

Table 2.23 Fetal lamb movements

0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	2	2	0	0	0	0	0	1	0	0	0
1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	7	3	2	3	2	4	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	2	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0
1	0	0	1	0	0	0	1	2	0	0	0	1	0	1	0	1	0
1	0	1	0	0	2	0	1	2	1	1	2	1	0	1	0	1	0
1	0	0	1	1	0	0	0	1	1	1	0	4	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.24 Fog and mist occurrence at Christchurch aerodrome

		Next		
		Clear	Mist	Fog
Now	Clear	13,650	245	30
	Mist	480	427	71
	Fog	11	171	198

overall distribution of weather type has clear weather 95.1% of the time, mist 3.4% of the time, and fog 1.5% of the time. Test whether the hours around sunrise are reasonably described by these stationary probabilities.

CHAPTER 3

Continuous time Markov chains

A natural generalization of the Markov chain in discrete time is to allow the times between transitions to be continuous random variables. The motivating example concerns multiplicative showers of cosmic radiation. The difference between deterministic and stochastic models is illustrated using linear and nonlinear birth processes. The relation to the previous chapter is brought out in the discrete skeleton and the jump chain. Some queueing models are presented, and the statistical inference for both nonparametric and parametric models of continuous time Markov processes is discussed. Partially observed processes are used to model the activity of synaptic nerve cells. The blood production in cats is described using a hidden Markov process.

3.1. The soft component of cosmic radiation

High-energy particles, many near the speed of light, continuously bombard the atmosphere from all directions. This cosmic radiation, discovered by V. F. Hess during a balloon flight in 1912, consists primarily of nuclei of the most abundant elements. It has two clearly distinguishable parts, the **soft** and **hard** components. The former, on which we shall concentrate, consists of very energetic electrons together with gamma ray photons, while the latter consists of mesons.

When soft component cosmic rays pass through the atmosphere, they collide with atmospheric atoms, producing showers of additional particles. If the number of shower particles is measured as a function of height above the ground, it is found to have a maximum at about 16 km. Very few of the primary rays reach the surface of the earth. When an energetic photon travels a distance Δt through some substance, such as lead or air, it has a certain chance of being absorbed by an atom and emitting a pair of particles consisting of one positron and one electron. By doing so it loses all its energy. This process is called **pair production**. The probability is proportional to Δt for small distances, and the

factor of proportionality is a function of the energies of the parent photon and the resulting electron offspring. A fast electron (or positron) radiates high energy photons when it is absorbed in the nucleus of an atom (**Bremsstrahlung**). Again, the probability is proportional to the distance traveled, and the factor of proportionality is again a function of the energies involved.

In order to measure the cosmic radiation one places a number of particle counters (such as Geiger–Müller tubes) in a pattern around an absorber, such as a block of lead. The absorber is expected to produce showers as described above, and if the counters are wired to register only simultaneous registration at several counters, one can be fairly confident that a shower phenomenon has been observed. The number of coincidences per unit time is measured and plotted against the thickness of the lead. Figure 3.1 illustrates the results of an experiment of Schwegler, with curve III corresponding to the soft component and curve II to the hard component.

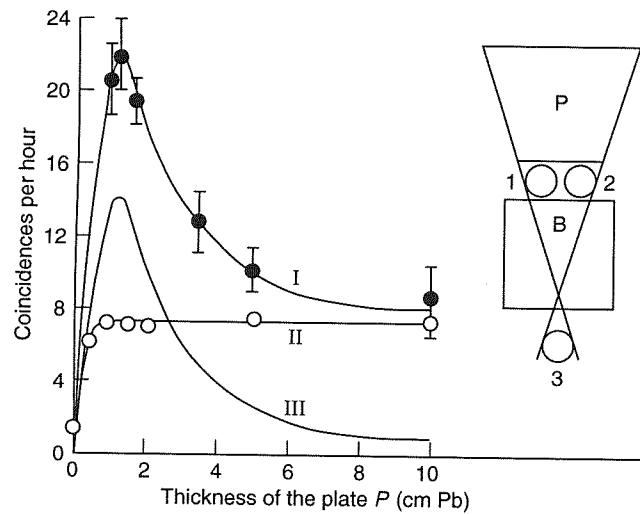


Figure 3.1. Coincidence counts of cosmic radiation as a function of lead absorber thickness. Curve I is obtained without the lead block B between the counters, curve II with B placed between the counters. Curve III is the difference between the two curves. This figure is an adaptation of one published in Arley (1943). Reproduced with permission from L. Janossy: *Cosmic Rays*, published by Oxford University Press, 1948; and with the permission of the estate of Dr. Arley.

We start with a very simple model that was proposed by Furry (1937). He disregarded the photons, and assumed that one electron by traveling a distance Δt could be converted into two electrons, with a probability asymptotically (as $\Delta t \rightarrow 0$) proportional to Δt . All electrons were assumed to act independently. Let $P_{1n}(t)$ denote the probability, starting from a single electron, of finding n electrons at depth t , where depth 0 is the top of the atmosphere. Then Furry's model says that the conditional probability of finding $n+m$ electrons at depth $t+\Delta t$, given that there were n at depth t , is given by

$$\frac{P_{1,n+m}(t+\Delta t)}{P_{1n}(t)} = \begin{cases} n\lambda\Delta t + o(\Delta t) & \text{if } m=1 \\ o(\Delta t) & \text{if } m>1 \\ 1-n\lambda\Delta t + o(\Delta t) & \text{if } m=0. \end{cases} \quad (3.1)$$

Rewriting this slightly, we see that

$$\begin{cases} P_{10}(t) \equiv 0 \\ P_{11}(t+\Delta t) = P_{11}(t)(1-\lambda\Delta t) + o(\Delta t) \\ P_{1n}(t+\Delta t) = P_{1n}(t)(1-n\lambda\Delta t) + P_{1,n-1}(t)(n-1)\lambda\Delta t + o(\Delta t), \quad n \geq 2. \end{cases} \quad (3.2)$$

By manipulating these equations and letting $\Delta t \rightarrow 0$ we see that

$$\begin{cases} P_{10}(t) \equiv 0 \\ \frac{dP_{11}(t)}{dt} = -\lambda P_{11}(t) \\ \frac{dP_{1n}(t)}{dt} = \lambda(n-1)P_{1,n-1}(t) - \lambda n P_{1n}(t). \end{cases} \quad (3.3)$$

Adding the initial condition $P_{1n}(0)=1(n=1)$ it is easy to solve these equations recursively, getting

$$\begin{cases} P_{10}(t) \equiv 0 \\ P_{1n}(t) = e^{-\lambda t}(1-e^{-\lambda t})^{n-1}. \end{cases} \quad (3.4)$$

In other words, the number of electrons at depth t has a positive geometric distribution with parameter $\exp(-\lambda t)$. The mean number of electrons therefore is $\exp(\lambda t)$, which is increasing in t , and therefore does not describe the observed mean values very well. This process is an example of a **pure birth process** (a process that only moves upwards in unit steps), and was first introduced in a paper by McKendrick (1914), and rediscovered by Yule¹ (1924). The process is now often called the Yule–Furry process, or the linear birth process.

¹Yule, George Udny (1871–1951). English statistician. Student of Karl Pearson. Introduced the multiple correlation coefficient.

There are at least three features of the actual shower process that are disregarded in Furry's model. First, the absorption of electrons is disregarded, second, the alternating character of the successive generations (photons \rightarrow electrons \rightarrow photons \cdots) is not taken into account, and, third, the energy decrease resulting from the production of photons is neglected. Later we will modify the model to produce something closer to the observed data.

3.2. The pure birth process

Now consider a generalization of the Yule–Furry model, by letting the conditional probability of a birth in the time interval $(t, t + \Delta t)$, given n births at a given time t , be $\lambda_n \Delta t$, with an error of smaller order in Δt . This is called the **general birth process**. The linear model (3.1) has $\lambda_n = \lambda n$. Arguing as when deriving (3.3), and starting from $X(0) = 1$, we obtain the system of equations

$$\begin{cases} P_{10}(t) \equiv 0 \\ \frac{dP_{11}(t)}{dt} = -\lambda_1 P_{11}(t) \\ \frac{dP_{1n}(t)}{dt} = \lambda_{n-1} P_{1,n-1}(t) - \lambda_n P_{1n}(t) \end{cases} \quad (3.5)$$

From the second equation in the system (3.5) we see that $P_{11}(t) = \exp(-\lambda_1 t)$, whence

$$\frac{dP_{12}(t)}{dt} = \lambda_1 \exp(-\lambda_1 t) - \lambda_2 P_{12}(t). \quad (3.6)$$

It turns out to be convenient to use Laplace transforms (much as we needed generating functions in the previous chapter). The Laplace transform f^* of a function $f : \mathbf{R}_+ \rightarrow \mathbf{R}$ is defined by

$$f^*(\theta) = \int_0^\infty e^{-\theta x} f(x) dx \quad (3.7)$$

which exists for $\theta > 0$ whenever f is integrable. Under some regularity conditions involving continuity, the Laplace transform determines f uniquely. In fact, f can be computed from f^* using integration in the complex plane. Here we are mainly interested in the following properties:

Lemma 3.1

- (a) If $h(x) = f(x) + g(x)$, then $h^*(\theta) = f^*(\theta) + g^*(\theta)$.
- (b) If $h(x) = \int_0^\infty g(x-y)f(y)dy$, then $h^*(\theta) = g^*(\theta)f^*(\theta)$.
- (c) $\left[f'(t) \right]^*(\theta) = \theta f^*(\theta) - f(0)$.
- (d) $\left[\int_0^\infty f(s)ds \right]^*(\theta) = f^*(\theta)/\theta$.

The proofs are straightforward computations using integration by parts, and are left as Exercise 1.

Example (Some Laplace transforms) If $f(x) = 1$ then $f^*(\theta) = 1/\theta$, while if $f(x) = x^\alpha \exp(-\beta x)$, then $f^*(\theta) = \Gamma(\alpha+1)(\beta+\theta)^{-(\alpha+1)}$. Also if $f(x) = g(\alpha x)$, we have $f^*(\theta) = g^*(\theta/\alpha)/\alpha$. \square

The Laplace transform (over t) of the third equation in (3.5) is (for $n > 1$)

$$(\theta + \lambda_n) P_{1n}^*(\theta) = \lambda_{n-1} P_{1,n-1}^*(\theta). \quad (3.8)$$

This is a difference equation in n with nonconstant coefficients, so the theory in Appendix B does not apply. However, the solution is easily obtained recursively as

$$P_{1n}^*(\theta) = \frac{\lambda_1}{\theta + \lambda_1} \frac{\lambda_2}{\theta + \lambda_2} \cdots \frac{1}{\theta + \lambda_n}. \quad (3.9)$$

Notice that the first $n-1$ terms in (3.9) are Laplace transforms of exponential densities with parameter λ_i , while the last term is the Laplace transform of the function $\exp(-\lambda_n x)$, which is the probability that an $\text{Exp}(\lambda_n)$ -distributed random variable exceeds x . Using Lemma 3.1(b), and recalling that convolution of densities corresponds to a sum of the corresponding independent random variables, we see that the process reaches n in time t provided the sum of $n-1$ exponential random variables with parameters λ_i is less than t , and that adding an independent $\text{Exp}(\lambda_n)$ -random variable makes the sum greater than t . We shall return to this construction of the process later.

