{\cdot}223.$$

(c) 10.00 to 12.00: the number of messages, N say, has distribution $\text{Pois}(3 \times 2) = \text{Pois}(6)$, so

$$\begin{aligned} &P(N \le 3) \\ &\quad = P(N=0) + P(N=1) + P(N=2) + P(N=3) \\ &\quad = e^{-6}\frac{6^0}{0!} + e^{-6}\frac{6^1}{1!} + e^{-6}\frac{6^2}{2!} + e^{-6}\frac{6^3}{3!} \\ &\quad = e^{-6}(1+6+18+36) = 61e^{-6} \doteq 0{\cdot}151. \end{aligned}$$

□

1.3 The distribution function of T

T is the time until the next point of the process from a given starting time. As the process is 'memoryless' it does not matter when this starting time is. T is a continuous random variable, unlike $X(t)$, which is discrete.

Consider the event $\{T > t\}$, i.e. the event that T is larger than some specified t. This is identical to $\{X(t) = 0\}$, the event that there is no point of the process up to time t. Thus

$$P(T > t) = P(X(t) = 0) = e^{-\lambda t}.$$

The distribution function of T is thus $P(T \le t) = 1 - e^{-\lambda t}$, i.e. T has an exponential distribution with parameter λ.

EXAMPLE 1.3.1. *In Example 1.2.2, what is the probability that the first message after 10.00 occurs by 11.00?*

SOLUTION. We have a Poisson process of rate 3 per hour and want the probability that the interval from our chosen starting time until the first point of the process is at most 1 hour, so

$$P(T \le 1) = 1 - e^{-3\times 1} \doteq 0{\cdot}9502.$$

□

1.4 The pooled Poisson process

Suppose that a bank branch has k service points in use and operates with a single long queue, so that when a server finishes serving a customer the person at the head of the queue replaces that customer. Further suppose that each service time is exponential with parameter λ, independently of all other service times.

The time that a person spends at the head of the queue depends upon k Poisson processes each with parameter λ.

FIGURE

Let T be the length of the waiting time for the person at the front of the queue until service starts. If T_i is the time until the person currently at server i is served, then $T = \min(T_1, \ldots, T_k)$. So

$$\begin{aligned} P(T > t) &= P(\min(T_1, \ldots, T_k) > t) \\ &= P(T_1 > t, T_2 > t, \ldots, T_k > t) \\ &= P(T_1 > t)P(T_2 > t) \cdots P(T_k > t) \\ &= e^{-\lambda t} \times \cdots \times e^{-\lambda t} = e^{-k\lambda t}, \end{aligned}$$

i.e. $P(T \leq t) = 1 - e^{-k\lambda t}$. Thus T is exponential with parameter $k\lambda$. Equivalently T is the time to the first point of a Poisson process of rate $k\lambda$.

DEFINITION 1.4.1. *k Poisson processes are said to be independent if the points in one process during any interval are independent of the points in any of the other processes during any intervals.*

Consider k independent Poisson processes, where process i has rate λ_i. We shall use the original formulation of the Poisson process to show that if we consider all the points they occur according to a single *pooled* process.

THEOREM 1.4.2. *If k independent Poisson processes with rates $\lambda_1, \ldots, \lambda_k$ occur simultaneously, the combined points follow a Poisson process with rate $\lambda_1 + \cdots + \lambda_k$.*

PROOF. In the interval $(t, t + \delta t]$,

$$\begin{aligned} P(\text{one point of type } i \text{ occurs}) &= \lambda_i \delta t + \mathrm{o}(\delta t); \\ P(\text{more than one point of type } i \text{ occurs}) &= \mathrm{o}(\delta t); \\ P(\text{no points of type } i \text{ occur}) &= 1 - \lambda_i \delta t + \mathrm{o}(\delta t). \end{aligned}$$

Combining these events, no points occur if and only if no points of type i occur for all i, i.e.

$$\begin{aligned} &P(\text{no point}) \\ &\quad = \big(1 - \lambda_1 \delta t + \mathrm{o}(\delta t)\big) \cdots \big(1 - \lambda_k \delta t + \mathrm{o}(\delta t)\big) \\ &\quad = 1 - \lambda_1 \delta t - \cdots - \lambda_k \delta t + \mathrm{o}(\delta t) \\ &\quad = 1 - (\lambda_1 + \cdots + \lambda_k)\delta t + \mathrm{o}(\delta t). \end{aligned}$$

One point of any type can occur in any of k ways: 1 point of type i and no points of any other type, for each i, i.e.

$$\begin{aligned} &P(\text{one point}) \\ &\quad = \big(\lambda_1 \delta t + \mathrm{o}(\delta t)\big)\big(1 - \lambda_2 \delta t + \mathrm{o}(\delta t)\big) \cdots \big(1 - \lambda_k \delta t + \mathrm{o}(\delta t)\big) \\ &\qquad + \big(1 - \lambda_1 \delta t + \mathrm{o}(\delta t)\big)\big(\lambda_2 \delta t + \mathrm{o}(\delta t)\big) \cdots \big(1 - \lambda_k \delta t + \mathrm{o}(\delta t)\big) \\ &\qquad + \cdots \\ &\qquad + \big(1 - \lambda_1 \delta t + \mathrm{o}(\delta t)\big)\big(1 - \lambda_2 \delta t + \mathrm{o}(\delta t)\big) \cdots \big(\lambda_k \delta t + \mathrm{o}(\delta t)\big) \\ &\quad = \lambda_1 \delta t + \cdots + \lambda_k \delta t + \mathrm{o}(\delta t) \\ &\quad = (\lambda_1 + \cdots + \lambda_k)\delta t + \mathrm{o}(\delta t). \end{aligned}$$

Finally the probability that more than one point occurs is one minus the probability that either no points or one point occurs, i.e.

$$\begin{aligned} &P(\text{more than one point}) \\ &\quad = 1 - \big(1 - (\lambda_1 + \cdots + \lambda_k)\delta t + \mathrm{o}(\delta t)\big) \end{aligned}$$

$$
\begin{aligned}
&- \big((\lambda_1 + \cdots + \lambda_k)\delta t + o(\delta t)\big) \\
&= o(\delta t).
\end{aligned}
$$

Thus we have satisfied condition (b) in our definition of a Poisson process of rate $\lambda_1 + \cdots + \lambda_k$. Condition (a) is satisfied trivially, and since all the individual processes are independent and obey (c) the combined process also obeys (c). □

EXAMPLE 1.4.3. *In the same office as in the previous two examples, telephone messages arrive at the mean rate of six per hour.*

(a) *Find the probability that exactly two messages (phone or fax) are received between 9.00 and 9.40.*
(b) *Find the probability that the first message after 10.00 occurs before 10.10.*

SOLUTION.

(a) We have two independent Poisson processes with rates 3 and 6 respectively, i.e. the pooled process has rate $3+6=9$. So in a 40-minute period, the number of calls follows a Poisson distribution with parameter

$$9 \times \frac{2}{3} = 6.$$

Therefore

$$P(X=2) = e^{-6}\frac{6^2}{2!} \simeq 0{\cdot}0446.$$

(b)

$$
\begin{aligned}
&P(\text{first message after 10.00 is before 10.10}) \\
&= P(T \le 1/6) \\
&= 1 - e^{-9\times 1/6} = 1 - e^{-3/2} \simeq 0{\cdot}7769.
\end{aligned}
$$

□

1.5 Breaking down a Poisson process

Suppose that a Poisson process is occurring and that a point can be of k different types. For example, traffic passes us as a Poisson process and we record whether each car is European, American, Japanese or some other.

Suppose that we know that the probability that a point is of type i is p_i for $i = 1, \ldots, k$, where $\sum p_i = 1$. Assume that the type of any given point is independent of the type of all other points.

Consider the occurrence of points of type i:

$$
\begin{aligned}
&P(\text{1 point occurs in } (t, t+\delta t], \text{ and it's of type } i) \\
&\quad = P(\text{1 point occurs})P(\text{point is of type } i|\text{1 point occurs}) \\
&\quad = \big(\lambda\delta t + o(\delta t)\big)p_i.
\end{aligned}
$$

Now for one point of type i to occur in $(t, t+\delta t]$, either it's the only point in the time-interval, as above, or it's one of several points of the overall Poisson process to occur. The latter event has probability $o(\delta t)$. So

$$
\begin{aligned}
&P(\text{1 point of type } i \text{ occurs in } (t, t+\delta t]) \\
&\quad = \big(\lambda\delta t + o(\delta t)\big)p_i + o(\delta t) \\
&\quad = p_i\lambda\delta t + o(\delta t).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&P(\text{more than 1 point of type } i \text{ occurs in } (t, t+\delta t]) \\
&\quad \le P(\text{more than 1 point occurs in } (t, t+\delta t]) = o(\delta t),
\end{aligned}
$$

and finally

$$\begin{aligned}
&P(0 \text{ points of type } i \text{ occur in } (t, t+\delta t]) \\
&\quad = 1 - \big(p_i\lambda\delta t + \mathrm{o}(\delta t)\big) - \mathrm{o}(\delta t) \\
&\quad = 1 - p_i\lambda\delta t + \mathrm{o}(\delta t).
\end{aligned}$$

Thus points of type i occur according to a Poisson process with rate λp_i. The single Poisson process is thus equivalent to a pooled process of k independent Poisson processes.

EXAMPLE 1.5.1. *Customers arrive at a bank according to a Poisson process at a mean rate of ten per minute. A proportion* 0·6 *wish to draw out money (type A),* 0·3 *wish to pay in money (type B) and* 0·1 *wish to do something else (type C).*

(a) *What is the probability that at least 5 customers arrive in 30 seconds?*
(b) *What is the probability that in one minute, six customers of type A, three of type B and one of type C arrive?*
(c) *If twenty customers arrive in two minutes, what is the probability that just one is of type C?*
(d) *What is the probability that the first three customers of the day want to draw out money?*
(e) *How long a time must elapse before there is a probability of* 0·9 *that at least one customer each of types A and B will have arrived?*

SOLUTION.

(a) The number of customers, N say, is Poisson with parameter $10 \times 0{\cdot}5 = 5$, so

$$P(N \geq 5) = 1 - \sum_{i=0}^{4} P(N = i)$$

$$= 1 - \sum_{i=0}^{4} \frac{5^i}{i!} e^{-5} = 1 - \frac{523}{8} e^{-5} \simeq 0{\cdot}5595.$$

(b) Customers of type A arrive according to a Poisson process of rate $0{\cdot}6 \times 10 = 6$. The number of customers in a minute is thus Poisson parameter 6. Similarly for type B the number of customers is Poisson with parameter 3 and for type C the number of customers is Poisson with parameter 1. The probability of 6 As, 3 Bs and 1 C is then

$$\begin{aligned}
P(6A,\ 3B,\ 1C) &= e^{-6}\frac{6^6}{6!} e^{-3}\frac{3^3}{3!} e^{-1}\frac{1^1}{1!} \\
&= \frac{1458}{5} e^{-10} \simeq 0{\cdot}0132.
\end{aligned}$$

(c) Twenty points have occurred, and the time of their occurrence is irrelevant. Each customer has a probability of 0·1 of being of type C. Thus we have a Binomial distribution, and

$$P(1 \text{ C out of } 20) = \binom{20}{1}(0{\cdot}1)(0{\cdot}9)^{19} \simeq 0{\cdot}270.$$

(d) The probability of any particular customer being of type A is 0·6. Thus the probability that the first three are of type A is $(0{\cdot}6)^3 = 0{\cdot}216$.
(e) Let t denote the required time. The two processes are independent, so that

$$\begin{aligned}
P(\geq 1A,\ \geq 1B) &= P(\geq 1A)P(\geq 1B) \\
&= (1 - e^{-6t})(1 - e^{-3t}).
\end{aligned}$$

Letting $y = e^{-3t}$, we need to solve

$$(1 - y^2)(1 - y) = 0{\cdot}9;$$

i.e.

$$y^3 - y^2 - y + 0{\cdot}1 = 0.$$

A solution may be found iteratively using, for example, the Newton-Raphson method, giving $y \simeq 0{\cdot}092$, i.e. $t \simeq 0{\cdot}795$ minutes (about 48 seconds).

□

Chapter 2: Birth processes

2.1 The simple birth process

Consider a population where each individual alive in the population generates further offspring according to a Poisson process at rate β (new individuals are produced asexually). We assume that the initial population size is x_0 and that there are no deaths, so that the population increases with time.

This is not a very realistic model, but it serves as a starting point. If the size of the population at time t is x, then we have x different Poisson processes each with rate β. When considering the next birth we thus have a pooled Poisson process with rate $x\beta$ (until the next arrival, when this becomes a Poisson process with rate $(x+1)\beta$). So we have

$$\begin{aligned} P(X(t+\delta t) = x+1 | X(t) = x) &= \beta x \delta t + \mathrm{o}(\delta t); \\ P(X(t+\delta t) = x | X(t) = x) &= 1 - \beta x \delta t + \mathrm{o}(\delta t); \\ P(X(t+\delta t) = y | X(t) = x) &= \mathrm{o}(\delta t) \quad \text{for other } y, \end{aligned}$$

where $X(t)$ is the population size at time t.

To find the distribution of $X(t)$ we proceed as before and find differential-difference equations. Note that at $t = 0$ there must be at least one member of the population, otherwise $X(t) = 0$ for all t, and since there are no deaths, $X(t) \geq 1$ for all t. Again letting $P(X(t) = x)$ be represented by $p_x(t)$,

$$p_x(t+\delta t) = \sum_{y=x_0}^{x} P(X(t+\delta t) = x | X(t) = y) P(X(t) = y),$$

since $X(t)$ is an increasing function which is never 0. Thus

$$p_{x_0}(t+\delta t) = p_{x_0}(t)(1 - \beta x_0 \delta t + \mathrm{o}(\delta t));$$

$$\therefore \qquad p'_{x_0}(t) = -\beta x_0 p_{x_0}(t). \tag{2.1.1}$$

So $\ln p_{x_0}(t) = -\beta x_0 t + c$. Since the initial population size is x_0, $p_{x_0}(0) = 1$, and so $c = 0$, giving

$$p_{x_0}(t) = e^{-\beta x_0 t}.$$

For general x, $P(X(t+\delta t) = x | X(t) = y) = \mathrm{o}(\delta t)$ if y is not x or $x-1$, i.e.

$$\begin{aligned} p_x(t+\delta t) = p_x(t)\big(1 - \beta x \delta t + \mathrm{o}(\delta t)\big) \\ + p_{x-1}(t)\big(\beta(x-1)\delta t + \mathrm{o}(\delta t)\big) + \mathrm{o}(\delta t), \end{aligned}$$

and so

$$\begin{aligned} \frac{p_x(t+\delta t) - p_x(t)}{\delta t} &= -\beta x p_x(t) + \beta(x-1) p_{x-1}(t) + \frac{\mathrm{o}(\delta t)}{\delta t}; \\ \therefore \quad p'_x(t) &= -\beta x p_x(t) + \beta(x-1) p_{x-1}(t) \\ & \qquad (x = x_0, x_0+1, \ldots). \end{aligned} \tag{2.1.2}$$

THEOREM 2.1.1. *The distribution of $X(t)$ is given by*

$$p_x(t) = \binom{x-1}{x_0-1} e^{-\beta t x_0} (1 - e^{-\beta t})^{x - x_0} \quad (x = x_0, x_0+1, \ldots).$$

This is sometimes called a ‘negative binomial distribution with scale parameter $e^{-\beta t}$ and index parameter x_0’. However the name is better used for the distribution of $X(t) - x_0$, taking values in $\{0, 1, 2, \ldots\}$.

If $x_0 = 1$, the distribution of $X(t)$ becomes

$$p_x(t) = e^{-\beta t}(1 - e^{-\beta t})^{x-1} \qquad (x = 1, 2, \ldots),$$

sometimes called the ‘geometric distribution with parameter $e^{-\beta t}$’. However, again the name is better used for the distribution of $X(t) - 1$, taking values in $\{0, 1, 2, \ldots\}$.

PROOF. We know that $p_{x_0}(t) = e^{-\beta x_0 t}$, which is of the correct form for $x = x_0$. We again suppose that the solution is of the correct form for $x - 1$, i.e.

$$p_{x-1}(t) = \binom{x-2}{x_0-1} e^{-\beta t x_0}(1 - e^{-\beta t})^{x-1-x_0}.$$

This gives

$$\begin{aligned} &p'_x(t) + \beta x p_x(t) \\ &\quad = \beta(x-1)\frac{(x-2)!}{(x_0-1)!(x-x_0-1)!}e^{-\beta t x_0}(1-e^{-\beta t})^{x-x_0-1}. \end{aligned}$$

Multiplying by the integrating factor $e^{\beta x t}$ we obtain

$$\begin{aligned} &\frac{d}{dt}\left(p_x(t)e^{\beta x t}\right) \\ &\quad = \frac{(x-1)!}{(x_0-1)!(x-x_0-1)!}\beta e^{\beta t(x-x_0)}(1-e^{-\beta t})^{x-x_0-1}, \end{aligned}$$

so

$$\begin{aligned} &p_x(t)e^{\beta x t} \\ &= \frac{(x-1)!}{(x_0-1)!(x-x_0-1)!}\int \beta e^{\beta t}(e^{\beta t}-1)^{x-x_0-1}\,dt + A \\ &= \frac{(x-1)!}{(x_0-1)!(x-x_0)!}\int (x-x_0)(e^{\beta t}-1)^{x-x_0-1}\beta e^{\beta t}\,dt + A \\ &= \binom{x-1}{x_0-1}(e^{\beta t}-1)^{x-x_0} + A. \end{aligned}$$

Therefore

$$p_x(t) = \binom{x-1}{x_0-1}e^{-\beta t x}(e^{\beta t}-1)^{x-x_0} + Ae^{-\beta t x},$$

and $A = 0$ since $p_x(0) = 0$ for $x > x_0$. The theorem is proved by induction. □

EXAMPLE 2.1.2. *A population starts at time 0 with a single individual. Let the birth rate be two per week.*

(a) What is the probability that after three weeks there are exactly two individuals?
(b) What is the probability that after one week there are between two and four individuals (inclusive)?

SOLUTION.

(a) $x_0 = 1$, $\beta = 2$ per week, and $t = 3$, i.e.

$$\begin{aligned} p_2(3) &= \binom{1}{0}e^{-2\times 3}(1-e^{-2\times 3})^1 \\ &= e^{-6}(1-e^{-6}) \simeq 0{\cdot}00247. \end{aligned}$$

(b) $x_0 = 1$, $\beta = 2$, $t = 1$, so

$$\begin{aligned} &P(2 \le X \le 4) \\ &\quad = P(X=2) + P(X=3) + P(X=4) \\ &\quad = e^{-2}(1-e^{-2}) + e^{-2}(1-e^{-2})^2 + e^{-2}(1-e^{-2})^3 \\ &\quad \simeq 0{\cdot}3057. \end{aligned}$$

□

2.2 The pure death process

We consider a population in which there are no births, just deaths. Observations start with x_0 individuals alive at time 0—these individuals die independently of each other, and eventually the population dies out completely.

In a similar way to the simple birth process, we assume that the probability of an individual dying in time interval $(t, t+\delta t]$ is $\nu\delta t + \mathrm{o}(\delta t)$.

Some questions of interest are

- What is the distribution of the population size at time t?
- How long does it take the population to die out?

This model is approached best by considering every individual separately. The probability that a given individual is alive at time t, which we label $P_a(t)$, is found as follows. The probability that the individual, if alive at time t, is still alive at time $t+\delta t$ is one minus the probability that it dies in this interval, so that

$$\begin{aligned} P(\text{alive at } t+\delta t|\text{alive at } t) &= 1-\nu\delta t+\mathrm{o}(\delta t); \\ \therefore \qquad P_a(t+\delta t) &= P_a(t)\big(1-\nu\delta t+\mathrm{o}(\delta t)\big); \\ \therefore \qquad P_a'(t) &= -\nu P_a(t). \end{aligned}$$

This differential equation has solution $P_a(t) = Ae^{-\nu t} = e^{-\nu t}$ since the individual is alive at time 0 with probability 1. We can use the binomial theorem to deduce that the probability that j individuals are still alive at time t is given by

$$p_j(t) = \binom{x_0}{j}(e^{-\nu t})^j(1-e^{-\nu t})^{x_0-j}.$$

In particular the probability that the population is extinct by time t is

$$p_0(t) = (1-e^{-\nu t})^{x_0}.$$

EXAMPLE 2.2.1. *A population starts at time 0 with 4 individuals. The population follows a pure death process at a rate of 1 every 2 days.*

(a) Find the probability that there is exactly one individual alive after a week.

(b) Find the probability that the population has died out after a week.

(c) Find the probability that the population has died out after two weeks, given that the total number of survivors after one week was 2.

SOLUTION. $x_0 = 4$, $\nu = 0{\cdot}5$ per day.

(a) $t = 7$, so

$$p_1(7) = \binom{4}{1}e^{-3\cdot5}(1-e^{-3\cdot5})^3 \simeq 0{\cdot}1102.$$

(b) Again $t = 7$, so

$$p_0(7) = (1-e^{-3\cdot5})^4 \simeq 0{\cdot}8846.$$

(c) The process is memoryless, so that

$$\begin{aligned} &P(0 \text{ after 2 weeks}|2 \text{ after 1 week}) \\ &\quad = P(0 \text{ after 1 week}|2 \text{ after 0 weeks}) \\ &\quad = (1-e^{-3\cdot5})^2 \simeq 0{\cdot}9405. \end{aligned}$$

□

Chapter 3: Birth and death processes

3.1 Introduction

We shall now consider a population model, similar to those of the last chapter, but with both births and deaths. In the Poisson process and the simple birth process the only change possible is an increase in the population. In the pure death process the population could only be reduced. Here both may occur.

Let $p_{y,x}(t, t+\delta t]$ be the probability that the population size changes from y at time t to x at time $t+\delta t$. This probability is negligible ($\mathrm{o}(\delta t)$) unless $y = x$ or $x \pm 1$, since the probability of more than one point of the process occurring is negligible.

Let the birth rate at $X = x$ be β_x, and the death rate be ν_x. Both β_x and ν_x depend upon x. Thus

$$\begin{aligned} p_{x-1,x}(t, t+\delta t] &= \beta_{x-1}\delta t + \mathrm{o}(\delta t); \\ p_{x+1,x}(t, t+\delta t] &= \nu_{x+1}\delta t + \mathrm{o}(\delta t); \\ p_{x,x}(t, t+\delta t] &= 1 - \beta_x \delta t - \nu_x \delta t + \mathrm{o}(\delta t). \end{aligned} \tag{3.1.1}$$

Note that all of the models that we have considered up until now are special cases of the birth-death process:

- Poisson process $\beta_x = \lambda$, $\nu_x = 0$;
- Simple birth process $\beta_x = \beta x$, $\nu_x = 0$;
- Pure death process $\beta_x = 0$, $\nu_x = \nu x$.

Thus in all the previous models either β_x or ν_x is zero.

3.2 The Kolmogorov equations

Remember that we define the function $p_x(t) = P(X(t) = x)$. For each value of x,

$$P\big(X(t+\delta t) = x\big)$$
$$= \sum_{k=0}^{\infty} P(X(t) = k) P(X(t+\delta t) = x | X(t) = k),$$

or in our simpler notation,

$$p_x(t+\delta t) = \sum_{k=0}^{\infty} p_k(t) p_{k,x}(t, t+\delta t].$$

These are the *Chapman-Kolmogorov equations.*

Substituting (3.1.1) in them gives

$$\begin{aligned} p_x(t+\delta t) &= p_{x-1}(t)\beta_{x-1}\delta t \\ &\quad + p_x(t)(1 - \beta_x\delta t - \nu_x\delta t) + p_{x+1}(t)\nu_{x+1}\delta t + \mathrm{o}(\delta t), \end{aligned}$$

whence

$$\begin{aligned} &\frac{p_x(t+\delta t) - p_x(t)}{\delta t} \\ &\quad = p_{x-1}(t)\beta_{x-1} - p_x(t)(\beta_x + \nu_x) + p_{x+1}(t)\nu_{x+1} + \frac{\mathrm{o}(\delta t)}{\delta t}. \end{aligned}$$

Therefore

$$\begin{aligned} p_x'(t) &= p_{x-1}(t)\beta_{x-1} - p_x(t)(\beta_x + \nu_x) + p_{x+1}(t)\nu_{x+1} \\ &\quad (x = 0, 1, 2, \ldots). \end{aligned}$$

Note that we define $p_{-1}(t) = 0$. This term occurs in the equation for $x = 0$, but isn't really there.

These are the *Kolmogorov forward equations* (there are also the *Kolmogorov backward equations*, but we shall not discuss them in this course).

By assigning values to β_x and ν_x we can obtain specific differential-difference equations for any particular process of this type.

EXAMPLE 3.2.1. *What are the Kolmogorov forward equations for*

(a) the Poisson process?
(b) the simple birth process?

SOLUTION.

(a) $\beta_x = \lambda$, $\nu_x = 0$, $p_0(0) = 1$. Thus

$$p_0'(t) = -\lambda p_0(t),$$
$$p_x'(t) = \lambda p_{x-1}(t) - \lambda p_x(t) \qquad (x = 1, 2, \ldots),$$

which are indeed the equations that we found in Chapter 1.

(b) $\beta_x = \beta x$, $\nu_x = 0$, $p_{x_0}(0) = 1$, so

$$p_{x_0}'(t) = -\beta x_0 p_{x_0}(t),$$
$$p_x'(t) = \beta(x-1)p_{x-1}(t) - \beta x p_x(t) \qquad (x = x_0 + 1, x_0 + 2, \ldots),$$

which are again the equations found in Chapter 2; see (2.1.1) and (2.1.2).

□

3.3 The simple birth-death process

In the previous chapter we considered the simple birth process and the pure death process. Now we shall combine the two.

In the simple birth process, each individual gives birth at rate β, so that when the population is of size x, the birth rate is βx. In the pure death process individuals die at rate ν, so that the death rate is νx. We wish to find an expression for $X(t)$, the number of individuals alive at time t. We shall find the probability generating function for $X(t)$.

Substituting βx for β_x and νx for ν_x in the Kolmogorov equations gives

$$p_0'(t) = \nu p_1(t).$$
$$p_x'(t) = p_{x-1}(t)\beta(x-1) - p_x(t)(\beta x + \nu x) + p_{x+1}(t)\nu(x+1) \qquad (x = 1, 2, \ldots),$$

Thus

$$\begin{aligned}\sum_{x=0}^{\infty} p_x'(t)s^x &= \sum_{x=1}^{\infty} \beta(x-1)p_{x-1}(t)s^x - \sum_{x=1}^{\infty}(\beta+\nu)xp_x(t)s^x \\ &\quad + \sum_{x=0}^{\infty}\nu(x+1)p_{x+1}(t)s^x \\ &= \sum_{y=0}^{\infty}\beta y p_y(t)s^{y+1} - \sum_{x=1}^{\infty}(\beta+\nu)xp_x(t)s^x \\ &\quad + \sum_{y=1}^{\infty}\nu y p_y(t)s^{y-1} \\ &= \beta s^2\sum_{x=1}^{\infty} xp_x(t)s^{x-1} - (\beta+\nu)s\sum_{x=1}^{\infty}xp_x(t)s^{x-1} \\ &\quad + \nu\sum_{x=1}^{\infty}xp_x(t)s^{x-1}.\end{aligned}$$

Now we introduce the probability generating function (p.g.f.) of $X(t)$:

$$\begin{aligned}\Pi(s,t) &= Es^{X(t)} \\ &= \sum_{x=0}^{\infty} s^x P(X(t) = x) = \sum_{x=0}^{\infty} s^x p_x(t) \quad (|s| \le 1).\end{aligned}$$

The above is then just

$$\frac{\partial \Pi}{\partial t} = \big(\beta s^2 - (\beta+\nu)s + \nu\big)\frac{\partial \Pi}{\partial s}.$$

We shall simply state the solution, which can be verified by substitution. There are two different cases;

$$\Pi(s,t) = \begin{cases} \left(\dfrac{\nu(1-s) - (\nu-\beta s)e^{(\nu-\beta)t}}{\beta(1-s) - (\nu-\beta s)e^{(\nu-\beta)t}}\right)^{x_0} & \text{if } \beta \neq \nu, \\ \left(\dfrac{\beta t - s\beta t + s}{\beta t - s\beta t + 1}\right)^{x_0} & \text{if } \beta = \nu. \end{cases}$$

Note that all the required information about $X(t)$ is in the p.g.f. and that no two discrete distributions can have the same p.g.f. We proceed to extract some of this information.

We can rewrite the probability generating function as

$$\Pi(s,t) = \left(\frac{a-bs}{c-ds}\right)^{x_0},$$

where a, b, c, d are all functions of t. When $x_0 = 1$ we have

$$\begin{aligned} \Pi(s,t) &= \frac{a-bs}{c-ds} \\ &= \frac{a-bs}{c}\left(1-\frac{ds}{c}\right)^{-1} \\ &= \frac{a-bs}{c}\sum_{i=0}^{\infty}\left(\frac{ds}{c}\right)^i \\ &= \frac{a}{c} + \sum_{i=0}^{\infty} s^{i+1}\left(-\frac{b}{c}\left(\frac{d}{c}\right)^i + \frac{a}{c}\left(\frac{d}{c}\right)^{i+1}\right) \\ &= \frac{a}{c} + \sum_{j=1}^{\infty} s^j \left(\frac{d}{c}\right)^{j-1}\frac{ad-bc}{c^2}. \end{aligned}$$

Thus, using the generating function to find the probabilities of the different population sizes,

$$P(X(t)=0) = \frac{a}{c},$$

$$P(X(t)=j) = \left(\frac{d}{c}\right)^{j-1}\frac{ad-bc}{c^2} \qquad (j = 1,2,\ldots),$$

which is similar to a geometric distribution, except that we have an added probability of 0.

Let us suppose that $\beta \neq \nu$. Letting $p = e^{(\nu-\beta)t}$, we have $a = \nu - \nu p$, $b = \nu - \beta p$, $c = \beta - \nu p$ and $d = \beta - \beta p$.

Exercise 3.3.1. *For both $\beta > \nu$ and $\beta < \nu$, show that*

$$0 < \frac{d}{c} < 1$$

for $t > 0$, and so the above expansion gives a valid distribution.

Exercise 3.3.2. *Show that $ad - bc = p(\nu-\beta)^2$.*

Thus we have

$$p_0(t) = \frac{\nu(1-p)}{\beta - \nu p}, \tag{3.3.1}$$

$$p_x(t) = \frac{\big(\beta(1-p)\big)^{x-1}p(\nu-\beta)^2}{(\beta-\nu p)^{x+1}} \quad (x = 1,2,\ldots). \tag{3.3.2}$$

If $x_0 > 1$ then each individual can be thought of as the founder of its own dynasty, and so $X(t)$ is the sum of x_0 independent observations from the above distribution.

If $\beta = \nu$ then when $x_0 = 1$,

$$\Pi(s,t) = \frac{\beta t - s\beta t + s}{\beta t - s\beta t + 1} = \frac{a-bs}{c-ds},$$

and so $a = \beta t$, $b = \beta t - 1$, $c = 1 + \beta t$, $d = \beta t$, and again $c > d$, giving

$$p_0(t) = \frac{\beta t}{1+\beta t}, \qquad p_x(t) = \frac{(\beta t)^{x-1}}{(1+\beta t)^{x+1}} \quad (x = 1, 2, \ldots).$$

EXAMPLE 3.3.3. *Find the mean and variance of a random variable with p.g.f.*

$$\Pi(s) = \frac{a - bs}{c - ds},$$

and hence find the mean and variance of the simple birth-death process with $\beta = \nu$.

SOLUTION.

$$\Pi'(s) = \frac{-(c-ds)b + (a-bs)d}{(c-ds)^2} = \frac{ad-bc}{(c-ds)^2},$$

$$\therefore \quad \Pi''(s) = \frac{2d(ad-bc)}{(c-ds)^3}.$$

Letting $s \uparrow 1$ we obtain

$$EX = \frac{ad-bc}{(c-d)^2},$$

and $\operatorname{var} X = \Pi''(1) + EX - (EX)^2$, so

$$\operatorname{var} X = \frac{2d(ad-bc)}{(c-d)^3} + \frac{ad-bc}{(c-d)^2} - \frac{(ad-bc)^2}{(c-d)^4}.$$

For the simple birth-death process with $\beta = \nu$, we have $ad - bc = 1$ and $c - d = 1$, so that $EX(t) = 1$ and $\operatorname{var} X(t) = 2d + 1 - 1 = 2\beta t$, if $x_0 = 1$. This implies that for a population starting with x_0 individuals,

$$EX(t) = x_0, \quad \operatorname{var} X(t) = 2x_0\beta t.$$

□

3.4 Simple birth-death: extinction

The probability of the population being extinct at time t is $P(X(t) = 0) = p_0(t)$, which is $\Pi(0, t)$ since

$$\Pi(s,t) = \sum_{x=0}^{\infty} p_x(t) s^x.$$

The probability that the population eventually becomes extinct is thus $\lim_{t\to\infty} p_0(t) = \lim_{t\to\infty} \Pi(0, t)$. Recall that if $\nu \neq \beta$ then

$$\Pi(s,t) = \left(\frac{\nu(1-s) - (\nu - \beta s)e^{(\nu-\beta)t}}{\beta(1-s) - (\nu-\beta s)e^{(\nu-\beta)t}} \right)^{x_0}.$$

If $\nu > \beta$, then as $t \to \infty$, $e^{(\nu-\beta)t} \to \infty$ and so $\Pi(0, t) \to 1$.