Summing (3.5) over n we get

$$\sum_{n=1}^N \frac{dP_{1n}(t)}{dt} = -\lambda_N P_{1N}(t), \quad (3.10)$$

so by Laplace transforming and using (3.9) we obtain

$$\theta \sum_{n=1}^N P_{1n}^*(\theta) = 1 - \lambda_N P_{1N}^*(\theta) = 1 - \prod_{n=1}^N \left[1 + \frac{\theta}{\lambda_n} \right]^{-1}. \quad (3.11)$$

By letting $N \rightarrow \infty$ we see that $\sum P_{1n}^*(\theta) \rightarrow 1/\theta$ if and only if the product diverges to zero, i.e., iff $\sum 1/\lambda_i = \infty$. In other words, $\sum P_{1n}(t) = 1$. This proves the following result.

Proposition 3.1 For any $t > 0$ $\sum_{n=1}^\infty P_{1n}(t) = 1$ iff $\sum \frac{1}{\lambda_i} = \infty$.

The interpretation of this result is that if the λ_i grow too fast, so that $\sum 1/\lambda_i < \infty$, then there is positive probability that the process at time t is not anywhere in the state space! In other words, it has gone off to infinity, or **exploded**. Such processes are called **dishonest**, and create a lot of problems in the theory. In particular, what ought to happen after an explosion?

Example (Linear birth process) In the case of a linear birth process we have $\lambda_i = i\lambda$, so that $\sum 1/\lambda_i = 1/\lambda \sum 1/i$ which is infinite. Hence the linear birth process does not explode in finite time. \square

Example (The Pólya process) Consider the Pólya urn model described in section 2.5. Fix $\tau > 0$, and assume that the urn scheme is carried out every τ time units. Let R_t be the number of red balls drawn at time $[t/\tau]$. Recall from equation (2.129) that

$$\mathbf{P}(R_{(n+1)\tau} = i+1 \mid R_{n\tau} = i) = \frac{R+i}{N+d}. \quad (3.12)$$

Let $p(\tau) = R/N$, $c(\tau) = d/N$, and $t = n\tau$. Then $p(\tau)/\tau$ is the rate of red balls drawn per unit time, and we may set things up so that $p(\tau)/\tau \rightarrow \lambda$ as $\tau \rightarrow 0$. Assume further that $c(\tau)/\tau \rightarrow \alpha > 0$, and consider the limit of R_t as $\tau \rightarrow 0$, which we, abusing notation, also call R_t . This birth process has transition probabilities

$$P_{i,i+1}(t, t+dt) = \lambda_i(t)dt + o(dt) \text{ as } dt \rightarrow 0 \quad (3.13)$$

where $\lambda_i(t) = \lambda(1+i\alpha)/(1+t\alpha)$. We see that the rate of births is decreasing with time, but for any fixed time roughly linear in population size. Starting from $R_0=0$ it is straightforward to verify (Exercise 2) that

$$\mathbf{P}^0(R_t = n) = \frac{(\lambda t)^n \prod_{k=1}^{n-1} (1+k\alpha)}{n! (1+\alpha\lambda t)^{n+1/\alpha}} \quad (3.14)$$

where, as before, \mathbf{P}^x denotes probability starting from $X(0)=x$. The Pólya process can be used as a model of neutron showers, which takes into account the loss of energy in particles as the shower proceeds. From (3.14) we see that $\mathbf{E}^0(R_t) = \lambda t$, which is again an increasing function of t , albeit growing much slower than the mean of a Yule–Furry process. Consequently the Pólya process is also an inadequate model of cosmic radiation. \square

The deterministic counterpart to a linear birth process satisfies the differential equation

$$\frac{dn_t}{dt} = \lambda n_t \quad (3.15)$$

with solution $n_t = n_0 \exp(\lambda t)$. One may expect that the deterministic equation would describe the average behavior of the stochastic model. Since

$X_t \sim \text{Geom}(\exp(-\lambda t))$, as shown in section 3.1, it has mean $\exp(\lambda t)$ when starting from one individual. Since we can think of the process starting from n_0 individuals as the sum of n_0 iid processes, the mean then is $n_0 \exp(\lambda t)$. Hence, in this case, the deterministic equation describes the average behavior of the system. The next example shows, however, that this is not generally the case.

Example (Sociology) Consider a social group of N individuals. Following Bartholomew (1973, Ch. 9) we shall build a simple model for the transmission of information through the group. The information originates from a source such as a television commercial, and is spread by word of mouth or by direct contact with the source. We assume that any individual in the group encounters the source at a constant rate α , and that the group is homogeneous, so that any pair of individuals has the same rate β of exchanging information.

Let $X(t)$ be the number of individuals that have received the information by time t , starting with $X(0)=0$. Presumably individuals will not forget the information, so $X(t)$ should be a birth process. The intensity of transition from state n to state $n+1$ is

$$\lambda_n = (N-n)\alpha + n(N-n)\beta = (N-n)(\alpha+n\beta) \quad (3.16)$$

since the $N-n$ individuals who are not yet informed either can receive information from the source or from any of the n informed individuals. Using (3.9) and Lemma 3.1 we see that

$$\mathbf{P}(X(t)=n) = (-1)^n \left[\prod_{i=0}^{n-1} \lambda_i \right] \sum_{j=0}^n \frac{e^{-\lambda_j t}}{\prod_{j \neq k} (\lambda_j - \lambda_k)}, \quad n \leq N. \quad (3.17)$$

A quantity of interest is the expected value of $X(t)$. Its derivative is (in a different application) often called the **epidemic curve** and depicts the average rate of growth of the process. While it is possible to compute $\mathbf{E}X(t)$ directly from (3.17) (see Haskey, 1954), we will take a roundabout route to obtain an approximate expression for it. Let T_n be the time when the n th person becomes informed. We can write $T_n = \sum_0^{n-1} \tau_i$ where the $\tau_i \sim \text{Exp}(\lambda_i)$ are independent. Hence

$$\begin{aligned} ET_n &= \sum_0^{n-1} \frac{1}{\lambda_i} = \sum_0^{n-1} \frac{1}{(N-i)(\alpha+i\beta)} \\ &= \frac{1}{\alpha+N\beta} \sum_{i=0}^{n-1} \frac{1}{N-i} + \frac{\beta}{\alpha+N\beta} \sum_{i=0}^{n-1} \frac{1}{\alpha+i\beta} \\ &= \frac{1}{\beta(N+\delta)} (\psi(n) - \psi(N-n) - \psi(\delta-1) + \psi(n+\delta-1)) \end{aligned} \quad (3.18)$$

where $\delta = \alpha/\beta$ and ψ is the **digamma function**, given by

$$\psi(x) = \sum_{i=1}^{\infty} \left[\frac{1}{i} - \frac{1}{i+x} \right]. \quad (3.19)$$

For large x we can approximate $\psi(x) \approx \log x + \gamma$ where γ is Euler's constant $0.5772\ldots$ (Abramowitz and Stegun, 1965, p. 259; our $\psi(x)$ is their $\psi(x+1)+\gamma$). Consider large values of N , and let $n=Np$ for some $0 < p < 1$. Then we have the approximation

$$ET_{Np} \approx \frac{1}{\beta(N+\delta)} \left(\log \left(\frac{\beta N + \delta - 1}{1-p} \right) + \gamma - \psi(\delta-1) \right). \quad (3.20)$$

Now let $F(t)$ be the distribution function for the time until a randomly chosen individual receives the information. From the times T_1, \dots, T_N of exposures we would estimate F by $F_N(t) = \#\{T_i \leq t\}/n$. In particular, we can write $T_{Np} = F_N^{-1}(p)$. Assuming that N is large we have $E F_N^{-1}(p) \approx F^{-1}(p)$, so $ET_{Np} \approx F^{-1}(p)$, where the approximation is quite good for values of p away from 0 or 1. Hence

$$F(p) \approx \frac{\exp(\beta(N+\delta)(p-\gamma+\psi(\delta-1)) + 1-\delta)}{\exp(\beta(N+\delta)(p-\gamma+\psi(\delta-1)) + N)}. \quad (3.21)$$

Finally $EX(t) = NF(t)$.

A deterministic counterpart to this problem would assume that (3.16) holds exactly in each infinitesimal time interval $(t, t+dt)$, so

$$\frac{dx(t)}{dt} = (N-x(t))(\alpha + x(t)\beta). \quad (3.22)$$

The solution, with $x(0)=0$, is

$$x(t) = N \frac{\exp(\beta(N+\delta)t - 1)}{\exp(\beta(N+\delta)t) + N/\delta} \quad (3.23)$$

which is different from the stochastic counterpart $EX(t)$. Figure 3.2 shows the two curves in one case.

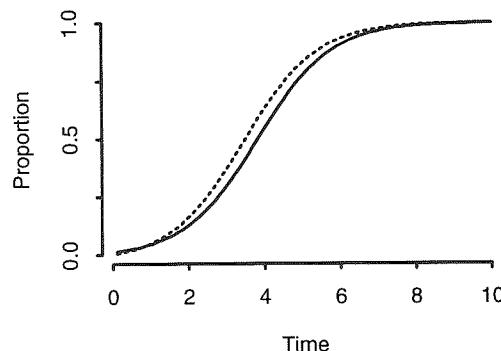


Figure 3.2. Deterministic description (dotted line) and mean of stochastic description (solid line) for the information transmission model. Here $N=50$, $\alpha=0.0292$, and $\beta=0.02$.

Thus we see that it is not generally the case that a deterministic counterpart to a stochastic model is an expression for the mean of the stochastic model. This is, in fact, only the case for models that are linear in population size. \square

3.3. The Kolmogorov equations

In this chapter we are concerned with Markovian stochastic processes with continuous time and discrete state space S , represented as the integers (or a subset of the integers) as for the discrete time chains in the previous chapter. We define the Markov property of a process $X(t)$ by

$$\mathbf{P}(X(t) = k \mid X(t_1) = k_1, \dots, X(t_n) = k_n) = P_{k_n, k}(t_n, t) \quad (3.24)$$

for $0 \leq t_1 \leq \dots \leq t_n \leq t$, and any integers n, k, k_1, \dots, k_n . If $P_{ij}(s, t) = P_{ij}(|t-s|)$ the process has **stationary transition probabilities**. Unless otherwise stated this property will be assumed henceforth. As we saw in the previous section, there is a possibility that these processes may reach infinity in finite time. We will assume that if this happens the process stays at infinity for ever (infinity is then called a **coffin state**). This is called the **minimal construction** (see, e.g., Asmussen, 1987, section II.2, for a discussion). The following facts are easily established.