If $\nu < \beta$, then as $t \to \infty$, $e^{(\nu-\beta)t} \to 0$ and so

$$\Pi(0,t) \to \left(\frac{\nu}{\beta}\right)^{x_0}.$$

If $\nu = \beta$ then

$$\Pi(0,t) = \left(\frac{\beta t}{\beta t + 1}\right)^{x_0} = \left(\frac{\beta}{\beta + t^{-1}}\right)^{x_0} \to 1$$

as $t \to \infty$.

Thus if $\beta \le \nu$, the population is certain to become extinct.

Note that if $\beta > \nu$, letting $t \to \infty$ we obtain

$$\lim_{t\to\infty} \Pi(s,t) = \left(\frac{\nu}{\beta}\right)^{x_0}$$

for every value of $s \in [0, 1)$. But for a p.g.f., $\lim_{s\uparrow 1} \Pi(s) = 1$ and so this limit is not a proper p.g.f. (letting $s \to 1$ first gives a different result!). In fact the probability of the population having a finite size in the limit is the same as the probability of extinction. This is because as t tends to infinity, either the population goes extinct or becomes very large, and the probability of it taking any given finite value except 0 tends to zero.

We have considered the probability that a population becomes extinct, but also of interest is the time to extinction. A population may be certain to become extinct but have a high probability of lasting a long time before that extinction occurs. Letting T be the time to extinction,

$$P(T \le t) = P(X(t) = 0),$$

the probability that the population has no members at time t.

We know that if $x_0 = 1$, then

$$P(X(t) = 0) = p_0(t) = \frac{\nu(1-p)}{\beta - \nu p} \qquad (p = e^{(\nu-\beta)t},\ \beta \ne \nu).$$

Considering the population as the combination of x_0 different sub-populations starting with one individual, the probability that there are no individuals in the whole population is the probability that there are none in any of the sub-populations, which is to say

$$P(T \le t) = \left(\frac{\nu(1-p)}{\beta - \nu p}\right)^{x_0} = \left(\frac{\nu(1-e^{(\nu-\beta)t})}{\beta - \nu e^{(\nu-\beta)t}}\right)^{x_0}.$$

If $\beta < \nu$, $P(T \le t) \to 1$ as $t \to \infty$ so extinction is certain and T has a proper distribution.

If $\beta > \nu$, $P(T \le t) \to (\nu/\beta)^{x_0}$ as $t \to \infty$ so extinction is not certain and T does not have a proper distribution.

If $\beta = \nu$ then

$$P(T \le t) = \left(\frac{\beta t}{1 + \beta t}\right)^{x_0} \to 1 \qquad (t \to \infty).$$

It is easy to show that although extinction is certain in this case, the expected time to extinction is infinite.

Example 3.4.1. *If a population has birth rate $\beta = 3$, death rate $\nu = 2$ and starts with $x_0 = 4$ individuals, what is the probability that it will be extinct*

(a) by time 1?
(b) by time 3?
(c) at all?

Find $P(T \le 3 | \text{becomes extinct})$.

Solution.

(a)

$$P(T \le 1) = \left(\frac{2(1-e^{-1})}{3 - 2e^{-1}}\right)^4 \doteq 0{\cdot}0972.$$

(b)

$$P(T \le 3) = \left(\frac{2(1-e^{-3})}{3 - 2e^{-3}}\right)^4 \doteq 0{\cdot}1843.$$

(c)

$$P(\text{extinction}) = \left(\frac{2}{3}\right)^4 \doteq 0{\cdot}1975.$$

Finally,

$$P(T \le 3|\text{extinction}) = \frac{P(T \le 3 \text{ and extinction})}{P(\text{extinction})}$$
$$= \frac{P(T \le 3)}{P(\text{extinction})} \doteq \frac{0{\cdot}1843}{0{\cdot}1975} = 0{\cdot}933.$$

Thus if the population does become extinct, it is likely to happen early. □

3.5 An embedded process

Suppose that we are not interested in the time of a particular point of the process, but only in its type (is it a birth or a death?). Relabelling the time of the i^{th} point as i, we obtain a new random process $(X_i)_{i=0,1,2,\ldots}$ in discrete time. This process is said to be *embedded* in the original process, i.e. is an *embedded process*. X_i is the size of the embedded process at time i, and of the population immediately after the i^{th} change.

In this case all we are concerned with is the probability that the next point will be a birth. This is

$$\frac{\beta_x}{\beta_x + \nu_x} = \frac{\beta x}{\beta x + \nu x} = \frac{\beta}{\beta + \nu}.$$

We write

$$p = \frac{\beta}{\beta + \nu},\ q = 1 - p = \frac{\nu}{\beta + \nu}.$$

Thus we have a simple random walk with

$$P(X_i = x + 1|X_{i-1} = x) = p,\ P(X_i = x - 1|X_{i-1} = x) = q,$$

and an absorbing barrier at zero (which is a gambler's ruin problem). This means that we can use standard results such as the following.

- If the gambler starts with j units of money and has probability of winning each game p, then the probability that the gambler is ruined is

$$\begin{cases} \left(\dfrac{q}{p}\right)^j & \text{if } p > q, \\ 1 & \text{otherwise.} \end{cases}$$

This is equivalent to the probability of extinction with

$$j = x_0, \quad p = \frac{\beta}{\beta + \nu} \implies q = \frac{\nu}{\beta + \nu}.$$

Thus the probability of extinction is

$$\begin{cases} \left(\dfrac{\nu}{\beta}\right)^{x_0} & \text{if } \beta > \nu, \\ 1 & \text{otherwise.} \end{cases}$$

- The expected duration of the whole game (before the gambler is ruined) is

$$\begin{cases} \dfrac{j}{q - p} & \text{if } p < q, \\ \infty & \text{if } p \ge q. \end{cases}$$

So for a simple birth-death process starting with x_0 individuals, the expected number of points of the process to extinction is

$$\begin{cases} \dfrac{x_0(\nu + \beta)}{\nu - \beta} & \text{if } \beta < \nu, \\ \infty & \text{if } \beta = \nu. \end{cases}$$

It is obviously infinite for $\beta > \nu$, since there is a positive probability that the population will never go extinct.

- In the classical gambler's ruin problem, there are two players with £m between them, so that the probability that our gambler wins is the probability that (s)he reaches m before 0, which, starting at $j \le m$, is

$$\begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^j}{1-\left(\frac{q}{p}\right)^m} & \text{if } p \neq q, \\ \dfrac{j}{m} & \text{if } p = q. \end{cases}$$

This corresponds exactly to the probability that the size of our population reaches $m \ge x_0$ individuals at some point (before possibly becoming extinct), which thus has probability

$$\begin{cases} \dfrac{1-\left(\frac{\nu}{\beta}\right)^{x_0}}{1-\left(\frac{\nu}{\beta}\right)^m} & \text{if } \beta \neq \nu, \\ \dfrac{x_0}{m} & \text{if } \nu = \beta. \end{cases}$$

Example 3.5.1. *If a simple birth-death process starts with $x_0 = 5$ individuals, what is the probability that it reaches 10 given that it becomes extinct, in the cases*

(a) $\beta = 4$, $\nu = 6$?
(b) $\beta = 9$, $\nu = 6$?

Solution.

(a)

$$P(\text{reaches } 10) = \frac{1-\left(\frac{6}{4}\right)^5}{1-\left(\frac{6}{4}\right)^{10}} \simeq 0{\cdot}1164.$$

The process is certain to become extinct, so that

$$\begin{aligned} P(\text{reaches } 10|\text{becomes extinct}) &= P(\text{reaches } 10) \\ &\simeq 0{\cdot}1164. \end{aligned}$$

(b)

$$P(\text{reaches } 10) = \frac{1-\left(\frac{6}{9}\right)^5}{1-\left(\frac{6}{9}\right)^{10}} \simeq 0{\cdot}8836.$$

Now

$$P(\text{reaches } 10|\text{becomes extinct}) = \frac{P(\text{extinction}|\text{reaches } 10)P(\text{reaches } 10)}{P(\text{extinction})},$$

and

$$\begin{aligned} P(\text{extinction}) &= \left(\frac{6}{9}\right)^5 \simeq 0{\cdot}1317, \\ P(\text{extinction}|\text{reaches } 10) &= \left(\frac{6}{9}\right)^{10} \simeq 0{\cdot}0173, \end{aligned}$$

so

$$\begin{aligned} P(\text{reaches } 10|\text{becomes extinct}) &= \frac{0{\cdot}0173 \times 0{\cdot}8836}{0{\cdot}1317} \\ &= 0{\cdot}116. \end{aligned}$$

So if we know that the population becomes extinct, it is unlikely to have reached 10 before doing so.

□

3.6 The immigration-death model

In the immigration-death model, individuals join the population according to a Poisson process with parameter λ, and live an exponential amount of time with parameter ν. We shall only briefly look at this model.

It's a realistic model for some practical situations, e.g. the number of gas molecules in a particular volume of space, the number of people in a shop.

A feature of the model is that there is no extinction: new individuals will keep arriving. Similarly the population will not tend to infinity since $\nu_x = \nu x$ will eventually overpower $\beta_x = \lambda$ as x becomes large.

The Kolmogorov equations give

$$\begin{aligned} p_0'(t) &= \nu p_1(t) - \lambda p_0(t), \\ p_x'(t) &= \lambda p_{x-1}(t) - (\lambda + \nu x)p_x(t) + \nu(x+1)p_{x+1}(t) \\ &\qquad (x = 1, 2, \ldots). \end{aligned}$$

We can sensibly consider an equilibrium distribution for this model, since the population can neither become extinct nor tend to infinity. Setting

$$p_x'(t) = 0 \quad \forall x,$$

and letting p_x represent the equilibrium probability of the population being of size x, we obtain

$$\nu p_1 - \lambda p_0 = 0 \implies p_1 = \frac{\lambda}{\nu} p_0$$

and

$$\lambda p_{x-1} - (\lambda + \nu x)p_x + \nu(x+1)p_{x+1} = 0 \qquad (x = 1, 2, \ldots).$$

Note that if $\lambda p_{x-1} = \nu x p_x$ then $\lambda p_x = \nu(x+1)p_{x+1}$, which is the same expression with x replaced by $x+1$. Since this is true for $x = 1$, it is true for all x and so

$$p_x = \frac{\lambda}{\nu}\frac{p_{x-1}}{x} \implies p_x = \left(\frac{\lambda}{\nu}\right)^x \frac{p_0}{x!}.$$

We need this to be a proper probability distribution, i.e.

$$1 = \sum_{x=0}^{\infty} p_x = p_0 \sum_{x=0}^{\infty} \frac{1}{x!}\left(\frac{\lambda}{\nu}\right)^x = p_0 e^{\lambda/\nu}.$$

So

$$p_x = e^{-\lambda/\nu}\frac{1}{x!}\left(\frac{\lambda}{\nu}\right)^x \qquad (x = 0, 1, 2, \ldots),$$

which is a Poisson distribution with parameter λ/ν.

EXAMPLE 3.6.1. *For an immigration-death process with immigration rate 6 per week and death rate 10 per year, what are the mean and variance of the population size X?*

SOLUTION. $\lambda = 6 \times 52 = 312$ per year, $\nu = 10$ per year, so

$$\frac{\lambda}{\nu} = \frac{312}{10} = 31{\cdot}2.$$

Thus we have a Poisson distribution with parameter 31·2 so that the mean and variance of X are both 31·2. □

Chapter 4: Queues

4.1 Introduction

In many aspects of everyday life we encounter queues: telephone queueing systems (becoming increasingly common), bank/shop queues, traffic jams, queues of aircraft circling an airport, etc. Patients have to 'queue' to wait for an operation. Sometimes we can see the size of the queue, sometimes not. It may or may not be possible to make a good guess at the length of the time to wait. We may not have to wait at all, or the queue could be so long that we decide to give up and try again some other time.

From the point of view of the server/shop owner it might be important to consider breaks (when nobody is queueing) or the possibility that arrivals come too quickly to be served (i.e. the queue gets longer and longer and/or people give up).

The queueing process is unpredictable in that the rate of customer arrivals and the time it takes for a customer to be served are both random. It is of interest to model queues as a random process because they are common and because some of the parameters can be controlled, e.g. by varying the number of servers, so it would be useful to understand the consequences of such variations.

There are three features of a queue which we shall consider (in reality, of course, there are many):

- The arrival mechanism—how do customers arrive, singly or in groups, randomly or by appointment?
- The service time—constant or random, what distribution?
- The number of servers.

Another question is that of queue discipline. Are customers served in the actual order by which they arrive? If there is a single queue in a bank, the answer is usually yes, but in a pub there is often a random element (which person the barman sees first) or a not-so-random element (Joe might be served earlier because he's in there every day). We shall assume that customers are served *strictly in the order of arrivals.*

If there is more than one server, we assume a central queueing system, so that customers move forward as servers become free.

We shall consider only queueing models formed by varying the three features mentioned above. In general we assume that arrivals occur singly and at random.

A queue will be characterised as follows to describe its three features. Firstly we specify the inter-arrival time, e.g. Poisson process/exponential, written as M (for Markovian), deterministic (fixed interval) D, general (unspecified) G, etc. Similarly we specify the service time. Finally we state the number of servers.

EXAMPLE 4.1.1. *A local bank has two cash dispensers. A customer arrives and joins a central queue for both machines (or uses a machine if one is free). Assuming it always takes the same time to use a machine, and that customers arrive at random, specify the queueing system.*

SOLUTION. 'Arrive at random' means arrivals are a Poisson process, 'M'. Service time constant: 'D'. Two servers. The queue is M/D/2. □

4.2 The simple queue

Customers arrive singly, independently of one another at a service point. We assume that they arrive as a Poisson process, with arrival rate λ. This is equivalent to the distribution of the inter-arrival time being exponential with parameter λ.

This is a reasonable model for banks and supermarkets, but not for cinemas, where arrivals cluster near the time when films start.

If the server is free, the customer goes straight to the server and is served immediately. Otherwise they join the end of the queue. Customers are served in the order of their arrival. We define the following.

- The *service time* is the time the customer takes to be served once he/she reaches the server.
- The *waiting time* is the time for all customers ahead of the new arrival to be served (including the one at the server).
- The *queueing time* = waiting time + service time.

Suppose that the service time is distributed exponentially, with parameter μ (i.e. customers leave according to a Poisson process if the queue is non-empty). μ is the *service rate* of the queue.

For our simple model, the queue specification is M/M/1. This is called the *simple queue*.

4.3 Queue size

We shall consider the size of the queue at time t (this is of particular interest to a customer arriving at time t). Queue size (including the customer being served if there is one) is an integer-valued random process $\big(X(t)\big)_{t\geq 0}$.

This can be modelled in a similar way to the birth and death process. Let $p_x(t) = P(X(t) = x)$ and consider an arrival as a birth and a departure as a death. The general Kolmogorov equations are

$$p_x'(t) = \beta_{x-1}p_{x-1}(t) - (\beta_x + \nu_x)p_x(t) + \nu_{x+1}p_{x+1}(t) \qquad (x = 0, 1, 2, \ldots),$$

where $\beta_{-1} = p_{-1}(t) = 0$. For the simple queue, $\beta_x = \lambda$ for all $x \geq 0$ and $\nu_x = \mu$ for all $x > 0$. Thus

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) + \mu p_1(t). \\ p_x'(t) &= \lambda p_{x-1}(t) - (\lambda + \mu)p_x(t) + \mu p_{x+1}(t) \quad (x = 1, 2, \ldots), \end{aligned}$$

This turns out to be a difficult set of equations to solve. We shall consider a more tractable problem.

In general we are interested in how the queue behaves after a long time; e.g. will the server be overwhelmed and if not what does the steady state distribution of the queue size look like?

To find a steady state, we must set all the derivatives in the Kolmogorov equations to zero, giving

$$\begin{gathered} -\lambda p_0 + \mu p_1 = 0, \\ \lambda p_{x-1} - (\lambda + \mu)p_x + \mu p_{x+1} = 0 \quad (x = 1, 2, \ldots). \end{gathered}$$

The latter gives $\lambda p_{x-1} - \mu p_x = \lambda p_x - \mu p_{x+1}$ for $x = 1, 2, \ldots$, and the former says that this quantity equals 0. Thus

$$p_{x+1} = \frac{\lambda}{\mu} p_x \qquad (x = 0, 1, 2, \ldots),$$

i.e.

$$p_x = \left(\frac{\lambda}{\mu}\right)^x p_0 \qquad (x = 0, 1, 2, \ldots).$$

Then (if $\lambda < \mu$)

$$1 = \sum_{x=0}^{\infty} p_x = \sum_{x=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^x p_0 = \frac{1}{1 - \lambda/\mu} p_0,$$

and so $p_0 = 1 - \lambda/\mu$. Thus

$$p_x = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^x \qquad (x = 0, 1, 2, \ldots).$$

Setting $\rho = \lambda/\mu$, this says

$$p_x = (1-\rho)\rho^x \qquad (x = 0, 1, 2, \ldots).$$

This is a steady-state distribution if and only if $\lambda < \mu$; otherwise no steady state exists. $\rho = \lambda/\mu$ is the *traffic intensity* of the queue.

EXAMPLE 4.3.1. *Suppose that the arrival and departure of customers at a village post office may be modelled as a simple queue, and that customers arrive every 12 minutes on average.*

(a) What is the condition for the queue to attain stochastic equilibrium?
(b) For (b)–(d) assume the mean service time is 8 minutes. What proportion of the time is the counter empty?
(c) What is the average queue length?
(d) If you enter the post office (some time after it has opened), what is the probability that there are more than two people in the queue already?

SOLUTION.

(a) $\lambda = 1/12$. Thus we need $\mu > 1/12$, i.e. the expected service time must be less than 12 minutes.
(b) $\mu = 1/8$, so

$$P(X = 0) = 1 - \frac{\lambda}{\mu} = 1 - \frac{1/12}{1/8} = 1 - \frac{2}{3} = \frac{1}{3}.$$

(c) The mean queue length is

$$\begin{aligned}\sum_{x=0}^{\infty} x p_x &= \sum_{x=0}^{\infty} x(1-\rho)\rho^x \\ &= (1-\rho)\rho\sum_{x=1}^{\infty} x\rho^{x-1} \\ &= (1-\rho)\rho\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{2/3}{1-2/3} = 2.\end{aligned}$$

(d)

$$\begin{aligned}P(X > 2) &= 1 - p_0 - p_1 - p_2 \\ &= 1 - \frac{1}{3} - \frac{1}{3}\left(\frac{2}{3}\right) - \frac{1}{3}\left(\frac{2}{3}\right)^2 \\ &= 1 - \frac{1}{3} - \frac{2}{9} - \frac{4}{27} = \frac{8}{27} \simeq 0{\cdot}296.\end{aligned}$$

□

We'll now consider the distribution of the queueing time (waiting time + service time).

The waiting time depends upon the size of the queue. If there are $x > 0$ in the queue, the waiting time is the time to serve the person at the head of the queue (who may have been there some time) plus the time to serve the other $x - 1$. Since our service times are exponential, however, they have the lack of memory property and so the distribution of the remaining time to serve the first person is still Expon(μ). Thus the waiting time is the sum of x independent random variables, each Expon(μ)-distributed.

This conclusion remains true when $x = 0$. In all cases the queueing time is thus the sum of $x+1$ independent Expon(μ)-distributed random variables, and so has the Gamma distribution $\Gamma(x+1,\mu)$.

For the queue in equilibrium, let $f(t)$ denote the density of the queueing time. Let $f_x(t)$ denote the density conditional upon the queue size being x. Then

$$f_x(t) = \frac{\mu^{x+1}t^x e^{-\mu t}}{x!} \qquad (t > 0),$$

so

$$\begin{aligned} f(t) &= \sum_{x=0}^{\infty} f_x(t)P(X=x) \\ &= \sum_{x=0}^{\infty} \frac{\mu^{x+1}t^x e^{-\mu t}}{x!}\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^x \\ &= \mu e^{-\mu t}\left(1-\frac{\lambda}{\mu}\right)\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{x!} \\ &= \mu e^{-\mu t}\left(1-\frac{\lambda}{\mu}\right)e^{\lambda t} = (\mu-\lambda)e^{-(\mu-\lambda)t}, \end{aligned}$$

so the queueing time for a typical customer is exponential, with parameter $\mu - \lambda$.

Another feature of interest is the *busy period*, the time spent serving customers between successive periods of rest.

- If $\lambda > \mu$ then initially there may be times when the queue is empty, but eventually there will be a busy period which never ends.
- $\lambda = \mu$ gives a strange situation, where any busy period is certain to end, but the expected length of time to this ending is infinite.
- $\lambda < \mu$ means that alternating periods of rest/busy periods are certain.

The last case is the one of interest, where the queue is stable. Consider that case. Because the probability that the server is unoccupied ('idle') at any given time is $p_0 = 1-\rho$, in a long period of time T the server will be idle for approximate time length $(1-\rho)T$.

An arrival terminates a period of idleness, so that periods of idleness are Expon(λ). So the number of idle intervals in time T is approximately

$$\frac{(1-\rho)T}{1/\lambda} = \lambda(1-\rho)T.$$

Busy periods alternate with periods of rest, so that there are $\lambda(1-\rho)T$ such periods during the total busy time ρT, so the average length of a busy period is

$$\frac{\rho T}{\lambda(1-\rho)T} = \frac{\lambda/\mu}{\lambda(1-\lambda/\mu)} = \frac{1}{\mu-\lambda}.$$

Note that this is equal to the expected time that a customer would have to spend in a queue.

The expected number of customers served in a busy period is given, roughly, by the expected number of arrivals divided by the expected number of busy periods, which is

$$\frac{\lambda T}{\lambda(1-\rho)T} = \frac{1}{1-\rho} = \frac{\mu}{\mu-\lambda}.$$

4.4 The M/M/n queue

We shall expand the 'simple queue' model to include more than one server (in reality a single server is relatively rare, except in the smallest shops). We assume a central queue where, when a server becomes free, the head of the queue goes to that server. Thus for an n-server queue we need over n people 'in the queue' to have any real queueing at all. People begin to be served in the order of their arrival, although they may finish out of turn. We assume that each server has the same service rate μ.

We shall again find the differential-difference equations for the number $X(t)$ in the queue.

Note that the rate of service depends upon the queue length, namely $\nu_x = x\mu$ if $x < n$, but $\nu_x = n\mu$ if $x \geq n$, since all the servers are occupied. Thus

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p_x'(t) &= \lambda p_{x-1}(t) - (\lambda + x\mu)p_x(t) + (x+1)\mu p_{x+1}(t) \\ &\qquad (x = 1, 2, \ldots, n-1), \\ p_x'(t) &= \lambda p_{x-1}(t) - (\lambda + n\mu)p_x(t) + n\mu p_{x+1}(t) \\ &\qquad (x = n, n+1, \ldots). \end{aligned}$$

To find the steady state, we set all the above derivatives to zero, giving first

$$p_1 = \frac{\lambda}{\mu} p_0.$$

Next, $\lambda p_{x-1} - x\mu p_x = \lambda p_x - (x+1)\mu p_{x+1}$ for $x = 1, \ldots, n-1$, and this equals 0 by the above, so that

$$p_{x+1} = \frac{\lambda}{(x+1)\mu} p_x,$$

i.e.

$$p_x = \left(\frac{\lambda}{\mu}\right)^x \frac{1}{x!} p_0 \qquad (x = 0, 1, \ldots, n).$$

If $x \geq n$ then $\lambda p_{x-1} - n\mu p_x = \lambda p_x - n\mu p_{x+1}$, which again equals 0 by the above, and so

$$\begin{aligned} p_{x+1} &= \frac{\lambda}{n\mu} p_x, \\ \therefore \quad p_x &= \left(\frac{\lambda}{n\mu}\right)^{x-n} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n p_0 \\ &= \left(\frac{\lambda}{n\mu}\right)^x \frac{n^n}{n!} p_0 \quad (x = n, n+1, \ldots). \end{aligned}$$

This sequence converges if and only if $\lambda < n\mu$, which is thus the condition for a steady-state distribution to exist.

Finally we must find the value of p_0. We require the probabilities to add to 1, so

$$\begin{aligned} p_0 \left(1 + \sum_{x=1}^{n-1} \left(\frac{\lambda}{\mu}\right)^x \frac{1}{x!} + \sum_{x=n}^{\infty} \left(\frac{\lambda}{n\mu}\right)^x \frac{n^n}{n!}\right) &= 1, \\ \therefore \quad p_0 = \left(1 + \sum_{x=1}^{n-1} \frac{(n\rho)^x}{x!} + \frac{(n\rho)^n}{n!(1-\rho)}\right)^{-1} & \end{aligned}$$

where

$$\rho = \frac{\lambda}{n\mu}$$

is the *traffic intensity* (note: redefined) of the process.

Because of the complexity of the above, we shall leave the term p_0 in our expressions for the equilibrium probabilities, which are thus

$$p_x = \begin{cases} p_0 \dfrac{n^x}{x!} \rho^x & \text{for } x = 0, 1, \ldots, n-1, \\ p_0 \dfrac{n^n}{n!} \rho^x & \text{for } x = n, n+1, \ldots. \end{cases}$$

For example, if there are two servers, i.e. $n = 2$,

$$p_0 = \left(1 + 2\rho + \frac{(2\rho)^2}{2!(1-\rho)}\right)^{-1}$$
$$= \frac{1-\rho}{(1-\rho)(1+2\rho) + 2\rho^2} = \frac{1-\rho}{1+\rho},$$

and so

$$p_1 = \frac{2\rho(1-\rho)}{1+\rho}, \quad p_x = \frac{2\rho^x(1-\rho)}{1+\rho} \quad (x = 2, 3, \ldots).$$

The expectation of X is thus given by

$$EX = \sum_{x=1}^{\infty} x \frac{2\rho^x(1-\rho)}{1+\rho} = \frac{2\rho(1-\rho)}{1+\rho} \sum_{x=1}^{\infty} x\rho^{x-1}$$
$$= \frac{2\rho(1-\rho)}{1+\rho} \frac{1}{(1-\rho)^2} = \frac{2\rho}{1-\rho^2}.$$

EXAMPLE 4.4.1. *Customers enter a supermarket at a rate of three per minute, each customer taking an average of two minutes to serve.*

(a) *What is the minimum number of servers to make the queue stable?*
(b) *If there are eight checkouts, what proportion of customers receive immediate service?*

SOLUTION. $\lambda = 3$, $\mu = 1/2$.

(a) For equilibrium we need

$$\frac{\lambda}{n\mu} < 1 \implies \frac{3}{\frac{1}{2}n} < 1 \implies n > 6,$$

so that 7 servers are required.

(b) If there are 8 servers, the probability of being served immediately is the probability that there are 7 customers or less upon arrival. Now

$$\rho = \frac{\lambda}{n\mu} = \frac{3}{1/2 \times 8} = \frac{3}{4} \implies n\rho = 6.$$

so

$$P(\leq 7 \text{ customers}) = p_0 + p_1 + \cdots + p_7$$
$$= p_0 \left(1 + 6 + \frac{6^2}{2!} + \frac{6^3}{3!} + \frac{6^4}{4!} + \frac{6^5}{5!} + \frac{6^6}{6!} + \frac{6^7}{7!}\right)$$
$$= \frac{2101}{7} p_0.$$

Therefore

$$p_0 = \left(\frac{2101}{7} + \frac{6^8}{8!(1 - 3/4)}\right)^{-1} = \frac{35}{16337}.$$

So the probability of being served immediately is

$$\frac{10505}{16337} \simeq 0{\cdot}643.$$

□

We now consider the waiting time. Suppose that when a customer arrives, there is a queue of length x ahead of him or her.

If $x < n$ then the waiting time is zero and the customer goes straight to a server. Denoting the waiting time by T, we thus have

$$P(T = 0) = P(X < n) = 1 - P(X \geq n) = 1 - p_0 \frac{(n\rho)^n}{n!(1-\rho)}.$$

Otherwise n customers are being served and $x-n$ are ahead of the customer in the queue, so that $1+x-n$ more departures are required before the customer is served. There are n servers, so that departures occur as a Poisson process with rate $n\mu$.

If $f(t)$ is the density of T, excluding the case $T=0$ above, and $f(t|x)$ the density of T conditional on X taking value $x(\geq n)$, then

$$f(t) = \sum_{x=n}^{\infty} f(t|x)P(X=x). \qquad (4.4.1)$$

When $X = x \geq n$, the arriving customer must wait for $1+x-n$ departures, each of which is Expon($n\mu$) distributed, so that the waiting time is $\Gamma(1+x-n, n\mu)$. Thus

$$\begin{aligned} f(t|x) &= \frac{t^{x-n}e^{-n\mu t}(n\mu)^{1+x-n}}{(x-n)!} \\ &= \frac{t^{y}e^{-n\mu t}(n\mu)^{1+y}}{y!} \quad \text{where } y = x-n. \end{aligned}$$

Substituting into (4.4.1) yields

$$\begin{aligned} f(t) &= \sum_{y=0}^{\infty} \frac{t^{y}e^{-n\mu t}(n\mu)^{1+y}}{y!} p_0 \frac{n^n \rho^{n+y}}{n!} \\ &= n\mu p_0 \frac{(n\rho)^n}{n!} e^{-n\mu t} \sum_{y=0}^{\infty} \frac{(tn\mu\rho)^y}{y!} \\ &= n\mu p_0 \frac{(n\rho)^n}{n!} e^{-n\mu t} e^{n\mu t\rho} \\ &= p_0 \mu \frac{(n\rho)^n}{(n-1)!} e^{-n\mu(1-\rho)t}, \end{aligned}$$

which is of exponential form.

Note that this will integrate to a number less than 1, namely to $1-P(T=0)$.

Thus the waiting time is a combination of a discrete and a continuous random variable. The queueing time is just the sum of this waiting time and an independent exponential random variable with parameter μ.

4.5 The M/M/∞ queue

If the number of servers is so large it can be assumed to be infinite, then no customer ever has to wait. Note that this would not be viable for a shop to put into practice, in general.

This is exactly the same process as the immigration-death process from the previous chapter. In a steady state the queue size thus has a Poisson distribution with parameter λ/μ, i.e.

$$P(X=x) = e^{-\lambda/\mu}\left(\frac{\lambda}{\mu}\right)^x \frac{1}{x!} \qquad (x = 0, 1, 2, \ldots).$$

We now go on to consider queues where the service time is not exponentially distributed. We first consider the case when it is constant.

4.6 The M/D/1 queue

The assumptions of the M/D/1 queue are as follows.

(1) Customers arrive according to a Poisson process.
(2) There is a single server.
(3) It takes exactly the same time to serve each customer.

We denote the service time by $T = 1/\mu$ (equivalent to a service rate of μ, since $ET = 1/\mu$ for the M/M/1 queue).

An example for which this is a reasonable model is customers arriving at a cash machine. Every customer inputs a PIN number, asks for a given amount of cash and waits for the cash to be delivered. The time taken is roughly the same for all customers.

We again wish to find the probability $p_x(t)$ that there are exactly x people in the queue at time t. We know that every service takes exactly $1/\mu$ units of time, so that in the interval $(t, t+1/\mu]$ exactly one customer is served, unless the queue was empty at time t, in which case no customers will be served in the interval.

Suppose that the number of arrivals in $(t, t+1/\mu]$ is A, then A has a Poisson distribution with parameter $\rho = \lambda/\mu$. If the number in the queue at time t is x, the number at time $t+1/\mu$ is

$$\begin{cases} X\big(t + {}^1\!/\!_\mu\big) = 0 + A = A & \text{if } X(t) = x = 0, \\ X\big(t + {}^1\!/\!_\mu\big) = x + A - 1 & \text{if } X(t) = x > 0. \end{cases}$$

The approach to finding $p_x(t)$ will be different from before. We shall use the natural relationship between $p_x(t)$ and $p_x(t+1/\mu)$. First,

$$\begin{aligned} p_0(t+1/\mu) &= p_0(t)P(A=0) + p_1(t)P(A=0) \\ &= \big(p_0(t) + p_1(t)\big)e^{-\rho}. \end{aligned}$$

Similarly

$$\begin{aligned} &p_1(t+1/\mu) \\ &\quad = p_0(t)P(A=1) + p_1(t)P(A=1) + p_2(t)P(A=0) \\ &\quad = \big(p_0(t) + p_1(t)\big)\rho e^{-\rho} + p_2(t)e^{-\rho}. \end{aligned}$$

In general

$$\begin{aligned} &p_x(t+1/\mu) \\ &= p_0(t)P(A=x) + p_1(t)P(A=x) \\ &\qquad + p_2(t)P(A=x-1) + \cdots + p_{x+1}(t)P(A=0) \\ &= e^{-\rho}\left(\big(p_0(t)+p_1(t)\big)\frac{\rho^x}{x!} + p_2(t)\frac{\rho^{x-1}}{(x-1)!} + \cdots + p_{x+1}(t)\right), \end{aligned}$$

for $x = 0, 1, \ldots$.