Proposition 3.2

$$0 \leq P_{ij}(t) \leq 1. \quad (3.25)$$

$$\sum_j P_{ij}(t) \leq 1. \quad (3.26)$$

$$P_{ik}(s+t) = \sum_j P_{ij}(s) P_{jk}(t). \quad (3.27)$$

$$P_{ij}(0) = 1(i=j). \quad (3.28)$$

Proof The first and fourth statements are trivial, while the second follows from

$$\sum_j P_{ij}(t) = \mathbf{P}(0 \leq X(t) < \infty \mid X(0)=i). \quad (3.29)$$

If the inequality is strict, the process is dishonest. To show the third equation, the **Chapman–Kolmogorov equation**, we compute

$$\begin{aligned} P_{ik}(t+s) &= \mathbf{P}^i(X(t+s)=k) = \sum_j \mathbf{P}^i(X(s)=j) \mathbf{P}^i(X(t+s)=k \mid X(s)=j) \\ &= \sum_j P_{ij}(s) P_{jk}(t), \end{aligned} \quad (3.30)$$

noticing that the coffin state is ruled out by our construction. \square

As in the case of discrete time it is convenient to express things in matrix notation. Let $(\mathbb{P}(t)) = (P_{ij}(t), i, j \in S)$. Then (3.27) can be written

$$\mathbb{P}(s+t) = \mathbb{P}(s)\mathbb{P}(t). \quad (3.31)$$

Incidentally, $(\mathbb{P}(t); t \geq 0)$ is therefore a **semigroup**. It is **stochastic** if there is equality in (3.26), and **substochastic** otherwise. In order to proceed we need to assume some regularity. We call the process (or the semigroup) **standard** if the transition probabilities are continuous at 0, i.e., if

$$\lim_{t \downarrow 0} P_{ij}(t) = P_{ij}(0). \quad (3.32)$$

We will restrict attention to standard stochastic semigroups. Even then pathologies, such as a process which never stays in any state, can happen (e.g., Blackwell, 1958).

Lemma 3.2 Let $(\mathbb{P}(t))$ be a standard stochastic semigroup. Then $P_{ij}(t)$ is a continuous function for all i, j .

Proof We show that for any j

$$|P_{ij}(t+h) - P_{ij}(t)| \leq 1 - P_{ii}(h), \quad h > 0. \quad (3.33)$$

From the Chapman–Kolmogorov equation (3.27) we have that

$$P_{ij}(t+h) = \sum_k P_{ik}(h)P_{kj}(t) \quad (3.34)$$

so

$$P_{ij}(t+h) - P_{ij}(t) = (P_{ii}(h) - 1)P_{ij}(t) + \sum_{k \neq i} P_{ik}(h)P_{kj}(t). \quad (3.35)$$

But $P_{kj}(t) \leq 1$, so the second term on the right-hand side is bounded by $\sum_{k \neq i} P_{ik}(h) = 1 - P_{ii}(h)$. Hence

$$P_{ij}(t+h) - P_{ij}(t) \leq (1 - P_{ii}(h))(1 - P_{ij}(t)), \quad (3.36)$$

whence the claim follows. \square

In fact, it is possible to prove more (although we omit the proof here; see, e.g., Freedman, 1983, Theorem 21 in Chapter 5).

Proposition 3.3 For a standard stochastic semigroup $(\mathbb{P}(t))$ we have

- (i) $P_{ii}'(0)$ exists and is nonpositive (but not necessarily finite);
- (ii) $P_{ij}'(0)$ exists and is finite for $i \neq j$.

Let $\mathbf{Q} = (Q_{ij}) = (P_{ij}'(0))$. \mathbf{Q} is sometimes called the **generator** of $(\mathbb{P}(t))$. The following result is left as Exercise 3.

Lemma 3.3 If S is finite then $\sum_j Q_{ij} = 0$, while if S is countably infinite we have $\sum_j Q_{ij} \leq 0$.

For the moment, let us concentrate on the case of finite S . We will look at extensions to countable state space a little later. Using Proposition 3.3 and Lemma 3.3 we can define $q_i = \sum_{j \neq i} Q_{ij}$, so $q_i = -Q_{ii}$. For any t write the Taylor expansions

$$P_{ij}(h) = P_{ij}(t, t+h) = Q_{ij}h + o(h) \quad (3.37)$$

and

$$P_{ii}(h) = P_{ii}(t, t+h) = 1 - q_i h + o(h). \quad (3.38)$$

It follows that

$$\begin{aligned} \mathbf{P}(\text{transition } i \rightarrow j \text{ in } (t, t+h) \mid X(t)=i; \text{ transition occurred}) \\ = \frac{P_{ij}(h)}{1 - P_{ii}(h)} = \frac{Q_{ij}h + o(h)}{q_i h + o(h)} \rightarrow \frac{Q_{ij}}{q_i} \text{ as } h \rightarrow 0, \end{aligned} \quad (3.39)$$

provided that $q_i > 0$. We call Q_{ij} the **intensity** of the transition $i \rightarrow j$.

Example (The Yule–Furry process) We have $P_{j,j+1}(\Delta t) = j\lambda\Delta t + o(\Delta t)$ from the development in section 3.1, and $P_{jk}(\Delta t) = o(\Delta t)$ for $k \neq j+1$. Hence

$$Q_{j,j+1} = j\lambda \quad (3.40)$$

and

$$q_j = \sum_{k \neq j} Q_{jk} = j\lambda. \quad (3.41)$$

The intensity of new events increases with the number of events that have happened in the past. In other words, the probability of an event happening in the next time interval is proportional to how many events have occurred in the lifetime of the process. \square

The following result is central to the theory of continuous time Markov chains.

Theorem 3.1 The transition probabilities of a finite state space, continuous time Markov chain satisfy

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \in S} P_{ik}(t)Q_{kj} = \sum_{k \in S} Q_{ik}P_{kj}(t). \quad (3.42)$$

Proof

By the Chapman–Kolmogorov equations

$$P_{ij}(t+h) = \sum_k P_{ik}(t)P_{kj}(h). \quad (3.43)$$

Hence

$$P_{ij}(t+h) = P_{ij}(t)P_{jj}(h) + \sum_{k \neq j} P_{ik}(t)P_{kj}(h) \quad (3.44)$$

so

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = -P_{ij}(t)\left(\frac{1 - P_{jj}(h)}{h}\right) + \sum_{k \neq j} P_{ik}(t)\frac{P_{kj}(h)}{h}. \quad (3.45)$$

The right-hand side of (3.45) converges to

$$-P_{ij}(t)q_j + \sum_{k \neq j} P_{ik}(t)Q_{kj} = \sum_k P_{ik}(t)Q_{kj}. \quad (3.46)$$

The left-hand side therefore converges to the derivative of P_{ij} . This system of equations is called Kolmogorov's **forward equations**. To get the second equality in the statement of the theorem we proceed in a similar fashion, but now writing

$$P_{ij}(t+h) = \sum_{k \in S} P_{ik}(h)P_{kj}(t). \quad (3.47)$$

The resulting set of equations is called the **backward equations**. \square

Remark

(i) The theorem is true also for a large class of processes with infinite state space. The necessary assumptions have to do with assuring the smoothness of the functions $P_{ij}(t)$. Call the semigroup $(\mathbb{P}(t))$ **uniform** if $P_{ii}(t) \rightarrow 1$ as $t \downarrow 0$ uniformly in i . There is an easy criterion for uniformity (for a proof, see Freedman, 1983, Theorem 29 in Chapter 5).

Lemma 3.4 $(\mathbb{P}(t))$ is uniform if

$$\sup_{i \in S} q_i < \infty. \quad (3.48)$$

In particular, if S is finite, the semigroup is uniform. Theorem 3.1 is true for uniform semigroups. In addition, $(\mathbb{P}(t))$ is uniform whenever $\sum_j Q_{ij} = 0$ (this is part of the result in Freedman, 1983) so, in essence, the process behaves just like the finite state space case.

(ii) In matrix notation, writing

$$\frac{d\mathbb{P}(t)}{dt} = \begin{pmatrix} dP_{jk}(t) \\ \vdots \\ dP_{0k}(t) \end{pmatrix}, \quad (3.49)$$

we can restate Theorem 3.1 in the following fashion. The forward equation becomes

$$\frac{d\mathbb{P}(t)}{dt} = \mathbb{P}(t)Q, \quad (3.50)$$

and the backward equation becomes

$$\frac{d\mathbb{P}(t)}{dt} = Q\mathbb{P}(t). \quad (3.51)$$

\square

Example (General birth process) Note that equations (3.3) and (3.5) are the forward equations for the Yule–Furry and the general birth processes, respectively. For the general birth process the intensity matrix is

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \quad (3.52)$$

and $q_i = \lambda_i$. If $\sum_i \lambda_i^{-1} < \infty$, i.e., the process is dishonest, then $\lambda_i^{-1} \rightarrow 0$ so $\sup_i q_i = \infty$ and $(\mathbb{P}(t))$ is not uniform. An honest birth process (such as the Yule–Furry process) may or may not be uniform; in fact, the Yule–Furry process is not uniform since $q_i = i\lambda$. Nevertheless, the Yule–Furry process satisfies both the backward and forward equations. A general honest birth process has a unique solution to the forward equation, but may have many solutions to the backward equation. \square

We often prefer to define a process in terms of its intensities. For a very simple example, assume that we have a birth process which has constant intensity of births. That means that

$$Q_{j,j+1} = \lambda, \quad Q_{jj} = -\lambda \quad (3.53)$$

with all other entries being zero. This is a uniform semigroup (provided that $\lambda < \infty$), and the forward equation yields

$$\frac{dP_{jk}(t)}{dt} = -\lambda P_{jk}(t) + \lambda P_{j,k-1}(t). \quad (3.54)$$

In particular, if $j = 0$ we get

$$\frac{dP_{0k}(t)}{dt} = -\lambda P_{0k}(t) + \lambda P_{0,k-1}(t). \quad (3.55)$$

The initial conditions are taken to be $P_{00}(0) = 1$ and $P_{0i}(0) = 0$ for $i \geq 1$, so that the process starts in state 0. In order to solve this differential equation, we will attempt to convert it into a partial differential equation for the probability

generating function

$$G(s;t) = \mathbb{E}s^{X(t)} = \sum_i P_{0i}(t)s^i. \quad (3.56)$$

Multiplying both sides of (3.55) by s^k and summing over k we compute

$$\frac{\partial G(s;t)}{\partial t} = -\lambda G(s;t) + \lambda s G(s;t) \quad (3.57)$$

and $G(s;0)=1$. For a fixed value of s we see that

$$\frac{\partial G(s;t)}{\partial t} = -\lambda(1-s)G(s;t) \quad (3.58)$$

so that

$$G(s;t) = A(s)e^{-\lambda(1-s)t}. \quad (3.59)$$

From the initial condition we must have $A(s)=1$, so $X(t) \sim \text{Po}(\lambda t)$ (i.e., $X(t)$ follows a Poisson distribution with mean λt). In general

$$P_{ij}(s,s+t) = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}, \quad j \geq i. \quad (3.60)$$

The process we just derived is called the **Poisson¹ process**. Because the intensity of events in this process is constant, regardless of which state the process is in, it is sometimes called a **totally random process**. It has the property that the numbers of events in disjoint intervals of time are independent. To see that, first note from (3.60) that $P_{j,j+i}(s) = P_{0,i}(s)$ for all j . Write $X(t,t+s] = X(t+s) - X(t)$, so that $X(t,t+s]$ counts the number of births in $(t,t+s]$. Then

$$\begin{aligned} \mathbb{P}(X(t,t+s] = l \mid X(t) = j) &= \mathbb{P}(X(t+s) = j+l \mid X(t) = j) \\ &= P_{j,j+l}(s) = P_{0,l}(s) \end{aligned} \quad (3.61)$$

independent of j . Thus the number of events in $(0,t]$ is independent of the number of events in $(t,t+s]$. The argument can be extended to any disjoint time sets. We say that the Poisson process has **independent increments**. Hence it has no memory: the chance of something happening at any instant in time is independent of what has happened in the past. It has therefore found uses in describing events such as alpha-particle emissions from radioactive substances, large earthquakes, volcanic eruptions, or arrival of phone calls to a large exchange.