To find the steady state we set $p_x(t) = p_x(t+1/\mu) = p_x$ for all x, so that

$$p_x = e^{-\rho}\left(p_0\frac{\rho^x}{x!} + \sum_{j=1}^{x+1} p_j \frac{\rho^{x+1-j}}{(x+1-j)!}\right) \qquad (x = 0, 1, 2, \ldots).$$

Once again considering probability generating functions,

$$\sum_{x=0}^{\infty} p_x s^x = e^{-\rho}\sum_{x=0}^{\infty}\left(p_0\frac{\rho^x}{x!} + \sum_{j=1}^{x+1} p_j\frac{\rho^{x+1-j}}{(x+1-j)!}\right)s^x.$$

In the inner sum on the right, set $i = j - 1$:

$$\begin{aligned} \Pi(s) &= e^{-\rho}p_0 e^{\rho s} + e^{-\rho}\sum_{x=0}^{\infty}\sum_{i=0}^{x} p_{i+1}\frac{\rho^{x-i}}{(x-i)!}s^x \\ &= e^{-\rho(1-s)}p_0 + e^{-\rho}\sum_{i=0}^{\infty}\sum_{x=i}^{\infty} p_{i+1}\frac{\rho^{x-i}}{(x-i)!}s^x \\ &= e^{-\rho(1-s)}p_0 + e^{-\rho}\sum_{i=0}^{\infty} p_{i+1}s^i e^{\rho s} \\ &= e^{-\rho(1-s)}p_0 + e^{-\rho(1-s)}\frac{\Pi(s) - p_0}{s} \\ &= e^{-\rho(1-s)}\left(p_0\Big(1 - \frac{1}{s}\Big) + \frac{\Pi(s)}{s}\right). \end{aligned}$$

Therefore

$$se^{\rho(1-s)}\Pi(s) = -p_0(1-s) + \Pi(s)$$
$$\therefore \quad \Pi(s) = \frac{p_0(1-s)}{1-se^{\rho(1-s)}}.$$

To specify the p.g.f. completely we must find the value of p_0. Rearranging the above expression gives

$$\Pi(s)(1-se^{\rho(1-s)}) = p_0(1-s).$$

Differentiating:

$$\Pi'(s)(1-se^{\rho(1-s)}) - \Pi(s)(1-\rho s)e^{\rho(1-s)} = -p_0. \qquad (4.6.1)$$

Setting $s=1$ gives

$$-(1-\rho) = -p_0.$$

So $p_0 = 1-\rho$, and we obtain finally

$$\Pi(s) = \frac{(1-\rho)(1-s)}{1-se^{\rho(1-s)}}.$$

We shall now find the mean of this distribution. Differentiating (4.6.1) again gives

$$\Pi''(s)(1-se^{\rho(1-s)}) - 2\Pi'(s)(1-\rho s)e^{\rho(1-s)} + \Pi(s)(2\rho-\rho^2 s)e^{\rho(1-s)} = 0.$$

Setting $s=1$, and using again $\Pi(1)=1$ and now also $\Pi'(1) = EX$,

$$-2(1-\rho)EX + 2\rho - \rho^2 = 0.$$

So

$$EX = \frac{2\rho-\rho^2}{2(1-\rho)} = \frac{\rho-\rho^2/2}{1-\rho}.$$

The waiting time (and hence the queueing time) can be considered as follows. A customer arriving at random is equally likely to arrive at any point during the $1/\mu$ it takes the current customer to be served, i.e. if the queue length is $x>0$, the total waiting time is

$$(x-1)\frac{1}{\mu} + \text{Unif}\left(0, \frac{1}{\mu}\right),$$

which has expectation

$$\left(x-\frac{1}{2}\right)\frac{1}{\mu}.$$

Thus the expected queueing time is

$$\begin{aligned}\frac{1}{\mu} + E(\text{waiting time}) &= \frac{1}{\mu} + 0\times p_0 + \sum_{x=1}^{\infty}\left(x-\frac{1}{2}\right)\frac{1}{\mu}p_x \\ &= \frac{1}{\mu} + \frac{1}{\mu}\sum_{x=1}^{\infty} x p_x - \frac{1}{2\mu}\sum_{x=1}^{\infty} p_x \\ &= \frac{1}{\mu} + \frac{EX}{\mu} - \frac{1-p_0}{2\mu} \\ &= \frac{1}{\mu}\left(1 + \frac{\rho-\frac{1}{2}\rho^2}{1-\rho} - \frac{\rho}{2}\right) \\ &= \frac{2-\rho}{2\mu(1-\rho)}.\end{aligned}$$

Note that the expected waiting time is

$$\frac{2-\rho}{2\mu(1-\rho)} - \frac{1}{\mu} = \frac{\rho}{2\mu(1-\rho)}.$$

4.7 The M/G/1 queue

This queue has a single server and arrivals occur as a Poisson process, as in the M/M/1 and M/D/1 queues. Here however the distribution of the service time is left unspecified. We again consider only the equilibrium distribution.

Let T have the service-time distribution. During a service time T, a random number A of customers arrive, and given T, A is Pois(λT) distributed. The Pois(θ) distribution has p.g.f. $e^{-\theta(1-s)}$, because

$$\sum_{j=0}^{\infty} e^{-\theta}\frac{\theta^j}{j!}s^j = e^{-\theta}\sum_{j=0}^{\infty}\frac{(\theta s)^j}{j!} = e^{-\theta}e^{\theta s} = e^{-\theta(1-s)},$$

so

$$E(s^A|T) = e^{-\lambda T(1-s)} \qquad (|s| \le 1).$$

It follows that A has p.g.f.

$$Es^A = E\big(E(s^A|T)\big) = Ee^{-\lambda T(1-s)} \qquad (|s| \le 1).$$

Let Q_n be the number of customers remaining in the system (waiting, or being served) just after the departure of the n^{th} customer.

If $Q_n > 0$ then the $(n+1)^{\text{th}}$ customer begins service immediately, and A customers arrive during this service time. So the $(n+1)^{\text{th}}$ customer leaves $A + Q_n - 1$ customers on departure. If $Q_n = 0$, then the server waits for the $(n+1)^{\text{th}}$ arrival, thus leaving $1 + A - 1 = A$ customers on departure. So

$$Q_{n+1} = \begin{cases} A + Q_n - 1 & \text{if } Q_n > 0, \\ A & \text{if } Q_n = 0. \end{cases}$$

The sequence of random variables $(Q_1, Q_2, \ldots)$ is thus a Markov chain. What are its transition probabilities? We've just seen that this chain can go up any distance in one transition, but down only by single steps. So

$$\begin{aligned} P(Q_{n+1} = j) &= \sum_{i=0}^{j+1} P(Q_n = i)P(Q_{n+1} = j|Q_n = i) \\ &= P(Q_n = 0)P(A = j) + \sum_{i=1}^{j+1} P(Q_n = i)P(A = j - i + 1). \end{aligned}$$

For a steady state we require that $P(Q_{n+1} = j) = P(Q_n = j) = p_j$, say. Letting

$$\delta_i = P(A = i) \qquad (i = 0, 1, 2, \ldots),$$

we obtain

$$p_j = p_0\delta_j + \sum_{i=1}^{j+1} p_i\delta_{j-i+1} \qquad (j = 0, 1, 2, \ldots). \tag{4.7.1}$$

We want to solve these for $p_j = P(Q = j)$, the stationary distribution of Q, the queue size just after a departure. Let $\Pi(s) = Es^Q = \sum_{j=0}^{\infty} p_j s^j$ be the p.g.f. of Q, and note that $\sum_{j=0}^{\infty} \delta_j s^j$ is the p.g.f. of A, which we found to equal $Ee^{-\lambda T(1-s)}$. Multiply (4.7.1) by s^j and sum over j, to get

$$\Pi(s) = p_0 Ee^{-\lambda T(1-s)} + \sum_{j=0}^{\infty}\sum_{i=1}^{j+1} p_i\delta_{j-i+1}s^j.$$

To deal with the double sum we go over to index variables i and k where $k = j - i + 1$. Their ranges of value are 1 to ∞ for i and 0 to ∞ for k, so the double sum becomes

$$\sum_{i=1}^{\infty}\sum_{k=0}^{\infty} p_i\delta_k s^{k+i-1} = \sum_{i=1}^{\infty} p_i s^{i-1}\sum_{k=0}^{\infty}\delta_k s^k$$

$$= \frac{\Pi(s) - p_0}{s} E e^{-\lambda T(1-s)}.$$

Thus

$$\Pi(s) = p_0 E e^{-\lambda T(1-s)} + \frac{\Pi(s) - p_0}{s} E e^{-\lambda T(1-s)};$$

$$\therefore \quad \Pi(s) = \frac{p_0(1-s) E e^{-\lambda T(1-s)}}{E e^{-\lambda T(1-s)} - s}.$$

To find p_0 we rewrite this as

$$(E e^{-\lambda T(1-s)} - s)\Pi(s) = p_0(1-s) E e^{-\lambda T(1-s)}$$

and differentiate:

$$\begin{aligned}(E e^{-\lambda T(1-s)} - s)\Pi'(s) &+ \big(\lambda E(T e^{-\lambda T(1-s)}) - 1\big)\Pi(s) \\ &= -p_0 E e^{-\lambda T(1-s)} + p_0(1-s)\lambda E(T e^{-\lambda T(1-s)}). \qquad (4.7.2)\end{aligned}$$

Setting $s = 1$ we obtain $\lambda ET - 1 = -p_0$, so $p_0 = 1 - \lambda ET$. We shall again let $\rho = \lambda ET$ be the traffic intensity (this is consistent with the definitions in the M/M/1 and M/D/1 queues). So finally

$$p_0 = 1 - \rho,$$

and (provided $\rho < 1$)

$$\Pi(s) = \frac{(1-\rho)(1-s) E e^{-\lambda T(1-s)}}{E e^{-\lambda T(1-s)} - s} \qquad (|s| \le 1).$$

To find the expected queue length, differentiate (4.7.2) once more:

$$\begin{aligned}(E e^{-\lambda T(1-s)} - s)\Pi''(s) &+ 2\big(\lambda E(T e^{-\lambda T(1-s)}) - 1\big)\Pi'(s) \\ &+ \lambda^2 E(T^2 e^{-\lambda T(1-s)})\Pi(s) \\ &= -2p_0\lambda E(T e^{-\lambda T(1-s)}) + p_0(1-s)\lambda^2 E(T^2 e^{-\lambda T(1-s)}).\end{aligned}$$

Setting $s = 1$ we obtain

$$2(\lambda ET - 1)EQ + \lambda^2 E(T^2) = -2p_0\lambda ET,$$

and so

$$\begin{aligned}2(1-\rho)EQ &= \lambda^2 E(T^2) + 2(1-\rho)\lambda ET \\ &= 2(1-\rho)\rho + \lambda^2\big(\operatorname{var} T + (ET)^2\big) \\ &= 2(1-\rho)\rho + \lambda^2 \operatorname{var} T + \rho^2.\end{aligned}$$

Thus

$$EQ = \frac{2\rho - \rho^2 + \lambda^2 \operatorname{var} T}{2(1-\rho)}.$$

This is known as the *Pollaczek-Khinchin formula* after Félix Pollaczek (1892–1981) and Aleksandr Yakovlevich Khinchin (1894–1959). It gives that the expected queue length depends not just upon the mean of the service time, but also on its variance! The larger the variance, the longer the queue.

EXAMPLE 4.7.1. *Use the above result to find the mean length in equilibrium, just after departures, of*

(a) *the M/M/1 queue;*
(b) *the M/U/1 queue, where T is* Unif$(0, 2/\mu)$*;*
(c) *the M/Γ/1 queue, where T is* $\Gamma(n, n\mu)$.

SOLUTION.

(a) For the M/M/1 queue, T is exponential with parameter μ, so that the variance of T is $1/\mu^2$, whence

$$EQ = \frac{2\rho - \rho^2 + \lambda^2/\mu^2}{2(1-\rho)} = \frac{2\rho - \rho^2 + \rho^2}{2(1-\rho)} = \frac{\rho}{1-\rho}$$

as before.

(b) For a uniform distribution on an interval (a,b) the variance is $(b-a)^2/12$, so here

$$\lambda^2 \operatorname{var} T = \lambda^2 \frac{1}{12}\left(\frac{2}{\mu}\right)^2 = \frac{\lambda^2}{3\mu^2} = \frac{\rho^2}{3}.$$

Therefore

$$EQ = \frac{2\rho - \rho^2 + \rho^2/3}{2(1-\rho)} = \frac{\rho - \frac{1}{3}\rho^2}{1-\rho}.$$

(c) T is $\Gamma(n, n\mu)$ so

$$\lambda^2 \operatorname{var} T = \lambda^2 \frac{n}{(n\mu)^2} = \frac{\rho^2}{n}.$$

So

$$EQ = \frac{2\rho - \rho^2 + \rho^2/n}{2(1-\rho)} = \frac{2\rho - \frac{n-1}{n}\rho^2}{2(1-\rho)}.$$

□

4.8 Equilibrium theory

M/M/n queues are particular cases of the general birth-and-death process, a Markov process with

$$P(X(t+\delta t) = x+1 | X(t) = x) = \beta_x \delta t + o(\delta t),$$
$$P(X(t+\delta t) = x-1 | X(t) = x) = \nu_x \delta t + o(\delta t),$$

in which

$$\beta_x = \lambda \qquad (x = 0,1,2,\ldots),$$
$$\nu_x = \begin{cases} x\mu & (x = 0,1,2,\ldots,n-1), \\ n\mu & (x = n, n+1, \ldots). \end{cases}$$

We will develop a general method for finding a steady-state distribution for the general birth-and-death process.

For the general birth-death process, the Kolmogorov equations are

$$p_x'(t) = \beta_{x-1}p_{x-1}(t) - (\beta_x + \nu_x)p_x(t) + \nu_{x+1}p_{x+1}(t)$$
$$(x = 0,1,2,\ldots),$$

defining $p_{-1}(t) = 0$. Thus at equilibrium we have

$$p_{x+1} = \frac{1}{\nu_{x+1}}\big((\beta_x + \nu_x)p_x - \beta_{x-1}p_{x-1}\big),$$

giving

$$p_1 = \frac{\beta_0}{\nu_1}p_0,$$
$$p_2 = \frac{1}{\nu_2}\big((\beta_1+\nu_1)p_1 - \beta_0 p_0\big)$$
$$= \frac{1}{\nu_2}\left((\beta_1+\nu_1)\frac{\beta_0}{\nu_1}p_0 - \beta_0 p_0\right) = \frac{\beta_0\beta_1}{\nu_1\nu_2}p_0.$$

In general, we could show using induction that

$$p_x = \frac{\beta_0\beta_1\cdots\beta_{x-1}}{\nu_1\nu_2\cdots\nu_x}p_0 = \rho_x p_0 \quad (x = 1,2,\ldots),$$

where

$$\rho_x = \frac{\beta_0\beta_1\cdots\beta_{x-1}}{\nu_1\nu_2\cdots\nu_x}.$$

Then

$$\sum_{x=0}^{\infty} \rho_x p_0 = 1 \implies p_0 = \frac{1}{\sum_{x=0}^{\infty}\rho_x}$$

if this sum is finite. Otherwise there is no equilibrium distribution.

EXAMPLE 4.8.1. *Check the above for the M/M/1 queue.*

SOLUTION.

$$p_x = \frac{\beta_0\beta_1\cdots\beta_{x-1}}{\nu_1\nu_2\cdots\nu_x}p_0 = \left(\frac{\lambda}{\mu}\right)^x p_0 = \rho^x p_0,$$

so

$$\sum_{x=0}^{\infty}\rho_x = \sum_{x=0}^{\infty}\rho^x = \frac{1}{1-\rho} \quad \text{if } \rho < 1,$$

i.e.

$$p_x = (1-\rho)\rho^x \qquad (x = 0, 1, 2, \ldots)$$

as before. □

4.9 Other queues

We shall consider two types of queue.

1) Queues with discouragement

Sometimes, if a queue is sufficiently long, arrivals simply go away (and possibly try again later) rather than join the queue. Suppose that if there are already x customers in the system (waiting or being served) a new arrival stays with probability

$$\frac{1}{x+1}.$$

Then, for the simple queue (thus modified),

$$\beta_x = \frac{\lambda}{x+1} \qquad (x = 0, 1, 2, \ldots),$$
$$\nu_x = \mu \qquad (x = 1, 2, \ldots).$$

Here

$$\rho_x = \frac{\beta_0\beta_1\cdots\beta_{x-1}}{\nu_1\nu_2\cdots\nu_x} = \frac{1}{x!}\left(\frac{\lambda}{\mu}\right)^x,$$

and

$$\sum_{x=0}^{\infty}\frac{1}{x!}\left(\frac{\lambda}{\mu}\right)^x = e^{\lambda/\mu},$$

which is finite no matter how large λ is compared to μ (the discouragement factor is sufficiently strong). Thus $p_0 = e^{-\lambda/\mu}$, and

$$p_x = \rho_x p_0 = \frac{1}{x!}\left(\frac{\lambda}{\mu}\right)^x e^{-\lambda/\mu},$$

which is a Poisson distribution with parameter λ/μ.

2) Finite waiting room

Sometimes there are only a finite number of spaces for arrivals to wait. Arriving customers who find no space just go away without joining the queue (a so-called *queue with balking*). Suppose that a queue can accommodate at most $c+1$ customers (1 being served and c waiting). Then, for the simple queue,

$$\beta_x = \begin{cases} \lambda & (x = 0, 1, 2, \ldots, c), \\ 0 & (x = c+1, c+2, \ldots), \end{cases}$$
$$\nu_x = \mu \qquad (x = 1, 2, \ldots).$$

To find the equilibrium distribution, first

$$\rho_x = \left(\frac{\lambda}{\mu}\right)^x = \rho^x \quad (x = 0, 1, \ldots, c+1),$$

and $\rho_x = 0$ otherwise, since $\beta_x = 0$ for $x > c+1$. To find the distribution, use

$$1 = \sum_{x=0}^{c+1}\rho^x p_0 = p_0\frac{1-\rho^{c+2}}{1-\rho} \quad \text{if } \rho \neq 1,$$

and so

$$p_x = \frac{(1-\rho)\rho^x}{1-\rho^{c+2}} \qquad (x = 0, 1, 2, \ldots, c+1).$$

See exercises for the case $\rho = 1$.

Chapter 5: Renewal processes

5.1 Introduction

Beginning at time zero, a sequence of *lifetimes* is laid end-to-end along the positive time-axis. When a lifetime ends the next begins immediately, so these time-points are called *renewals*. We shall assume the lifetimes are independent and identically distributed positive random variables. They are the waiting times between renewals. The Poisson process is an example, with lifetimes having an exponential distribution. There are many examples of *renewal processes* that can be modelled in this way:

- lifetimes of a component vital to the running of a machine (e.g. batteries, light bulbs), which is immediately replaced when it fails by an identical one, and the machine continues;
- any process where the Poisson process is appropriate;
- departures from the end of an assembly line (almost constant waiting time);
- the commencement of successive idle periods in a queue.

The theory of such processes is called *renewal theory*.

We shall discuss various features of this model, for example

- 'the time since the last renewal' = 'the age of the component',
- 'waiting time to the next renewal' = 'remaining lifetime of the component'.

5.2 Discrete-time renewal processes

Renewals can occur at discrete time points 1, 2, ... only. At any integer time-point there is either one renewal or none. Lifetimes are thus positive integers. We assume that the lifetimes are independent and identically distributed (i.i.d.).

Example 5.2.1. The Bernoulli process. *Suppose that the probability of a renewal occurring at any time point is p, independently of all other time points. Find the lifetime distribution.*

Solution.

$$P(T = n) = (1 - p)^{n-1}p \qquad (n = 1, 2, \ldots).$$

□

We shall assume that we start observing renewals at times greater than 0, but that we know that there was a renewal at time $n = 0$. We define two useful sequences:

- $f_n = P$(the first renewal since time 0 is at time n),
- $u_n = P$(a renewal occurs at time n).

The values of f_n are given by the distribution of the lifetime T, i.e. $P(T = n) = f_n$ and are thus known (f_0 is defined to be 0). Since a renewal is assumed to have occurred at 0, $u_0 = 1$. In general we wish to find a relationship between the f_n and the u_n, and thus find u_n (note that for the Bernoulli process $u_n = p$ if $n > 0$, but usually u_n is unknown). (u_n) is called the *renewal sequence*.

Define the generating functions

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

We assume that lifetimes are finite, i.e. $P(T = \infty) = 0$ or equivalently $F(1) = 1$. Note that $U(s)$ is not the generating function of a random variable, since the u_i sum to greater than one (the sum will be infinite). Now

$$\begin{aligned}u_n &= P(\text{renewal at time } n)\\ &= \sum_{j=1}^{n} P(\text{renewal at } n | \text{first renewal at } j) \times\\ &\qquad \times P(\text{first renewal at } j).\end{aligned}$$

If there is a renewal at time j, the conditional probability that there's also a renewal at time n is u_{n-j} (as a new lifetime starts at j), i.e.

$$\begin{aligned}u_n &= \sum_{j=1}^{n} u_{n-j} f_j\\ &= \sum_{j=0}^{n} u_{n-j} f_j \quad \text{as } f_0 = 0.\end{aligned}$$

For $n = 1, 2, 3, \ldots$, multiply this by s^n, then add the equations up:

$$\begin{aligned}\sum_{n=1}^{\infty} u_n s^n &= \sum_{n=1}^{\infty}\sum_{j=0}^{n} u_{n-j} f_j s^n\\ \therefore \quad \sum_{n=0}^{\infty} u_n s^n - u_0 &= \sum_{n=0}^{\infty}\sum_{j=0}^{n} u_{n-j} f_j s^n - u_0 f_0\\ \therefore \quad \sum_{n=0}^{\infty} u_n s^n - 1 &= \sum_{j=0}^{\infty} f_j s^j \sum_{n=j}^{\infty} u_{n-j} s^{n-j}\\ \therefore \quad U(s) - 1 &= F(s)U(s)\\ \therefore \quad U(s) &= \frac{1}{1-F(s)}.\end{aligned}$$

Example 5.2.2. *Check this for the Bernoulli process.*

Solution. As $P(T = n) = (1-p)^{n-1}p$, we get the p.g.f.

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

Thus

$$\begin{aligned}U(s) &= \frac{1}{1 - ps/\big(1-(1-p)s\big)}\\ &= \frac{1-(1-p)s}{1-s}\\ &= 1 + \frac{ps}{1-s} = 1 + p(s + s^2 + s^3 + \cdots),\end{aligned}$$

i.e. $u_0 = 1$ and $u_n = p$ for $n > 0$. That is,

$$P(\text{renewal at time } n) = p \text{ for all } n \geq 1$$

(which, of course, follows from the definition of the Bernoulli process). □

Example 5.2.3. *Let $P(T = 1) = P(T = 2) = 1/2$. Find the renewal sequence (u_n) explicitly, and its behaviour as $n \to \infty$.*

Solution. $F(s) = \frac{1}{2}(s + s^2)$ and

$$\begin{aligned}U(s) &= \frac{1}{1 - \frac{1}{2}s - \frac{1}{2}s^2}\\ &= \frac{1}{(1-s)(1+\frac{1}{2}s)}\\ &= \frac{2}{3(1-s)} + \frac{1}{3(1+\frac{1}{2}s)}\\ &= \frac{2}{3}\sum_{n=0}^{\infty} s^n + \frac{1}{3}\sum_{n=0}^{\infty}\left(-\frac{s}{2}\right)^n,\end{aligned}$$

which expands to give the coefficient of s^n to be

$$u_n = \frac{2}{3} + \frac{1}{3}\left(-\frac{1}{2}\right)^n \qquad (n = 0, 1, 2, \ldots).$$

This converges to $2/3$ as $n \to \infty$. □

In the above example we know that $ET = F'(1) = 3/2$. The convergence of u_n to $1/ET$ is a consequence of the following result, which we shall not prove:

THEOREM 5.2.4. ERDŐS-FELLER-POLLARD THEOREM (1949). *Suppose that there is no integer k larger than 1 such that T is restricted to the set $\{k, 2k, 3k, 4k, \ldots\}$ of multiples of k. Then*

$$u_n \to \frac{1}{ET} \qquad (n \to \infty).$$

Finding the r^{th} renewal

The waiting time W_r for the r^{th} renewal is the sum of the waiting time for the first renewal, the waiting time from the first to the second renewal, etc., i.e.

$$W_r = T_1 + T_2 + \cdots + T_r,$$

which is the sum of r i.i.d. random variables. Thus its generating function $\Pi(s)$ is given by

$$\Pi(s) = \big(F(s)\big)^r.$$

We can use this formula to find the distribution of W_r. Further, there is a direct relationship between X_n (the number of renewals after time 0 up to time n) and W_r, namely

$$P(X_n \geq r) = P(W_r \leq n).$$

EXAMPLE 5.2.5. *In the last example, find the probability that there are exactly 7 renewals in the first 12 time points.*

SOLUTION. We have

$$F(s) = \frac{1}{2}s + \frac{1}{2}s^2$$
$$\big(F(s)\big)^r = \left(\frac{1}{2}s + \frac{1}{2}s^2\right)^r = s^r\left(\frac{1}{2} + \frac{1}{2}s\right)^r.$$

This is the product of two p.g.f.s: s^r is the p.g.f. of the distribution degenerate at r and the second term is the p.g.f. of a Binomial random variable with parameters r and $1/2$, i.e.

$$W_r = r + V_r \quad \text{where} \quad V_r \sim \text{Binom}(r, 1/2).$$

So

$$P(X_n \geq r) = P(W_r \leq n) = P(r + V_r \leq n) = P(V_r \leq n - r).$$

The probability that there are exactly 7 renewals up to time 12 is

$$\begin{aligned} P(X_{12} = 7) &= P(X_{12} \geq 7) - P(X_{12} \geq 8) \\ &= P(V_7 \leq 5) - P(V_8 \leq 4) \\ &= 1 - P(V_7 \geq 6) - \big(1 - P(V_8 \geq 5)\big) \\ &= P(V_8 \geq 5) - P(V_7 \geq 6). \end{aligned}$$

As $V_7 \sim \text{Binom}(7, 1/2)$ and $V_8 \sim \text{Binom}(8, 1/2)$,

$$\begin{aligned} &P(X_{12} = 7) \\ &\quad = \sum_{j=5}^{8} P(V_8 = j) - \sum_{j=6}^{7} P(V_7 = j) \end{aligned}$$

$$= \sum_{j=5}^{8} \binom{8}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{8-j} - \sum_{j=6}^{7} \binom{7}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{7-j}$$
$$= \frac{56+28+8+1}{256} - \frac{7+1}{128}$$
$$= \frac{93}{256} - \frac{8}{128} = \frac{77}{256} \simeq 0{\cdot}301.$$

□

5.3 The ordinary renewal process

We now move on to renewal processes in continuous time, $\big(X(t)\big)_{t\geq 0}$. Here, $X(t)$ is the number of renewals in the time interval $(0,t]$, i.e. not counting the renewal at time 0. Renewals occur singly, separated by i.i.d. waiting times. Our assumption that at time $t = 0$ a renewal occurred (i.e. we have a brand new component), means that T_1 has the same distribution as T_2, T_3, ... This is the *ordinary renewal process.*

Note that a renewal process is not in general a Markov process (unlike most of the processes we have previously encountered). The future of the process depends upon its history (how much lifetime of the current component has been used up), not just the most recent state (how many renewals there have been up to now). The only renewal processes that are also Markov processes are the Bernoulli process and the Poisson process, where the waiting times are memoryless.

In the following definitions T is the lifetime, i.e. the waiting time between renewals.

DEFINITION 5.3.1. *The survivor function for T is $Q(t) = P(T > t)$.*

Thus if the distribution function of T is $P(T \leq t) = G(t)$ then the survivor function is given by $Q(t) = P(T > t) = 1 - G(t)$.

DEFINITION 5.3.2. *The residual lifetime distribution at age z is the distribution of $T - z$ given $T > z$.*

Let T_z be a random variable having the residual lifetime at age z. Then, for any $t \geq 0$,

$$\begin{aligned} P(T_z > t) &= P(T > z + t | T > z) \\ &= \frac{P(T > z + t \text{ and } T > z)}{P(T > z)} \\ &= \frac{P(T > z + t)}{P(T > z)} = \frac{Q(z+t)}{Q(z)}. \end{aligned}$$

In the case of exponential lifetimes, of parameter μ say, $Q(t) = e^{-\mu t}$ and so

$$P(T_z > t) = \frac{Q(z+t)}{Q(z)} = \frac{e^{-\mu(z+t)}}{e^{-\mu z}} = e^{-\mu t} = P(T > t),$$

i.e. the residual lifetime at any age z has the same distribution as the original lifetime.

More realistically, components are more likely to break down as they age, i.e. $P(T_z > t) \leq P(T > t)$ for all values of z and t, with strict inequality for some z and t. This property is referred to as *new better than used.*

EXAMPLE 5.3.3. *Show that the $\Gamma(2, \beta)$ lifetime distribution is new-better-than-used.*

SOLUTION. Let $T \sim \Gamma(2, \beta)$, i.e. $g(t) = G'(t) = \beta^2 t e^{-\beta t}$ for $t > 0$. Thus

$$Q(t) = \int_t^\infty g(x)\,dx = (1 + \beta t)e^{-\beta t},$$

and so

$$
\begin{aligned}
P(T_z > t) &= \frac{Q(z+t)}{Q(z)} \\
&= \frac{1+\beta z+\beta t}{1+\beta z} e^{-\beta t} \\
&= \frac{1+\beta z+\beta t}{(1+\beta z)(1+\beta t)} P(T>t) < P(T>t),
\end{aligned}
$$

so that this distribution has the 'new better than used' property.
□

The less usual property of the component being less likely to break down the older it gets, i.e. *new worse than used*, is when $P(T_z > t) \geq P(T > t)$ for all z and t, with strict inequality for some z and t.

EXAMPLE 5.3.4. *Show that the Pareto lifetime distribution*

$$
Q(t) = \left(\frac{1}{1+\beta t}\right)^c \quad (t>0),
$$

where $\beta > 0$ and $c > 0$, has the 'new worse than used' property.

SOLUTION.

$$
\frac{P(T_z > t)}{P(T>t)} = \left(\frac{(1+\beta z)(1+\beta t)}{1+\beta(z+t)}\right)^c > 1.
$$

□

EXAMPLE 5.3.5. *Consider the Pareto distribution as above.*

(a) Find the life expectancy of a component aged z, as a multiple of ET.

(b) If $c = 3$, $\beta = 0{\cdot}1$ per hour, find

(i) ET;

(ii) the life expectancy of a component 5 hours old;

(iii) beyond what age a component which has already operated successfully for 15 hours is just as likely as not to survive.

SOLUTION.

(a)

$$
\begin{aligned}
ET &= \int_0^\infty P(T>t)\,dt \\
&= \int_0^\infty (1+\beta t)^{-c}\,dt \\
&= \left[\frac{(1+\beta t)^{-c+1}}{(-c+1)\beta}\right]_0^\infty \\
&= \frac{1}{\beta(c-1)}
\end{aligned}
$$

if $c > 1$. If $0 < c \leq 1$ then $ET = \infty$. Assuming $c > 1$,

$$
\begin{aligned}
ET_z &= \int_0^\infty P(T_z > t)\,dt \\
&= (1+\beta z)^c \int_0^\infty (1+\beta z+\beta t)^{-c}\,dt \\
&= (1+\beta z)^c \left[\frac{(1+\beta z+\beta t)^{-c+1}}{\beta(-c+1)}\right]_0^\infty \\
&= \frac{(1+\beta z)^c}{\beta(c-1)(1+\beta z)^{c-1}} = \frac{1+\beta z}{\beta(c-1)},
\end{aligned}
$$

so $ET_z = (1+\beta z)ET$.

(b) With $\beta = 0{\cdot}1$ and $c = 3$,

(i) $ET = 1/\big(0{\cdot}1(3-1)\big) = 5$ hours.

(ii) $ET_5 = (1 + 0{\cdot}1 \times 5)ET = 1{\cdot}5 \times 5 = 7{\cdot}5$ hours.