Example (Subsampling Poisson process events) Consider events occurring according to a Poisson process of rate λ , and recorded using a device which detects each event with probability p , independently from event to event.

¹ Poisson, Siméon Denis (1781–1840). French mathematical physicist. Generalized Bernoulli's law of large numbers.

Thinking of $Q_{i,i+1}$ as the instantaneous rate of events, it is clear that the recorded process Y_t will be a birth process with intensity $Q_{i,i+1}^Y = p\lambda$, i.e., a Poisson process with rate $p\lambda$. The case where p varies with time occurs sometimes when using historical records of earthquakes (Lee and Brillinger, 1979) or volcanic eruptions (Guttorp and Thompson, 1991). \square

We can develop a formal solution to the forward equation in matrix form. Since $d\mathbb{P}(t)/dt = \mathbb{P}(t)\mathbf{Q}$, an analogy with the one-dimensional case suggests that

$$\mathbb{P}(t) = e^{t\mathbf{Q}} \equiv \sum_0^\infty t^n \mathbf{Q}^n / n!. \quad (3.62)$$

Indeed, this can be shown to be the unique stochastic solution to the forward equation satisfying $\mathbb{P}(0) = \mathbb{I}$ and $\mathbb{P}(u+t) = \mathbb{P}(u)\mathbb{P}(t)$, the Chapman–Kolmogorov equations. This way of solving the equation suggests a way of developing properties of the model (see, e.g., Feller, 1971, ch. X and XIV). Using a spectral decomposition of \mathbf{Q} one can implement this solution in practice, but here we will concentrate on solving some special cases using the method of probability generating functions.

Define a **linear death process** by the intensities

$$Q_{j,j-1} = \mu j; \quad Q_{jj} = -\mu j. \quad (3.63)$$

We see that the intensity of a decrease by one (i.e., a death) is linear in current population size, while all other changes have intensity zero. The forward equation yields, starting from $X(0) = N$,

$$dP_{Nk}(t)/dt = -\mu k P_{Nk}(t) + \mu(k+1) P_{N,k+1}, \quad k < N. \quad (3.64)$$

Note that $P_{NN}(t) = e^{-\mu N t}$. Writing $G(s;t) = \sum s^k P_{Nk}(t)$ we derive the partial differential equation

$$\frac{\partial G(s;t)}{\partial t} = -\mu \sum k s^k P_{Nk}(t) + \mu \sum (k+1) s^k P_{N,k+1}(t) \quad (3.65)$$

or

$$\frac{\partial G(s;t)}{\partial t} = \mu(1-s) \frac{\partial G(s;t)}{\partial s} \quad (3.66)$$

which is a Lagrange equation (Appendix C contains solution methods for such equations) with solution

$$G(s;t) = (1-(1-s)e^{-\mu t})^N, \quad (3.67)$$

the probability generating function of a binomial random variable. Thus

$$P_{Nk}(t) = \binom{N}{k} e^{-k\mu t} (1-e^{-\mu t})^{N-k}. \quad (3.68)$$

Note that if $T \sim \text{Exp}(\mu)$ we have that $\mathbb{P}(T \geq t) = \exp(-\mu t)$. In other words, the death

process is in state k if $N-k$ independent exponential random variates have occurred at time t . This is like having N independent individuals, each dying at an exponentially distributed time. The time to extinction has the distribution of the maximum of N exponential random variates (see Figure 3.3).

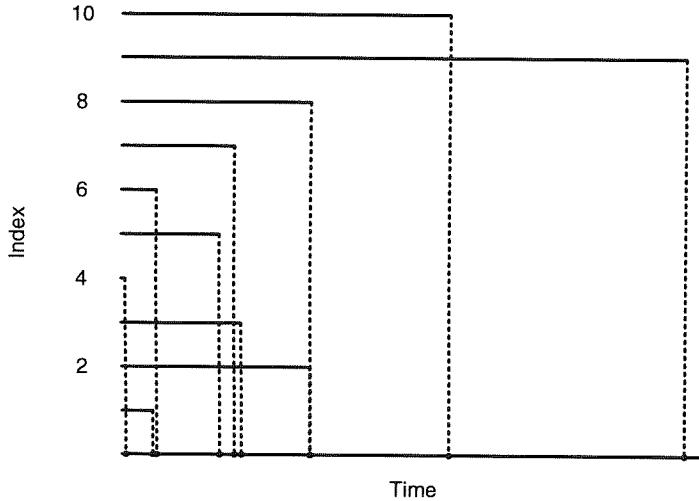


Figure 3.3. The relation between the exponential random variates and the death process.

For a more general death process, with death intensity μ_k , a similar result will hold: the next event at stage k will happen after an time which is exponentially distributed with parameter μ_k .

3.4. A general construction

The following result describes the structure of finite state space continuous time Markov chains.

Theorem 3.2 Let $(X(t), t \geq 0)$ be a finite state space, continuous time Markov chain.

(i) If $q_j=0$, and $X(t_0)=j$, then

$$\mathbf{P}(X(t)=j, t>t_0 \mid X(t_0)=j) = 1. \quad (3.69)$$

(ii) If $q_j>0$ and $X(t_0)=j$, then with probability one there is a jump discontinuity at some time $t>t_0$.

(iii) Let $0<\alpha<\infty$, $q_j>0$, and $X(t_0)=j$. Given that there is a discontinuity of $X(t)$

in $[t_0, t_0+\alpha]$, then the conditional probability that the first discontinuity is a jump to k is Q_{jk}/q_j .

(iv) $\mathbf{P}(X(u)=j, t \leq u \leq t+\alpha \mid X(t)=j) = \exp(-q_j\alpha)$, $\alpha>0$.

Proof (i) The backward equation is

$$\frac{dP_{jj}(t)}{dt} = \sum Q_{jr} P_{rj}(t) \equiv 0 \quad (3.70)$$

since $q_j=\sum_{r \neq j} Q_{jr}=0$. (ii) follows, if we can prove (iv), by letting $\alpha \rightarrow \infty$. Property (iii) was argued as (3.(Pn)). To verify (iv), let τ be the time spent in state j after time t until it makes a transition to a different state. By the Markov property, the probability that the process stays in j during the time interval $(u, u+v]$, given that it stays at least u time units, is precisely the unconditional probability that it stays in state j for v time units. In equation form

$$\mathbf{P}(\tau>u+v \mid \tau>u) = \mathbf{P}(\tau>v). \quad (3.71)$$

The result follows if we can prove that this equation only has the solution $\mathbf{P}(\tau>v) = \exp(-\alpha v)$. To see that $\alpha=q_j$, notice that by the backward equation $P_{jj}'(0)=-q_j$, and that $\mathbf{P}(\tau>t)=P_{jj}(t)$. It just remains to show that the only solution to the equation

$$g(u+v) = g(u)g(v) \quad (3.72)$$

is the exponential function. In order to verify this fact, we first compute $g(2/n)=g(1/n+1/n)=g(1/n)^2$. By repeating this calculation, we obtain $g(m/n)=g(1/n)^m$. Also $g(1)=g(1/n)^n$, so $g(m/n)=g(1)^{m/n}$. Since g is a distribution function, it is right continuous. Therefore (taking limits through the rational numbers) $g(x)=g(1)^x$. But $g(1)=g(1/2)^2 \geq 0$, so we can write $g(x)=\exp(-\alpha x)$ where $\alpha=-\log g(1)$. \square

Remark This result is true in some generality, although care is needed to deal with the possibility of explosion for infinite state spaces. A careful statement and proof is in Freedman (1983, section 5.6). \square

We can use Theorem 3.2 to construct the process $(X(u), 0 \leq u \leq t)$ with initial distribution \mathbf{p}^0 and intensities Q_{ij} . The procedure is illustrated in Figure 3.4.

- (1) Choose i_0 from \mathbf{p}^0 . Let $X(0)=i_0$.
- (2) If $q_{i_0}=0$, i.e., if i_0 is absorbing, we are done: $X(u)=i_0, 0 \leq u \leq t$.
- (3) If $q_{i_0}>0$, draw an $\exp(q_{i_0})$ -distributed random variable τ . If $\tau \geq t$ we are done: $X(u)=i_0, 0 \leq u \leq t$.

- (4) If $\tau < t$, choose i_1 according to the distribution $(R_{i_0 k})$, where $R_{jk} = Q_{jk}/q_j$.
 (5) If $q_{i_1} = 0$ the resulting path is $X(u) = i_0, 0 \leq u \leq \tau$ and $X(u) = i_1, \tau < u \leq t$.
 (6) If $q_{i_1} > 0$ go to (3).

The same algorithm applies in the case of infinite state space, as long as we stick to the minimal construction.

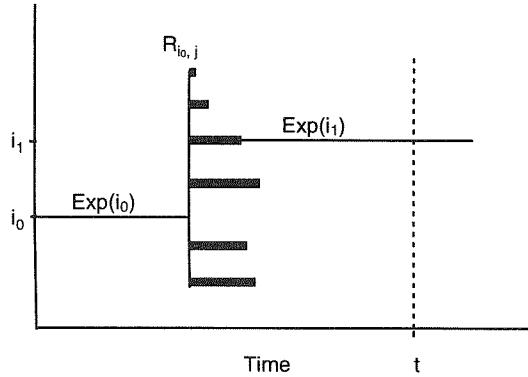


Figure 3.4. The construction using Theorem 3.2.

There are several important discrete time Markov chains associated with a given continuous time Markov chain. Let $Y_n = X(nh)$ for a fixed $h > 0$. This process is called the **discrete skeleton** of X (to scale h).

Proposition 3.4 The discrete skeleton (Y_n) is a Markov chain, with transition probabilities $P_{ij}(h)$.

Proof Using the Markov property of X we have

$$\begin{aligned} \mathbf{P}(Y_n = i \mid Y_{n-1} = j, Y_{n-2} = j_2, \dots, Y_0 = j_n) \\ = \mathbf{P}(X(nh) = i \mid X((n-1)h) = j) = P_{ij}(h). \end{aligned} \quad (3.73)$$

□

Define

$$R_{jk} = \begin{cases} (1 - \delta_{jk})Q_{jk}/q_j & q_j > 0 \\ \delta_{jk} & q_j = 0 \end{cases} \quad (3.74)$$

where $\delta_{jk} = 1(j=k)$. Then (R_{jk}) is a stochastic matrix, called the **jump matrix** of $(X(t))$. The jump matrix governs where the process goes when it leaves a state. Let τ_1, τ_2, \dots be the successive jump times of the process. Define the **jump**

chain $Z_k = X(\tau_k +)$. From the construction following Theorem 3.2 we see that this is a Markov chain with transition matrix R .

Example (The Yule–Furry process) The discrete skeleton to scale h of a Yule–Furry process has transition probabilities

$$P_{ij}(h) = \begin{bmatrix} j-1 \\ i-1 \end{bmatrix} p^i (1-p)^{j-i} \quad (3.75)$$

where $p = \exp(-\lambda h)$. The jump matrix has $R_{j,j+1} = 1$ and $R_{jk} = 0$ for $k \neq j+1$. This describes the process as a pure birth process, only moving upwards one step at the time. The jump chain is therefore very dull. □

The discrete skeleton $Y_n = X(nh)$ is useful in determining properties of the process $X(t)$. Since for Y_n we have $P_{jj}^{(n)} = P_{jj}(nh)$ the state j is persistent (in the discrete time sense) iff $\sum P_{jj}(nh) = \infty$. Looking at this sum as a Riemann sum, an equivalent condition is

$$\int P_{jj}(t) dt = \infty. \quad (3.76)$$

We say that a state j is persistent for a continuous time Markov chain if

$$\mathbf{P}^j(\sup\{t \geq 0 : X(t) = j\} = \infty) = 1. \quad (3.77)$$

Clearly this implies that $Y_n = j$ infinitely often, so j is persistent in the discrete time sense. If, on the other hand, j is persistent in the discrete time sense, we know that Y_n returns to j infinitely often, so $\mathbf{P}^j(\sup\{n : X(nh) = j\} = \infty) = 1$ for all h , and hence in the limit as $h \rightarrow 0$.

We define transience similarly: j is transient if

$$\mathbf{P}^j(\sup\{t \geq 0 : X(t) = j\} < \infty) = 1, \quad (3.78)$$

and again a state is transient in the continuous time sense iff it is transient in the discrete time sense for the skeleton chain.

There is a major difference between discrete and continuous time chains: there are no periodic states in continuous time. In fact, we have $P_{ii}(t) > 0$ for all $t \geq 0$. To see this, recall that (since we only consider standard semigroups) $P_{ii}(t) \rightarrow 1$ as $t \rightarrow 0$, so there is a $h > 0$ with $P_{ii}(s) > 0$ for all $s \leq h$. For any real t choose n large enough so that $t \leq nh$. By Chapman–Kolmogorov $P_{ii}(t) \geq (P_{ii}(t/n))^n > 0$, since $t/n < h$. We have shown the following result.

Theorem 3.3 (a) A state i is persistent (transient) for $X(t)$ iff it is persistent (transient) in the discrete skeleton.

(b) A state i is persistent iff $\int_0^\infty P_{ii}(t) dt = \infty$.

(c) $P_{ii}(t) > 0$ for all $t \geq 0$.

Remark For any two states i and j the **Lévy dichotomy** holds: either $P_{ij}(t) > 0$ for all $t > 0$, or $P_{ij}(t) = 0$ for all $t > 0$. A proof can be found in Bhattacharya and Waymire (1990, p. 304). As in the discrete time case we call a chain **irreducible** if $P_{ij}(t) > 0$ for all i, j . \square

We can use the discrete skeleton to assess stationarity as well. As in Chapter 2 we restrict attention to the irreducible case. Then if Y_n is non-null persistent we have $p_{ij}^{(n)} = P_{ij}(nh) \rightarrow \pi_j$; otherwise it goes to zero. Looking at two rational values h_1 and h_2 we see that the limit for each of these skeletons must be the same, and since the $P_{ij}(t)$ are continuous, we have a unique stationary distribution. Here is how you compute it: π_k is a solution to the equation

$$\sum \pi_j Q_{jk} = 0 \quad (3.79)$$

since, for any value of t , π solves

$$\sum_j \pi_j P_{jk}(t) = \pi_k \quad (3.80)$$

or, equivalently,

$$\pi_k(P_{kk}(t)-1) + \sum_{j \neq k} \pi_j P_{jk}(t) = 0. \quad (3.81)$$

Divide by t and let $t \rightarrow 0$ to get (3.79). The stationary distribution of the jump chain of $X(t)$ satisfies a different equation (Exercise 4).

Another way of obtaining (3.79) is to look at the forward equation $d\mathbb{P}(t)/dt = \mathbb{P}(t)Q$. We can think of the stationary distribution as one in which there is no probability flux, so $\mathbb{P}(t)$ is constant equal to π , and the time derivative is zero, whence $0 = \pi Q$. In the special case of reversible chains the law of detailed balance holds for the discrete skeleton. It translates into the requirement that

$$\pi_i Q_{ij} = \pi_j Q_{ji} \quad (3.82)$$

for all i and j , using the same argument as in deriving the equation for the stationary distribution.

Example (Birth and death process) Consider a process with intensities

$$\begin{cases} Q_{j,j+1} = \lambda_j \\ Q_{j,j-1} = \mu_j \\ Q_{jk} = 0 \text{ for } |j-k| \geq 2 \end{cases} \quad (3.83)$$

This is called a **birth and death process**. It was introduced by McKendrick (1925; in a special case in 1914), used by Feller (1939) to describe biological population growth, and studied in detail by Kendall (1948). In order to

determine the stationary distribution, the equation $\pi Q = 0$ yields

$$\begin{cases} -\lambda_0 \pi_0 + \mu_1 \pi_1 = 0 \\ \lambda_0 \pi_0 - (\mu_1 + \lambda_1) \pi_1 + \mu_2 \pi_2 = 0 \\ \dots \\ \lambda_{k-1} \pi_{k-1} - (\mu_k + \lambda_k) \pi_k + \mu_{k+1} \pi_{k+1} = 0 \end{cases} \quad (3.84)$$

with solution

$$\begin{cases} \pi_1 = \frac{\lambda_0}{\mu_1} \pi_0 \\ \pi_2 = \frac{\lambda_1}{\mu_2} \pi_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} \pi_0 \\ \dots \\ \pi_k = \frac{\lambda_{k-1} \dots \lambda_0}{\mu_k \dots \mu_1} \pi_0 \\ \dots \end{cases} \quad (3.85)$$

Since $\sum \pi_k = 1$ we obtain

$$\pi_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda_0 \dots \lambda_{k-1}}{\mu_1 \dots \mu_k} \right]^{-1} \quad (3.86)$$

and

$$\pi_k = \frac{\lambda_0 \dots \lambda_{k-1}}{\mu_1 \dots \mu_k \left[1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \dots \lambda_{n-1}}{\mu_1 \dots \mu_n} \right]}. \quad (3.87)$$

In order for these equations to be well defined, we must have

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \dots \lambda_{n-1}}{\mu_1 \dots \mu_n} < \infty. \quad (3.88)$$

This condition is also sufficient for the existence of a stationary distribution, which can easily be seen from the results for birth and death chains in the previous chapter upon noticing that the jump chain for a birth and death process is a birth and death chain. \square

Example (The linear BIDE process) While the birth and death process can be a reasonable description of many closed populations, a natural generalization allows for immigration and emigration as well. We restrict attention to the linear BIDE (birth, immigration, death, and emigration) process. As before, a birth or a death occurs with intensity proportional to the population size, while immigration occurs at constant rate, and emigration is similar to death. Thus the BIDE process is a nonlinear birth and death process with birth

intensity $\lambda_n = \lambda n + v$ and death intensity $\mu_n = (\mu + \eta)n$. Unless we have detailed information about which events are deaths and which are emigrations, we are unable to separate μ and η from data. By adding a constant to μ and subtracting it from η the death intensity remains the same. Thus we say that μ and η are not **identifiable** from the BIDE process $X(t)$. We may write $\mu_n = \kappa n$, where $\kappa = \mu + \eta$.

Applying the same type of argument as before we see that the pgf $G(s,t) = \sum P_{ij}(t)s^j$ satisfies the partial differential equation

$$\frac{\partial G(s,t)}{\partial t} = G(s,t)v(s-1) + \frac{\partial G(s,t)}{\partial s}(s-1)(\lambda s - \kappa). \quad (3.89)$$

Using the method of auxiliary equations the solution corresponding to the initial condition $X(0)=a$ is

$$G(s,t) = \frac{(\lambda - \kappa)^{v/\lambda} (\kappa(1-s)e^{(\lambda-\kappa)t} + \lambda s - \kappa)^a}{(\lambda(1-s)e^{(\lambda-\kappa)t} + \lambda s - \kappa)^{a+v/\lambda}}. \quad (3.90)$$

Letting $t \rightarrow \infty$, assuming that $\lambda < \kappa$, we get the pgf for the stationary distribution

$$G(s,\infty) = \left[\frac{\lambda - \kappa}{\lambda s - \kappa} \right]^{v/\lambda}, \quad (3.91)$$

which is the pgf of a negative binomial distribution with parameters v/λ and $(\kappa - \lambda)/\kappa$, the latter being the success probability. Thus the mean of the stationary distribution is $v/(\kappa - \lambda)$ and its variance is $v\kappa/(\kappa - \lambda)^2$.

If we know which events are of what type, a more detailed analysis is possible. Let $(B(t), I(t), D(t), E(t))$ be the number of births, immigrants, deaths, and emigrants, respectively, during the time interval $(0, t]$. These processes are called **counting processes**, and we shall consider them in more detail in Chapter 5. Note that $X(t) = X(0) + B(t) + I(t) - D(t) - E(t)$, and

$$\mathbb{P}(B(t+dt) - B(t) \mid X(t) = x) = x \lambda dt + o(dt), \quad (3.92)$$

so $B(t)$ is very similar to a linear birth process, except that its intensity does not only depend on the current state of itself, but also on the current state of other processes. Clearly, the four-dimensional process is a Markov process with discrete state space. As we shall see later, μ and η are both identifiable if we observe the four-dimensional process. \square

3.5. Queueing systems

Waiting in line is a common way of spending (some would say wasting) one's time in modern society. Queues occur in grocery stores, in multi-user computers, in satellites providing intercontinental telephone service, in hospitals, and in countless other activities. Queueing theory is one of the largest subfields of applied probability. Asmussen (1987) is a good source for a theoretical development. Here we shall concentrate on some very simple queueing models.