(iii) We need to find the median residual lifetime of a component that has worked for 15 hours, i.e. to find t such that $P(T_{15} > t) = 1/2$. So we need to solve

$$\frac{(1+\beta z)^c}{(1+\beta z+\beta t)^c} = \frac{1}{2}$$

$$\therefore \quad \frac{1+\beta z}{1+\beta z+\beta t} = \left(\frac{1}{2}\right)^{1/c},$$

and so

$$1+\beta z+\beta t = 2^{1/c}(1+\beta z),$$

$$\therefore \quad \beta t = (1+\beta z)(2^{1/c}-1),$$

and

$$\begin{aligned} t &= (\beta^{-1}+z)(2^{1/c}-1) \\ &= (10+15)(2^{1/3}-1) \simeq 6{\cdot}498 \text{ hours.} \end{aligned}$$

So the age which the component is as likely as not to reach is $15 + 6{\cdot}5 = 21{\cdot}5$ hours.

□

Hazard rate

We have considered an example where the chance of failure increases with age, and one where it decreases. More generally, the inequality relating T_z and T may change direction as the component ages; e.g. the risk of failure is high when the component is very new or very old (human body). We consider the rate at which failure occurs. The probability that a component fails in $(t, t+\delta t]$ given that it has survived until time t is given by

$$\begin{aligned} P(t < T \le t+\delta t | T > t) &= \frac{P(t < T \le t+\delta t)}{P(T > t)} \\ &= \frac{G(t+\delta t) - G(t)}{1 - G(t)} \\ &= \frac{g(t)\delta t}{1-G(t)} + \mathrm{o}(\delta t) \quad \text{as } \delta t \downarrow 0. \end{aligned}$$

Definition 5.3.6. *The* **hazard rate** *or* **age-specific failure rate** *is*

$$h(t) = \lim_{\delta t \downarrow 0} \frac{P(t < T \le t+\delta t | T > t)}{\delta t}.$$

The above calculation shows that the hazard rate is given by

$$h(t) = \frac{g(t)}{1-G(t)} \qquad (t > 0). \tag{5.3.1}$$

Note that this is clearly not a probability density.

The definition of hazard rate leads to a useful approximation:

$$P(t < T \le t+\delta t | T > t) \simeq h(t)\delta t \quad \text{for small } \delta t.$$

Example 5.3.7. *Again with the $\Gamma(2, \beta)$ distribution, let $\beta = 0{\cdot}5$ per hour.*

(a) Using the hazard rate, find the probability, approximately, that after 6 hours use the component fails in the next minute.

(b) Find the exact probability using the distribution function.

Solution. For the $\Gamma(2, \beta)$ lifetime, $G(t) = 1 - (1+\beta t)e^{-\beta t}$ and so

$$h(t) = \frac{\beta^2 t}{1+\beta t},$$

so here

$$h(t) = \frac{t}{4+2t}.$$

(a)

$$h(6) = \frac{6}{4+12} = \frac{3}{8}.$$

The probability that failure occurs in the next minute is approximately

$$\frac{3}{8} \times \frac{1}{60} = 0{\cdot}00625.$$

(b)

$$\begin{aligned} P(6 \le T < 6 + 1/60) &= \frac{Q(6) - Q(6 + 1/60)}{Q(6)} \\ &\simeq \frac{4e^{-3} - 4{\cdot}0083e^{-3{\cdot}0083}}{4e^{-3}} \\ &= 0{\cdot}00623. \end{aligned}$$

□

Formula (5.3.1) gives $h(t)$ in terms of the distribution, and we can invert it to give $G(t)$ or $Q(t)$ in terms of $h(t)$:

THEOREM 5.3.8. *For $t \ge 0$,*

$$Q(t) = e^{-\int_0^t h(u)\,du}.$$

PROOF. As $Q(t) = 1 - G(t)$, (5.3.1) gives

$$h(t) = \frac{-Q'(t)}{Q(t)}$$

(note that this is an alternative formula for the hazard rate). Change the variable to u and integrate from 0 to t:

$$\int_0^t h(u)\,du = [-\ln Q(u)]_0^t.$$

As $Q(0) = 1$, the right-hand side is $-\ln Q(t)$. The result follows.
□

COROLLARY 5.3.9. *If the hazard rate $h(t)$ is strictly increasing the lifetime distribution is new-better-than-used. If $h(t)$ is strictly decreasing the lifetime distribution is new-worse-than-used.*

PROOF.

$$P(T_z > t) = \frac{Q(z+t)}{Q(z)} = \frac{e^{-\int_0^{z+t} h(u)\,du}}{e^{-\int_0^{z} h(u)\,du}} = e^{-\int_z^{z+t} h(u)\,du}.$$

If h is strictly increasing then $\int_z^{z+t} h(u)\,du > \int_0^t h(u)\,du$ so

$$P(T_z > t) = e^{-\int_z^{z+t} h(u)\,du} < e^{-\int_0^{t} h(u)\,du} = P(T > t),$$

which is the new-better-than-used property. Similarly if h is strictly decreasing. □

EXAMPLE 5.3.10. *For the Pareto distribution as in Example 5.3.4, use the hazard rate to show 'new worse than used'.*

SOLUTION.

$$h(t) = \frac{-Q'(t)}{Q(t)} = \frac{c\beta}{(1+\beta t)^{c+1}} \Big/ \frac{1}{(1+\beta t)^{c}} = \frac{c\beta}{1+\beta t}.$$

This is a strictly decreasing function of t, so by the Corollary, the lifetime distribution is 'new worse than used'. □

The number of renewals of the ordinary renewal process
In a similar way to the discrete renewal process, we consider W_r, the waiting time until r renewals:

$$W_r = T_1 + T_2 + \cdots + T_r,$$

where $T_1, \ldots, T_r$ are i.i.d. with distribution function $G(t)$. Thus $EW_r = rET$.

Again similarly to the discrete case, we have

$$P(X(t) \geq r) = P(W_r \leq t) = G_r(t),$$

where $G_r(t)$ is the distribution function of W_r. Thus

$$EX(t) = \sum_{r=0}^{\infty} rP(X(t) = r) = \sum_{r=1}^{\infty} P(X(t) \geq r) = \sum_{r=1}^{\infty} G_r(t).$$

This is called the *renewal function* of the renewal process X.

(The middle equality in the above holds because of the following formula for the expectation of a general non-negative-integer-valued random variable X:

$$EX = \sum_{x=0}^{\infty} xP(X = x) = \sum_{x=1}^{\infty} P(X \geq x).$$

This can be deduced from the formula $EX = \int_0^{\infty} P(X > x)\,dx$ for the expectation of a general non-negative random variable, or can be proved directly.)

We let $g_r(t)$ denote the density of W_r. Then, differentiating the renewal function,

$$\frac{d}{dt}EX(t) = \frac{d}{dt}\sum_{r=1}^{\infty} G_r(t) = \sum_{r=1}^{\infty} g_r(t).$$

This, the derivative of the renewal function, is called the *renewal density* of X.

The probability of a renewal in $(t, t+\delta t]$ is

$$\begin{aligned}\sum_{r=1}^{\infty} P(t < W_r \leq t + \delta t) &= \sum_{r=1}^{\infty} g_r(t)\delta t + \mathrm{o}(\delta t) \\ &= \delta t \frac{d}{dt}EX(t) + \mathrm{o}(\delta t).\end{aligned}$$

The renewal density at time t is thus the rate at which renewals occur at t.

Example 5.3.11. *Find the renewal density for the Poisson process.*

Solution. As $W_r \sim \Gamma(r, \lambda)$,

$$\begin{aligned}\frac{d}{dt}EX(t) &= \sum_{r=1}^{\infty} g_r(t) \\ &= \sum_{r=1}^{\infty} \frac{t^{r-1}\lambda^r e^{-\lambda t}}{(r-1)!} \\ &= \lambda e^{-\lambda t}\sum_{r=1}^{\infty} \frac{(\lambda t)^{r-1}}{(r-1)!} = \lambda e^{-\lambda t}e^{\lambda t} = \lambda,\end{aligned}$$

as in the definition of the Poisson process. □

The *Renewal Theorem* (several forms) gives the asymptotic behaviour of the renewal function and renewal density:

$$\lim_{t\to\infty} \frac{EX(t)}{t} = \frac{1}{ET}, \quad \lim_{t\to\infty} \frac{d}{dt}EX(t) = \frac{1}{ET}.$$

The second of these limits is the rate of occurrence of renewals after the process has been running a long time. It is analogous to the result for the discrete renewal process.

5.4 The equilibrium renewal process

We consider a renewal process that has been running for some time. What is the total lifetime of the component currently in use? The lifetime of a component selected at random from a batch of identical components has survival function $Q(t) = 1 - G(t)$. It may seem that this should also describe the total lifetime of the component in use on inspection. However that's not the case.

EXAMPLE 5.4.1. *Suppose that the fuse for an electric circuit has the following lifetime distribution, in years:*

$$P(T=1) = \frac{1}{4}, \quad P(T=4) = \frac{3}{4}.$$

Find the probability that when the circuit is inspected, a long-life fuse will be found to be in use.

SOLUTION. A fuse selected from stock will last for 1 year with probability $1/4$ or for 4 years with probability $3/4$. However, if we inspect the circuit whilst one of the fuses is in use, the probability that a long-life fuse is observed is equal to the proportion of time that a long-life fuse is in use, which is

$$\frac{\frac{3}{4}\times 4}{\frac{3}{4}\times 4 + \frac{1}{4}\times 1} = \frac{12}{13}.$$

□

In general for components with discrete probability mass function $P(T=t) = g_t$, if W is the lifetime of the component in use,

$$P(W=w) = \frac{w g_w}{\sum t g_t} = \frac{w g_w}{ET}.$$

Similarly for components with a continuous distribution having density $g(t)$, the density $f_W(w)$ of the lifetime W of the component in use is

$$f_W(w) = \frac{w g(w)}{\int t g(t)\, dt} = \frac{w g(w)}{ET}.$$

EXAMPLE 5.4.2. *For a Poisson process, what is the density of W, the lifetime of the component in use on the observer's arrival?*

SOLUTION.

$$f_W(w) = \frac{w g(w)}{ET} = \frac{w\lambda e^{-\lambda w}}{1/\lambda} = \lambda^2 w e^{-\lambda w} \qquad (w > 0),$$

i.e. W is $\Gamma(2, \lambda)$. □

Suppose that nothing is known about the distribution of T. As T and W have different distributions, if sampling was based on examining the total lifetime of the element in use on a number of occasions we would not get a good estimate of the mean lifetime ET of a component. Our sampling of T is *length-biased.*

In general

$$\begin{aligned} EW &= \int_0^\infty w f_W(w)\, dw \\ &= \frac{\int_0^\infty w^2 g(w)\, dw}{ET} \\ &= \frac{E(T^2)}{ET} \\ &= \frac{(ET)^2 + \operatorname{var} T}{ET} \\ &= ET + \frac{\operatorname{var} T}{ET}. \end{aligned} \tag{5.4.1}$$

Thus W always has a higher mean than T.

We now consider the *residual lifetime* V of the component in use when we look at the process.

Since an observer is equally likely to arrive at any time during a component's lifetime, the time of an observer's arrival is uniform conditional upon W, i.e.

$$f(v|w) = \begin{cases} \dfrac{1}{w} & \text{if } 0 < v < w, \\ 0 & \text{otherwise.} \end{cases}$$

Letting $f(v, w)$ denote the joint density of v and w, the density $f_V(v)$ of V is given by

$$\begin{aligned} f_V(v) &= \int_0^\infty f(v, w)\, dw \\ &= \int_0^\infty f(v|w) f_W(w)\, dw \\ &= \int_v^\infty \frac{1}{w} \frac{wg(w)}{ET}\, dw \\ &= \frac{1 - G(v)}{ET}. \end{aligned}$$

In the previous sections, we have assumed that $t = 0$ is the time of a renewal, so that the waiting times $T_1, T_2, \ldots$ all have the same density function $g(t)$. Now we assume that we know nothing of the history of the process, but start at the observer's arrival (which will not generally be the time of a renewal). The density of the residual lifetime of the component in use is

$$g^*(t) = \begin{cases} \dfrac{1 - G(t)}{ET} = \dfrac{Q(t)}{ET} & \text{for } t > 0, \\ 0 & \text{for } t \le 0, \end{cases}$$

so that T_1 has density $g^*(t)$ while $T_2, T_3, \ldots$ all have density $g(t)$. All remain mutually independent. This is called the *equilibrium renewal process.*

EXAMPLE 5.4.3. *What is the equilibrium renewal process obtained from the Poisson process?*

SOLUTION. $g(t) = \lambda e^{-\lambda t}$, so

$$g^*(t) = \frac{1 - G(t)}{ET} = \frac{e^{-\lambda t}}{1/\lambda} = \lambda e^{-\lambda t} \quad (t > 0),$$

so that the Poisson renewal process and the Poisson equilibrium renewal process are identical (the only continuous renewal process with this property). □

EXAMPLE 5.4.4. *What is the distribution of T_1, the waiting time to the first failure, for the equilibrium renewal process where*

(a) $T \sim \Gamma(2, \lambda)$,
(b) $T \sim \text{Unif}(0, \theta)$*?*

SOLUTION.

(a) $G(t) = 1 - (1 + \lambda t)e^{-\lambda t}$ for $t > 0$, and $ET = 2/\lambda$, so

$$g^*(t) = \frac{1 - G(t)}{ET} = \frac{1}{2}\lambda(1 + \lambda t)e^{-\lambda t} \quad (t > 0).$$

(b) $G(t) = t/\theta$ for $0 < t < \theta$, and $ET = \theta/2$, so

$$g^*(t) = \frac{1 - G(t)}{ET} = \frac{2(\theta - t)}{\theta^2} \qquad (0 < t < \theta),$$

and $g^*(t) = 0$ for other t.

□

To find the expectation of T_1 recall that we developed its distribution by arguing that given W it will be $\mathrm{Unif}(0, W)$ distributed. This uniform distribution has expectation $W/2$, i.e.

$$E(T_1|W) = W/2.$$

Then

$$ET_1 = E\big(E(T_1|W)\big) = E(W/2) = \tfrac{1}{2}EW.$$

Example 5.4.5. *Find ET_1 for the two distributions in the previous example.*

Solution. Recall (5.4.1):

$$EW = ET + \frac{\operatorname{var} T}{ET}.$$

(a) For $T \sim \Gamma(2, \lambda)$, $ET = 2/\lambda$ and $\operatorname{var} T = 2/\lambda^2$, so

$$ET_1 = \frac{1}{2}\left(\frac{2}{\lambda} + \frac{2/\lambda^2}{2/\lambda}\right) = \frac{3}{2\lambda}.$$

(b) For $T \sim \mathrm{Unif}(0, \theta)$, $ET = \theta/2$ and $\operatorname{var} T = \theta^2/12$, so

$$ET_1 = \frac{1}{2}\left(\frac{\theta}{2} + \frac{\theta^2/12}{\theta/2}\right) = \frac{\theta}{3}.$$

□

It is not possible to say much in general about how the distributions of T_1 and T compare. In the above two examples, T_1 has a lower expectation than T. However for the Pareto distribution, for $c > 2$,

$$ET_1 = \frac{1}{\beta(c-2)} > \frac{1}{\beta(c-1)} = ET,$$

where the formula for ET_1 comes from the following

Exercise 5.4.6. *Show that the Pareto distribution of Example 5.3.4 has, for $c > 2$,*

$$E(T^2) = \frac{2}{\beta^2(c-1)(c-2)},$$

and hence that

$$\operatorname{var} T = \frac{c}{\beta^2(c-1)^2(c-2)}, \quad ET_1 = \frac{1}{\beta(c-2)}.$$

Chapter 6: Epidemics

6.1 Introduction

Throughout history epidemics have killed thousands of people and played decisive roles in wars and in civilisation in general (e.g. Bubonic plague). Today diseases such as malaria are widespread, and every few years there is a worldwide influenza epidemic. We shall consider some basic models of how epidemics spread.

First some definitions. It is assumed that initially there is at least one person with the disease. After the disease has been communicated by an *infected* individual (an *infective*) to a *susceptible* one, there is a *latent period* during which the newly infected individual cannot pass on the disease. This is followed by an *infectious period* during which the person can transmit the disease. (When the individual is first infected there is also an *incubation period*, the time until the first symptoms of the disease appear, which is generally a superset of the latent period.) At the end of the infectious period, the infected person is no longer capable of passing on the disease and *removal* occurs (the person is either dead, immune to the disease or otherwise has been put into isolation from the population). The important mathematical features are summarised in

FIGURE

6.2 The simple epidemic

This is a specially simple model in which we assume that removals do not occur and that an individual who catches the disease does not recover or die, shows no symptoms and is not isolated. The infected individual is thus capable of transmitting the disease to susceptibles forever. We also assume that the latent period is of zero length, so that we have only two categories of people, susceptibles and infectives.