A simple queue consists of an **arrival process** of customers, a set of **service facilities** dealing with the customers, and a **queue discipline**, determining the order in which customers are served. Following Kendall (1953) a queueing system is written $A/B/k$, where A describes the interarrival time distribution (most queues are assumed to have independent interarrival times), B the service time distribution (again, service from different servers and for different customers are often assumed independent), and k the number of servers. There are many possible choices of A and B ; we restrict attention to the exponential distribution, denoted M (for Markovian), a deterministic distribution denoted D , and a general (unspecified) distribution, written G . Thus $M/M/1$ is a single server queue with exponential interarrival times and service times, while $M/G/\infty$ is a queue with exponential interarrival times, general service time, and no waiting (each arriving customer is immediately served).

The $M/M/m$ processes are simply birth and death processes. These are the only Markovian queues. If $m=1$ we have $\lambda_n = \lambda$ and $\mu_n = \mu 1(n>0)$. The stationary distribution is easily seen to be geometric with parameter $\rho = \lambda/\mu$, provided that $\lambda < \mu$ (Exercise 6). The parameter ρ is called the **traffic intensity** and measures the ratio between input rate and output rate at full capacity. From the equilibrium distribution we see that

$$\mathbb{P}(\text{server idle}) = \pi_0 = 1 - \rho \quad (3.93)$$

and the mean queue length in equilibrium is

$$\mathbb{E}^\pi X(t) = \frac{\rho}{1 - \rho}. \quad (3.94)$$

As $\rho \uparrow 1$ the server is busy with high probability, and the mean queue length is large, as would be expected. Let W be the waiting time before service in equilibrium. Since each individual in the queue, at the arrival (which we take to be time 0) of the customer whose waiting time we are determining, will take an exponentially distributed time τ_i to be served, we have

$$(W \mid X(0) = i) = \tau_1 + \cdots + \tau_i \sim \Gamma(i, \mu) \quad (3.95)$$

with conditional Laplace transform $(1+t/\mu)^{-i}$. Hence

$$\mathbb{E}(e^{-W\theta}) = \sum (1-\rho) \rho^i (1+t/\mu)^{-i} = \frac{1-\rho}{1-\rho\mu/(1+t/\mu)} \quad (3.96)$$

which is the Laplace transform of a random variable which is 0 with probability $1-\rho$ and $\text{Exp}(\mu(1-\rho))$ with probability ρ .

Next consider the $M/M/\infty$ queue. Here $\lambda_n = \lambda$, $\mu_n = n\mu$, so we have an immigration-death process, whence the process is ergodic with equilibrium distribution $\text{Po}(\eta)$, where $\eta = \lambda/\mu$. The traffic intensity for this process is zero, since nobody ever needs to wait to be served.

Finally we look at an $M/M/m$ queue, with $1 < m < \infty$, so $\lambda_n = \lambda$ and $\mu_n = \min(m, n)\mu$. The traffic intensity is $\rho = \lambda/m\mu$. Let $\eta = \lambda\mu$. From equation (3.88) the ergodicity condition is the finiteness of

$$S = \sum_{n=0}^{m-1} \frac{\eta^n}{n!} + \frac{\eta^m}{m!} \sum_{n=0}^{\infty} \rho^n \quad (3.97)$$

or, equivalently, $\rho < 1$. The equilibrium distribution is

$$\pi_n = \begin{cases} \frac{1}{S} \frac{\eta^n}{n!} & 0 \leq n \leq m \\ \frac{1}{S} \frac{\eta^m}{m!} \rho^{n-m} & n > m. \end{cases} \quad (3.98)$$

Thus π is a combination of the $M/M/\infty$ solution on $\{0, \dots, m\}$, where there is no waiting, and the $M/M/1$ solution on $\{m+1, m=2, \dots\}$ where some customers wait. We see that

$$P^\pi(\text{all servers busy}) = \pi_m + \pi_{m+1} + \dots = \frac{1}{S} \frac{\eta^m}{m!} \frac{1}{1-\rho} \quad (3.99)$$

and

$$E^\pi X(t) = \sum k \pi_k = \frac{1}{S} \left(\sum_{n=1}^{m-1} \frac{\eta^n}{(n-1)!} + \frac{\eta^m}{m!} \left[\frac{\rho}{(1-\rho)^2} + \frac{m}{1-\rho} \right] \right). \quad (3.100)$$

As we saw in Chapter 2, every ergodic birth and death chain, and consequently every ergodic birth and death process, is time reversible. Suppose we have a stationary version of a Markovian queue $X(t), t \in \mathbb{R}$. Let $\tilde{X}(t) = X(-t-0) \equiv \lim_{s \uparrow -t} X(s)$. A departure from X at time t corresponds to an arrival to \tilde{X} at time $-t$. If $\lambda_n = \lambda$, so we have Poisson arrivals, then the instants $-t_i$ of arrivals to \tilde{X} , corresponding to departures from X at times t_i , form a Poisson process by reversibility. This Poisson process has the same intensity λ , and the departures up to time t are independent of $\tilde{X}(-t)$, which, with probability one, is equal to $X(t)$ since X has probability 0 of jumping at time t . We have shown the following result.

Proposition 3.5 The departure times of an ergodic $M/M/\bullet$ queue form a Poisson process with the same rate as the arrival process. The departure process at time t is independent of past departures up to time t . \square

Remark This result is astonishing. The output process does not depend on the service rate or the number of servers, just on the input rate. Of course, in equilibrium what comes in must go out, but intuition still balks at the result at first!

Consider a series of K Markovian queues, where arrivals to the first are Poisson of rate λ , and the servers in each queue act independently of the servers in the other queue, with service rate $\mu_n^{(k)}$ in the k th queue. If the input queue is ergodic, and we assume that the output from the k th queue is the input to the $(k+1)$ th queue, then the whole system is ergodic, with independent queue lengths each governed by its respective stationary distribution. This is immediate from repeated use of Proposition 3.5.

Application (Communications center) Hatzikostandis and Howe (1967) provide some data on manual message-handling operators in a military communications terminal. We restrict attention to messages of type A, which arrive at a receipt center (staffed by trained personnel) where they are logged and immediately sent to the processing center (staffed by relatively untrained personnel). After input to the processing center routing is determined by message content and multiple copies are made. The messages are then logged out and placed in a pigeon-hole box until messengers pick them up for distribution to the users.

The data in Table 3.1 are on priority 1 messages, which are immediately dealt with.

Table 3.1 Quarter-hourly arrival counts

No. of messages	0	1	2	3
Receipt center	39	33	19	5
Processing center	41	31	17	7

They consist of 24 hours' worth of 15-minute counts of arrivals to the receipt center and the processing center. If we can consider the transport time between the centers approximately constant, then the arrivals at the processing center have the same distribution as the departures from the receipt center. If the receipt center queue can be modeled as $M/M/\bullet$, the arrivals and departures should both be Poisson with mean $\lambda/4$, where λ is the hourly intensity of priority 1 message arrivals. Both the input and the output intensities are estimated to be $4 \times 86/96 = 3.6$ messages per hour. The variance of the input process is 0.8, and that of the output process is 0.9, both essentially the same as the mean 0.9, indicating that a Poisson distribution is a reasonable description. \square

While queues other than $M/M/\bullet$ are non-Markovian, it is sometimes the case that one can find a closely related process which is Markovian, thereby simplifying much of the analysis. We give an example below, and leave another as Exercise 7.

Example (The M/G/1 queue and its imbedded branching process) Consider an M/G/1 queue, starting with one individual arriving at time 0 and being served immediately. During the service time of this individual, a number of new customers arrive. We consider these customers the offspring of the original individual. Doing this for each customer, we obtain, by looking at the process only at the instances of departure, a Bienaymé-Galton-Watson process with offspring distribution

$$p_j = \int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} dB(t) \quad (3.101)$$

where B is the cdf of the service time. If the service time has mean μ , the offspring mean is $\lambda\mu$, and we have a subcritical process, provided that $\rho = \lambda\mu < 1$. Once the queue becomes empty, we get an independent realization of the process starting from the next arrival.

Letting Y_n denote the queue length just after the departure of customer $n-1$, and Z_n the number of offspring of the n th individual (i.e., the number of customers arriving during the n th service time), we can write

$$Y_{n+1} = Y_n - 1 + Z_n + 1(Y_n=0). \quad (3.102)$$

Hence Y_n has transition matrix

$$\mathbb{P} = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \\ 0 & 0 & 0 & p_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \quad (3.103)$$

which is irreducible, since all the p_i are positive. The stationary distribution satisfies the system of equations

$$\begin{cases} \pi_0 = \pi_0 p_0 + \pi_1 p_0 \\ \pi_1 = \pi_0 p_1 + \pi_1 p_1 + \pi_2 p_0 \\ \pi_2 = \pi_0 p_2 + \pi_1 p_2 + \pi_3 p_1 + \pi_4 p_0 \\ \dots \end{cases} \quad (3.104)$$

Writing $q_r = p_{r+1} + p_{r+2} + \dots$, adding the first n equations and solving for $\pi_{n+1} p_0$ we obtain

$$\begin{cases} \pi_1 p_0 = \pi_0 q_0 \\ \pi_2 p_0 = \pi_0 q_1 + \pi_1 q_1 \\ \pi_3 p_0 = \pi_0 q_2 + \pi_1 q_2 + \pi_2 q_1 \\ \dots \end{cases} \quad (3.105)$$

Equilibrium analysis of (3.102) yields that the equilibrium mean $\mathbf{E}Y$ satisfies

$$\mathbf{E}Y = \mathbf{E}Y - 1 + \mathbf{E}Z + \mathbf{P}(Y=0) \quad (3.106)$$

so, since $\mathbf{E}Z = \rho$, we have $\mathbf{P}(Y=0) = \pi_0 = 1 - \rho$, from which the system (3.105) can be solved recursively. \square

3.6. An improved model for cosmic radiation

At the end of Section 3.1 we were discussing shortcomings of the Furry model of cosmic radiation showers. Following Arley (1943), we will try to adapt the model to the physics of the situation. The first problem is that particles should disappear as their energy becomes depleted. The pure birth model therefore needs to be replaced by a birth and death model. Roughly speaking, the rate at which an individual dies should be proportional to the age of the family. This yields a time-dependent birth and death process with birth rate $\lambda_n(t) = \lambda n$ and death rate $\mu_n(t) = \mu n t$. The model is still overly simplified, since the probability of being absorbed (as well as the probability of giving birth to new particles) really depends on the energy of the particular particle at that time. This energy, which depends on the antecedents of each particle, will fluctuate around the average energy, while this model sets it equal to the average energy at that time. However, we will see that it nonetheless adds some amount of realism to the description.

Let us first compute an equation for the mean value $m(t)$ of the process. This mean value is $\sum k P_{1k}(t)$. From the forward equation we see that

$$\frac{dP_{1k}(t)}{dt} = \lambda(k-1)P_{1,k-1}(t) - (\lambda + \mu t)kP_{1k}(t) + \mu t(k+1)P_{1,k+1}(t) \quad (3.107)$$

from which we derive the following equation for the mean:

$$m'(t) = (\lambda - \mu t)m(t) \quad (3.108)$$

which has solution $(m(0)=1)$

$$m(t) = \exp(\lambda t - \frac{\mu}{2}t^2). \quad (3.109)$$

As shown in Figure 3.5, this curve has a maximum of $\exp(\lambda^2/2\mu)$ at $t = \lambda/\mu$. The probability generating function of $X(t)$ satisfies the partial differential equation

$$\frac{\partial G(s;t)}{\partial t} = (\lambda s^2 - (\lambda + \mu t)s + \mu t) \frac{\partial G(s;t)}{\partial s}. \quad (3.110)$$

The solution satisfying $G(s;0)=s$ is

$$G(s;t) = (-A)^{-1}(t - A(s)) \quad (3.111)$$

where

$$A(s) = \exp(-\mu(s-s^2/2)) \int_0^s \lambda u(1-u)\exp(\mu(u-u^2/2))du \quad (3.112)$$

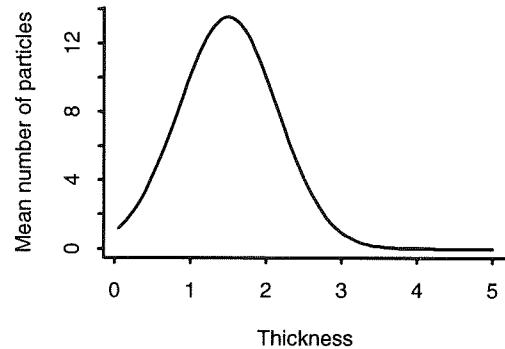


Figure 3.5. The mean curve for the modified model of cosmic radiation.

which can be expressed in terms of incomplete gamma functions.

The final step towards a realistic approximate model for cosmic radiation showers is to take into account the different kinds of particles (electrons and photons) that occur in alternating generations. We will assume that the process is symmetric with respect to the two kinds of particles, apart from the initial condition. In other words, we will assume that the birth and death process described above adequately describes both the production of electrons from photons and the production of photons from electrons. More precisely, we assume that particles have an intensity rate of λ to be absorbed and give birth to two particles of the other type, and that all particles have an intensity rate μt of being absorbed without giving birth to a new particle. Arley (1943, section 5.2) discusses the reasonableness of these assumptions. Let $(X(t), Y(t))$ be the two-dimensional process of electrons and photons, with $X(0)=1$ and $Y(0)=0$. The equations for mean values, letting $n(t)=\mathbb{E}X(t)$ and $m(t)=\mathbb{E}Y(t)$, are

$$n(t+dt) = n(t) + m(t)2\lambda dt - n(t)\lambda dt - n(t)\mu dt + o(dt) \quad (3.113)$$

and

$$m(t+dt) = m(t) + n(t)2\lambda dt - m(t)\lambda dt - m(t)\mu dt + o(dt). \quad (3.114)$$

Turning these equations into differential equations we see that

$$n'(t) = -(\lambda+\mu t)n(t) + 2\lambda m(t) \quad (3.115)$$

and

$$m'(t) = -(\lambda+\mu t)m(t) + 2\lambda n(t). \quad (3.116)$$

Solving the first equation for m and substituting that in the second equation yields

$$n''(t) + 2(\lambda+\mu t)n'(t) + ((\lambda+\mu t)^2 + \mu - 4\lambda^2)n(t) = 0. \quad (3.117)$$

The solutions are

$$n(t) = \frac{1}{2}\exp(\lambda t - \mu t^2/2) + \frac{1}{2}\exp(-3\lambda t - \mu t^2/2) \quad (3.118)$$

and

$$m(t) = \frac{1}{2}\exp(\lambda t - \mu t^2/2) - \frac{1}{2}\exp(-3\lambda t - \mu t^2/2). \quad (3.119)$$

When λt is large we have that $n(t) \approx m(t)$, each approximately being equal to half the mean value in the one-dimensional model. The effects of the initial conditions (which are the only asymmetries in the model) disappear exponentially fast.

Arley approximates the marginal time-dependent probability distributions by Pólya distributions, and obtains very good fits to experimental data. However, the tendency in physics has later been to attempt to exploit the branching structure of electron showers. An interesting description of such models is given in Harris (1963, Chapter VII).

3.7. Statistical inference for continuous time Markov chains

We restrict attention to uniform semigroups, so that, in particular, $\sum_j Q_{ij}=0$. In order to compute the likelihood from complete observation of a continuous time Markov chain $(X(t), t \leq T)$ we first note that all we need to know are the number of jumps $N(T)$, the successive jump times τ_i , and the jump chain. In other words, we need to know

$$V_T = \{N(T), X(0), \tau_1, X(\tau_1+), \tau_2, \dots, X(\tau_{N(T)}+)\}. \quad (3.120)$$

In statistical terminology, V_T is a **sufficient statistic**. The likelihood of Q at $v=(n, x_0, t_1, \dots, x_n)$ is

$$\begin{aligned} p^0(x_0)q(x_0)\exp(-q(x_0)t_1)\prod_{j=1}^{n-1} \frac{Q(x_{j-1}, x_j)}{q(x_{j-1})} q(x_j)\exp(-q(x_j)(t_{j+1}-t_j)) \\ \times \frac{Q(x_{n-1}, x_n)}{q(x_{n-1})}\exp(-q(x_n)(T-t_n)) \text{ if } n \geq 1 \end{aligned} \quad (3.121)$$

$$p^0(x_0)\exp(-q(x_0)T) \text{ if } n=0$$

Here q_k is written $q(k)$ and Q_{ij} is $Q(i,j)$ for typographic reasons. The derivation of (3.121) is immediate from Theorem 3.3: there are independent exponential holding times, and jumps according to the jump chain matrix. We can write the likelihood, given that $X(0)=x$, as

$$L(Q; V_T) = \left[\prod_{i=1}^{K_T} \prod_{j \neq i} Q(i,j)^{N_T(i,j)} \right] \exp(-\sum_i A_T(i)q_i) \quad (3.122)$$

where the number K_T of observed states up to time T is finite, since (with probability one) only a finite number of states are observed in finite time, $N_T(i,j)$ is the number of transitions from i to j up to time t (so $N_T(i,i)=0$), and $A_T(i)$ is the

total time spent in state i up to time t . It follows that $(A_T(i), N_T(i,j), i, j \in D)$ is a sufficient statistic (it is, in fact, minimal sufficient).

Theorem 3.4 Suppose that X is ergodic at the true parameter Q_0 , and that the set $D = \{(i,j) : i \neq j, Q_0(i,j) > 0\}$ is known. Then with probability one

$$\hat{Q}(i,j;T) = \frac{N_T(i,j)}{A_T(i)} \quad (3.123)$$

are unique solutions to the likelihood equations, and for T large enough they provide a maximum with probability one.

Proof Write the logarithm of the likelihood

$$\begin{aligned} \log L(Q) &= \sum_i 1(X(0)=i) \log p^0(i) - \sum_i A_T(i) q_i + \sum_{i \neq j} N_T(i,j) \log Q_{ij} \\ &= c - \sum_{i \neq j} A_T(i) Q_{ij} + \sum_{i,j} N_T(i,j) \log Q_{ij} \end{aligned} \quad (3.124)$$

where we used the fact that $\sum_j Q_{ij} = 0$. Taking derivatives with respect to Q_{ij} and setting them equal to zero yields the likelihood estimates

$$Q_{ij} = \frac{N_T(i,j)}{A_T(i)}. \quad (3.125)$$

The second derivative has diagonal elements $N_T(i,j)/Q_{ij}^2$, and off-diagonal elements 0, so the solution is a maximum whenever $N_T(i,j) > 0$, which will eventually happen with probability one since X is ergodic. \square

Corollary Let $N_T(i) = \sum_j N_T(i,j)$ be the number of transitions out of i in $(0, T)$. Then $\hat{q}_i = N_T(i)/A_T(i)$.

Proof By invariance of the mle under reparametrization we have that

$$\hat{q}_i = \sum_{j \neq i} \hat{Q}_{ij} = \sum_{j \neq i} \frac{N_T(i,j)}{A_T(i)} = \frac{N_T(i)}{A_T(i)}. \quad (3.126)$$

\square

It is fairly straightforward to derive first and second order properties (means, variances, and covariances) of the components of the sufficient statistic.

Proposition 3.6 Let $p_i(t) = \mathbb{P}(X(t)=i)$. Then for $i \neq j$,

$$\mathbb{E}N_t(i,j) = Q_{ij} \int_0^t p_i(u) du. \quad (3.127)$$

$$\begin{aligned} \mathbb{E}N_t(i,j)N_t(k,m) &= 1((i,j)=(k,m))Q_{ij} \int_0^t p_i(u) du \\ &\quad + Q_{ij}Q_{km} \int_0^t \int_0^v (P_{mi}(v-u)p_k(u) + P_{jk}(v-u)p_i(u)) du dv. \end{aligned} \quad (3.128)$$

Proof Divide $[0, t]$ into intervals $[(k-1)h, kh)$ for $k = 1, \dots, m$, and let $n_k(i,j)$ count the number of transitions $i \rightarrow j$ in the k th interval. Then $N_t(i,j) = \sum n_k(i,j)$. For h small enough, each $n_k(i,j)$ is 0 or 1 with probability $1-o(h)$. For $r \neq j$, $l \neq i$,

$$\begin{aligned} \mathbb{P}(n_k(i,j)=1, X(kh)=r, X((k-1)h)=l) \\ &= (Q_{li}h + o(h))(Q_{ij}h + o(h))(Q_{jr}h + o(h)) + o(h) \\ &= O(h^3) = o(h). \end{aligned} \quad (3.129)$$

Thus

$$\begin{aligned} \mathbb{P}(n_k(i,j)=1) &= \sum_{l,r} \mathbb{P}(n_k(i,j)=1, X((k-1)h)=l, X(kh)=r) \\ &= \mathbb{P}(n_k(i,j)=1, X((k-1)h)=i, X(kh)=j)) + o(h) \\ &= p_i((k-1)h)Q_{ij}h + o(h) \end{aligned} \quad (3.130)$$

and

$$\begin{aligned} \mathbb{E}N_t(i,j) &= \sum_k \mathbb{P}(n_k(i,j)=1) = Q_{ij} \sum_1^m p_i((k-1)h)h + o(h) \\ &\rightarrow Q_{ij} \int_0^t p_i(u) du \end{aligned} \quad (3.131)$$

as $h \rightarrow 0$. The second equality is similar, only having more terms to keep track of. \square

The following result deals with the times spent in different states.