Let $X(t)$ denote the number of susceptibles at time t. Similarly let $Y(t)$ denote the number of infectives. Let $X(0) = x_0$ and $Y(0) = y_0$. We shall always assume that we have a closed community of fixed size $(= n+1)$, so that

$$\begin{aligned} x_0 + y_0 = X(t) + Y(t) &= n+1, \\ \therefore \quad X(t) &= n + 1 - Y(t) \quad \forall t, \end{aligned}$$

so it suffices to work only with $Y(t)$.

The model

The model is a type of birth process ($Y(t)$ is always increasing), so we set

$$P(Y(t+\delta t) = y+1 | Y(t) = y) = \beta_y \delta t + o(\delta t).$$

What is the value of β_y, the epidemic rate?

We assume that for any two chosen members of the population, they come into contact according to a Poisson process with rate α/n. This is equivalent to the rate at which each individual comes into contact with the whole of the rest of the population being α.

Thus when there are y infectives, and so $n+1-y$ susceptibles (non-infectives), the rate at which infectives and susceptibles come into contact with each other is

$$y \times (n+1-y) \times \frac{\alpha}{n} = \frac{(n+1-y)y\alpha}{n}.$$

We shall assume that when an infective meets a susceptible the probability that the disease is passed on is β/α, where $\beta \le \alpha$.

Thus the rate of transmission from susceptibles to infectives is

$$\beta_y = \frac{(n+1-y)y\alpha}{n} \times \frac{\beta}{\alpha} = \frac{(n+1-y)y\beta}{n}.$$

Modelling as a birth process gives us the differential-difference equations for $p_y(t) = P\big(Y(t) = y\big)$:

$$p'_{y_0}(t) = -\frac{(n+1-y_0)y_0\beta}{n} p_{y_0}(t),$$
$$p'_y(t) = -\frac{(n+1-y)y\beta}{n} p_y(t) + \frac{(n+2-y)(y-1)\beta}{n} p_{y-1}(t)$$
$$(y = y_0+1, \ldots, n),$$
$$p'_{n+1}(t) = \beta p_n(t).$$

It is difficult to find a general solution to such a sequence of equations. However we can proceed iteratively for a small population (computationally for larger).

EXAMPLE 6.2.1. *In a household of size 5, initially 3 have an infectious disease. Assuming that $\beta = 2$, what is the probability distribution of the number of infected people at time t?*

SOLUTION. $n = 4$ and $y_0 = 3$, so $p_1(t) = p_2(t) = 0$ for all t. And

$$p'_3(t) = -\frac{2 \times 3 \times 2}{4} p_3(t) = -3p_3(t).$$

With $p_3(0) = 1$ this implies $p_3(t) = e^{-3t}$.

Next,

$$p'_4(t) = -2p_4(t) + 3p_3(t),$$
$$\therefore \quad \frac{d}{dt}\big(p_4(t)e^{2t}\big) = 3e^{-3t} \times e^{2t} = 3e^{-t},$$
$$\therefore \quad p_4(t)e^{2t} = A - 3e^{-t}.$$

With $p_4(0) = 0$ this implies $A = 3$, so $p_4(t) = 3e^{-2t} - 3e^{-3t}$.

Finally,

$$p'_5(t) = 6e^{-2t} - 6e^{-3t}, \quad p_5(0) = 0$$
$$\implies p_5(t) = 1 - 3e^{-2t} + 2e^{-3t}.$$

□

The duration of the simple epidemic

We define the *duration* of the epidemic to be the waiting time until the last remaining susceptible is infected. Then

$$P(W \le t) = P(Y(t) = n+1) = p_{n+1}(t).$$

Let T_y be the time taken for the number $Y(t)$ of infectives to rise from y to $y+1$, for $y = y_0, \ldots, n$. Then T_y is exponential with parameter β_y. The duration of the epidemic is the sum of these waiting times, which are independent:

$$W = \sum_{y=y_0}^{n} T_y.$$

The distribution of W is that of a sum of exponential r.v.s with different parameters and is thus complicated, but its expectation and variance are easier to find, using

$$EW = \sum_{y=y_0}^{n} \frac{1}{\beta_y}, \quad \operatorname{var} W = \sum_{y=y_0}^{n} \frac{1}{\beta_y^2}.$$

EXAMPLE 6.2.2. *Find the mean and variance of W in the previous example.*

SOLUTION.

$$EW = \sum_{y=3}^{4} \frac{1}{\beta_y} = \frac{1}{3} + \frac{1}{2} = \frac{5}{6},$$

$$\operatorname{var} W = \left(\frac{1}{3}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{13}{36}.$$

□

In general

$$\begin{aligned} EW &= \sum_{y=y_0}^{n} \frac{1}{\beta_y} \\ &= \sum_{y=y_0}^{n} \frac{n}{\beta y(n+1-y)} \\ &= \frac{n}{(n+1)\beta} \sum_{y=y_0}^{n} \left(\frac{1}{y} + \frac{1}{n+1-y}\right) \\ &= \frac{n}{(n+1)\beta} \left(\sum_{y=y_0}^{n} \frac{1}{y} + \sum_{i=1}^{n+1-y_0} \frac{1}{i}\right). \end{aligned}$$

If $y_0 = 1$ then

$$EW = \frac{2n}{(n+1)\beta} \sum_{y=1}^{n} \frac{1}{y}. \tag{6.2.1}$$

EXAMPLE 6.2.3. *What is the value of EW when $\beta = 1$, $y_0 = 1$ and*

(a) $n = 5$,
(b) $n = 10$,
(c) $n = 20$,
(d) $n = 200$,
(e) $n = 2000$?

SOLUTION ((a)–(c) only).

(a)

$$EW = \frac{137}{36} \simeq 3{\cdot}8056,$$

(b)

$$EW = \frac{671}{126} \simeq 5{\cdot}3254,$$

(c)

$$EW = \frac{279\,175\,675}{40\,738\,698} \simeq 6{\cdot}8528.$$

We'll find the others after a little theory. □

A useful result is that, for large n,

$$\sum_{y=1}^{n} \frac{1}{y} - \ln n \to \gamma \simeq 0{\cdot}57721 \quad (n \to \infty).$$

γ is called *Euler's constant.* So

$$\sum_{y=1}^{n} \frac{1}{y} \simeq \ln n + \gamma$$

for large n. The error is approximately $1/(2n)$.

SOLUTION CONTINUED. From the above we have the approximation

$$EW \simeq \frac{2n}{(n+1)\beta}(\gamma + \ln n),$$

and using this for parts (a)–(c) gives

(a) 3·6444,

(b) 5·2360,
(c) 6·8056

(to be compared with the exact evaluations above), while for parts (d) and (e) it gives

(d) $EW \simeq 11{\cdot}6926$,
(e) $EW \simeq 16{\cdot}3481$.

□

From (6.2.1) we find that as $n \to \infty$,

$$EW = \frac{2}{\beta}\left(1 - \frac{1}{n+1}\right)\big(\ln n + \gamma + \mathrm{o}(1)\big)$$
$$= \frac{2}{\beta}\big(\ln n + \gamma + \mathrm{o}(1)\big)$$

since

$$\frac{\ln n}{n+1} \to 0 \quad \text{as } n \to \infty.$$

The epidemic duration is thus approximately $2/\beta$ times $\ln n$.

6.3 The general epidemic

The general epidemic model is the same as the simple epidemic, except that we introduce 'removals' into the population. Individuals thus follow the path

Susceptible → Infective → Removed

or alternatively stay as susceptible if the disease runs its course before they become infected. This may occur if infectives become removed sufficiently quickly so that the infective category becomes empty before we run out of susceptibles.

We shall consider two questions:

- What is the probability that not everyone gets the disease?
- What is the expected number of people to get the disease?

Define $X(t)$, $Y(t)$ and $Z(t)$ as the number of susceptibles, infectives and removeds respectively at time t. Let $X(0) = x_0$, $Y(0) = y_0 \geq 1$, $Z(0) = z_0 = 0$.

Example 6.3.1. *Plague breaks out in a small hamlet, where there are 100 individuals. After a week 25 people have caught the disease, of which two have already died. After four weeks all trace of the disease has disappeared, leaving 65 people dead. Of the others 8 caught the disease, but recovered (it is assumed that they are now immune) and the remaining 27 never caught the disease. Express this information in terms of $X(t)$, $Y(t)$ and $Z(t)$.*

Solution. This says that with time measured in weeks,

$$X(1) = 100 - 25 = 75, \quad Y(1) = 25 - 2 = 23, \quad Z(1) = 2,$$

and

$$X(4) = 27, \quad Y(4) = 0, \quad Z(4) = 65 + 8 = 73.$$

□

The model

We know that $X(t) + Y(t) + Z(t) = n + 1$, so we need consider only the bivariate process $\big(X(t), Y(t)\big)$.

Values of $X(t)$ and $Y(t)$ may change in two ways:

- An infective meets a susceptible and passes the disease on. If there are x susceptibles and y infectives the transition is

$(x, y) \to (x-1, y+1)$, and this occurs at rate

$$\frac{\beta x y}{n}$$

as before.

- An infective is removed. We suppose that an individual who becomes infected remains an infective for a length of time exponentially distributed with parameter γ. Thus $(x, y) \to (x, y-1)$, and this occurs at rate $y\gamma$.

Expressed in the usual way,

$$\begin{aligned} &P(X(t+\delta t) = x-1, Y(t+\delta t) = y+1 | X(t) = x, Y(t) = y) \\ &\quad = \frac{\beta x y}{n}\delta t + o(\delta t), \\ &P(X(t+\delta t) = x, Y(t+\delta t) = y-1 | X(t) = x, Y(t) = y) \\ &\quad = \gamma y \delta t + o(\delta t). \end{aligned}$$

This process occurs on a triangular lattice

$$\{(x, y) : 0 \le x + y \le n+1\}.$$

The distribution of the process, i.e. the joint distribution of $X(t)$ and $Y(t)$, is complicated and we shall not consider it. We will consider the path of the population on the lattice as a two-dimensional (discrete-time) Markov chain.

Starting at (x, y) we either go to $(x-1, y+1)$ or $(x, y-1)$ if $y > 0$ (if $y = 0$ we are at the final position). Using similar reasoning to that for breaking down a Poisson process, the probability that we move to $(x-1, y+1)$ is

$$\frac{\frac{\beta x y}{n}}{\frac{\beta x y}{n} + \gamma y} = \frac{\beta x}{\beta x + n\gamma}$$

$$= \frac{x}{\rho + x}$$

where we set

$$\rho = \frac{n\gamma}{\beta}.$$

Thus the Markov chain has probabilities of transitions

$$\begin{aligned} P\big((x, y) \to (x-1, y+1)\big) &= \frac{x}{\rho + x}, \\ P\big((x, y) \to (x, y-1)\big) &= \frac{\rho}{\rho + x}. \end{aligned}$$

EXAMPLE 6.3.2. *Suppose that the population size is 4. Further suppose that $\gamma = 1{\cdot}2$ and $\beta = 1{\cdot}8$. After the end of the epidemic, $y = 0$. What is the probability distribution of the final value of x ($= X(\infty)$), if we start with 3 susceptibles and 1 infective?*

SOLUTION. $n = 3$. The starting state is $(x, y) = (3, 1)$ and we have $\rho = 3 \times 1{\cdot}2/1{\cdot}8 = 2$. The probabilities of the paths by which the epidemic runs its course are as follows:

FIGURE

The distribution of the final value of x is thus

$$\begin{aligned} &P\big(X(\infty) = 3\big) = P\big((3,1) \to (3,0)\big) = \frac{2}{5} = 0{\cdot}4, \\ &P\big(X(\infty) = 2\big) = P\big((3,1) \to (2,2) \to (2,1) \to (2,0)\big) \\ &\quad = \frac{3}{5} \times \frac{2}{4} \times \frac{2}{4} = \frac{3}{20} = 0{\cdot}15, \\ &P\big(X(\infty) = 1\big) \\ &\quad = P\big((3,1) \to (2,2) \to (1,3) \to (1,2) \to (1,1) \to (1,0)\big) \\ &\qquad + P\big((3,1) \to (2,2) \to (2,1) \to (1,2) \to (1,1) \to (1,0)\big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3} \;+\; \frac{3}{5}\times\frac{2}{4}\times\frac{2}{4}\times\frac{2}{3}\times\frac{2}{3} \\
&= \frac{4}{45}+\frac{3}{45}=\frac{7}{45}\doteqdot 0{\cdot}156, \\
P\big(X(\infty)=0\big) &= 1-\frac{2}{5}-\frac{3}{20}-\frac{7}{45}=\frac{53}{180}\doteqdot 0{\cdot}294.
\end{aligned}
$$

□

Example 6.3.3. *In the previous Example, what is the probability that not everyone gets the disease? What is the expected number who contract the disease?*

Solution. The probability that not everyone gets the disease is

$$1-P\big(X(\infty)=0\big)=1-\frac{53}{180}=\frac{127}{180}\doteqdot 0{\cdot}706.$$

Also

$$EX(\infty)=3\times\frac{2}{5}+2\times\frac{3}{20}+1\times\frac{7}{45}=\frac{149}{90},$$

so the expected number of people who contract the disease (not including the original infected individual) is

$$3-\frac{149}{90}=\frac{121}{90}\doteqdot 1{\cdot}344.$$

□

The distribution of $X(\infty)$ is called the *survivor distribution.* Its shape in the present case is as follows:

FIGURE

6.4 The threshold

Notice that the probability distribution of the number of those who do not catch the disease in our example is 'U-shaped' (the values at the extremes are larger than those in the middle).

For a large population a U-shaped distribution implies that either very few or very many people will catch the disease. The following charts show the survivor distributions for $n=20$, $x_0=20$, $y_0=1$, for values for ρ of 5, 10, 20 and 40.

FIGURE

The distributions for $\rho=5$ and $\rho=10$ are U-shaped, the others are not (there is likely to be only a small outbreak for these last two cases). In general it has been found that U-shaped survivor distributions occur when $x_0>\rho$. That is, there is a *threshold phenomenon.*

Why does this occur? We consider an approximation. Suppose y_0 is small compared to x_0. In the early stages of an epidemic, the rate of change of the number of infectives is as follows.

$$
\begin{aligned}
y\to y+1:&\ \text{ rate } \frac{\beta yx}{n}\doteqdot\frac{\beta yx_0}{n},\\
y\to y-1:&\ \text{ rate } \gamma y.
\end{aligned}
$$

This is equivalent to a simple birth-death process with birth rate $\beta x_0/n$ and death rate γ. For this birth-death process, if the birth rate is less than the death rate, i.e.

$$\frac{\beta x_0}{n}\le\gamma \iff x_0\le\rho$$

then the population is bound to become extinct. If the birth rate is greater than the death rate, there is probability

$$\left(\frac{\gamma}{\beta x_0/n}\right)^{y_0}=\left(\frac{\rho}{x_0}\right)^{y_0}$$

that the population becomes extinct, and otherwise the population tends to infinity.

Back to the epidemic: the 'population' is the population of infectives. Thus if $x_0 \le \rho$ there will probably be only a small outbreak, whereas if $x_0 > \rho$ there is approximate probability

$$\left(\frac{\rho}{x_0}\right)^{y_0}$$

of a small outbreak, and approximate probability

$$1 - \left(\frac{\rho}{x_0}\right)^{y_0}$$

of a large outbreak, with little chance of a medium-sized outbreak.

Note that in general, infection spreads at its fastest when about half the population are infected, so the disease is unlikely to die out at that point.

EXAMPLE 6.4.1. *Consider the following examples of a general epidemic model:*

(a) $x_0 = 20$, $y_0 = 2$, $z_0 = 0$, $\beta = 4{\cdot}2$, $\gamma = 1$;
(b) $x_0 = 222$, $y_0 = 4$, $z_0 = 0$, $\beta = 1$, $\gamma = 1$.

What is the probability, approximately, of a major outbreak in each case?

SOLUTION.

(a) $22 = x_0 + y_0 = n + 1$ so $n = 21$. Then

$$\rho = \frac{n\gamma}{\beta} = \frac{21 \times 1}{4{\cdot}2} = 5 < x_0.$$

Thus

$$P(\text{major outbreak}) \simeq 1 - \left(\frac{\rho}{x_0}\right)^{y_0} = 1 - \left(\frac{5}{20}\right)^2 = \frac{15}{16}.$$

(b) $226 = x_0 + y_0 = n + 1$ so $n = 225$. Then

$$\rho = \frac{n\gamma}{\beta} = \frac{225 \times 1}{1} = 225 > x_0.$$

Thus

$$P(\text{major outbreak}) \simeq 0.$$

□