Proposition 3.7

$$\mathbb{E}A_t(i) = \int_0^t p_i(u) du. \quad (3.132)$$

$$\mathbb{E}A_t(i)A_t(j) = \int_0^t \int_0^u (P_{ji}(u-v)p_j(v) + P_{ij}(u-v)p_i(v)) dv du. \quad (3.133)$$

Proof Use $A_t(i) = \int_0^t 1(X(u)=i)du$, so, upon taking expectations, we get $\mathbb{E}A_t(i) = \int_0^t \mathbf{P}(X(u)=i)du$. The second part is similar. \square

The joint behavior between transitions and time spent in each state is as follows.

Proposition 3.8

$$\mathbb{E}N_t(i,j)A_t(r) = Q_{ij} \int_0^t \int_0^u (P_{ri}(u-v)p_r(v) + P_{jr}(u-v)p_i(v))dv du. \quad (3.134)$$

$$\mathbb{E}(N_t(i,j)-Q_{ij}A_t(i))(N_t(r,s)-Q_{rs}A_t(r)) = \delta(i,j; r, s) \int_0^t p_i(u)du. \quad (3.135)$$

Proof Use arguments similar to those of the previous two proposition. \square

As in the discrete time case, an ergodic theorem holds. Let T_i be the time of the first return to i .

Theorem 3.5 Let $X(t)$ be an ergodic chain. Assume that f satisfies

$$\mathbb{E}^i \left| \int_0^{T_i} f(X(s))ds \right| < \infty. \quad (3.136)$$

Then

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X(s))ds = \frac{\mathbb{E}^i \int_0^{T_i} f(X(s))ds}{\mathbb{E}^i T_i} \text{ with probability one.} \quad (3.137)$$

\square

A proof is in Bhattacharya and Waymire (1990, p. 306). Now write $N_t(i,j) = \int_0^t f_{ij}(X(s-), X(s))ds$ where $f_{ij}(u,v) = 1(u=i, v=j, i \neq j)$. We deduce (with some handwaving) the following result.

Proposition 3.9 If X is ergodic then as $t \rightarrow \infty$

$$\mathbb{E}N_t(i,j)/t \rightarrow \pi_i Q_{ij}, \quad (3.138)$$

$$\mathbb{E}N_t(i,j)N_t(r,s)/t^2 \rightarrow \pi_i \pi_r Q_{ij} Q_{rs}, \quad (3.139)$$

$$\mathbb{E}A_t(i)/t \rightarrow \pi_i, \quad (3.140)$$

$$\mathbb{E}A_t(i)A_t(j)/t^2 \rightarrow \pi_i \pi_j, \quad (3.141)$$

$$\mathbb{E}N_t(i,j)A_t(r)/t^2 \rightarrow \pi_i \pi_r Q_{ij} \quad (3.142)$$

and

$$\mathbb{E}(N_t(i,j)-Q_{ij}A_t(i))(N_t(r,s)-Q_{rs}A_t(r))/t \rightarrow \delta(i,j; rs)\pi_i. \quad (3.143)$$

\square

It follows, using Chebyshev's inequality, that

$$N_t(i,j)/t \xrightarrow{P} \pi_i Q_{ij} \quad (3.144)$$

and

$$A_t(i)/t \xrightarrow{P} \pi_i. \quad (3.145)$$

It is possible to show (Adke and Manjunath, 1984, Theorem 5.4.2) that if

$$\xi_{ij}(t) = \frac{N_t(i,j)-A_t(i)Q_{ij}}{t^{\frac{1}{2}}}, \quad i \neq j \quad (3.146)$$

then the ξ_{ij} are asymptotically independent normal variates with mean 0 and variance $\pi_i Q_{ij}$. Hence the following result, summarizing the asymptotic behavior of the nonparametrics mle's.

Theorem 3.6 (i) $\hat{Q}(i,j; T) \xrightarrow{P} Q_{ij}$ for $i, j \in D$.

(ii) The $t^{\frac{1}{2}}(\hat{Q}(i,j; T) - Q_{ij})$, $i \in D$, converge in distribution to independent, mean zero, normal variates with variances Q_{ij}/π_i , respectively. \square

Corollary $(A_T(i)\hat{Q}(i,j; T))^{\frac{1}{2}}(\hat{Q}(i,j; T) - Q_{ij}) \xrightarrow{d} N(0, 1)$

Proof Use Slutsky's theorem, Theorem 3.6, and (3.145). \square

Application (Zoology) S. A. Altman studied the vital events in a baboon troupe (Cohen, 1969) in the Amboseli nature reserve in Kenya. The data are given in Table 3.2. Ignoring the information about which events are births, deaths, immigrations or emigrations, we compute the nonparametric mle. Since all changes are by one unit, the resulting \hat{Q} is tri-diagonal. Figure 3.6 shows the estimated intensities for population increases and decreases, plotted against population size. Popular parametric models for this type of data (birth and death or BIDE processes) are linear in population size. Thus, we may want to fit a line to the points in the figure. Note, however, that the estimators have different

Table 3.2 Vital events in a baboon troupe

After this many days	at this troop size	this event occurred.
41	40	B
5	41	B
22	42	B
2	43	D
17	42	D
26	41	I
0	42	I
55	43	B
35	44	I
20	45	E
5	44	D
6	43	E
32	42	D
4	41	D
0	40	D
22	39	D
10	38	B
0	39	B
7	40	D
4	39	B
17	40	D
11	39	E
3	38	B
4	39	D
8	38	D
2	37	D
5	36	B
10	37	B

variances. Using the corollary to Theorem 3.6 the asymptotic variances are estimated by $\hat{Q}(i,j;T)/A_T(i)=N_T(i,j)/A_T(i)^2$. A weighted least squares line with weights inversely proportional to the estimated variances is shown in the figure. The slope is estimated to be -0.004 , and the intercept to be 0.193 . Neither of these estimates is significantly different from zero. Of course, a linear model with negative slope is not acceptable, since eventually the estimated intensity may become negative. Fitting a line through the origin, the estimated slope is 5.9×10^{-4} with a standard error of 1.5×10^{-4} .

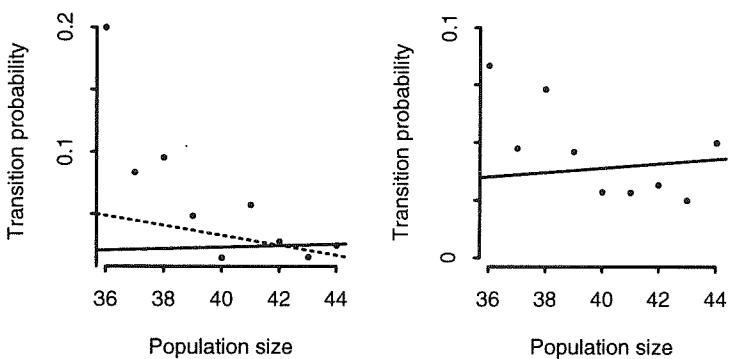


Figure 3.6. Estimated intensity of increase (left) and decrease (right) versus population size.

The estimated population decrease intensities are shown in Figure 3.6. The most common model for such data would be linear with zero intercept. The appropriate weighted least squares line has slope 9.7×10^{-4} , with standard error 2×10^{-4} . The quality of this fit is partly assessed by the (weighted) correlation coefficient of 0.91. These results indicate that a linear birth and death model may be sensible. \square

In many applications we are interested in parametric models for Q , and would only use the nonparametric estimators for preliminary data-analytic purposes, such as assessment of linearity. As in the discrete time case one must assume some regularity conditions for Q . Again, we consider only ergodic chains. Assume that the $k \times k$ dimensional parameter θ takes values in an open subset of \mathbb{R}^k .

Assumptions B

B1: The set $\{(i,j): Q_{ij}(\theta) > 0\}$ does not depend on θ .

B2: The functions $Q_{ij}(\theta)$ are three times continuously differentiable.

B3: The $k \times k$ matrix $\Sigma(\theta)$ with elements

$$\Sigma_{lm}(\theta) = \sum_{i,j} \frac{\pi_i(\theta)}{Q_{ij}(\theta)} \frac{\partial Q_{ij}(\theta)}{\partial \theta_l} \frac{\partial Q_{ij}(\theta)}{\partial \theta_m} \quad (3.147)$$

is positive definite for all θ , where $\pi(\theta)$ is the stationary distribution.

Theorem 3.7 Under Assumptions B the likelihood equations $(\partial/\partial \theta_i) \log L(\theta) = 0$ have, for sufficiently large t , a solution $\hat{\theta}(t)$ with probability one. This solution is consistent and maximizes the likelihood. If, in addition, $Q(\theta)$ has all eigenvalues distinct for all θ , then

$$\left[-D^2 \log L(\hat{\theta}) \right]^{\frac{1}{2}} (\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, \mathbb{I}) \quad (3.148)$$

where D^2 denotes the matrix of second order partial derivatives.

This result is a simple version of Theorem 7.3 in Billingsley (1961), where a proof can be found.

Application (Zoology, continued) Returning to the Altman baboon troupe data, and still neglecting the information about the character of each event, we may model the population sizes using a BIDE model (the birth and death model indicated earlier is a special case). The likelihood is

$$L(\lambda, v, \kappa) \propto \left[\prod_{i=n_{\min}(T)}^{n_{\max}(T)} (\lambda i + v)^{n_i(T)} \right] \kappa^{D(T)+E(T)} \exp(-vT - (\lambda + \kappa)S(T)) \quad (3.149)$$

where $n_{\min}(T)$ and $n_{\max}(T)$ are the smallest and largest observed population sizes for which a population increase has been observed in the observation time $(0, T)$, $n_i(T) = N_T(i, i+1)$, $S(T) = \int_0^T X(s) ds$, and all terms that do not depend on unknown parameters have been grouped into the constant of proportionality. Numerically we have $T=373$, $n_{\min}(T)=36$, $n_{\max}(T)=44$, $\mathbf{n}(T)=(n_{36}, \dots, n_{44}) = (1, 1, 2, 2, 1, 2, 2, 1, 1)$, $S(T)=15407$, and $D(T)+E(T)=15$. Maximizing (3.149) yields $(\hat{\lambda}, \hat{v}, \hat{\kappa}) = (1.65 \times 10^{-9}, 3.49 \times 10^{-2}, 9.74 \times 10^{-4})$. Note that the terms involving κ can be separated from those containing (λ, v) whence asymptotically $\hat{\kappa}$ is independent of $(\hat{\lambda}, \hat{v})$. Differentiating $\log L(\lambda, v, \kappa)$ twice yields

$$\begin{aligned} \frac{\partial^2}{\partial \lambda^2} \log L(\lambda, v, \kappa) &= - \sum_{i=n_{\min}(T)}^{n_{\max}(T)} \frac{i^2 n_i(T)}{(\lambda i + v)^2} \\ \frac{\partial^2}{\partial \lambda \partial v} \log L(\lambda, v, \kappa) &= - \sum_{i=n_{\min}(T)}^{n_{\max}(T)} \frac{i n_i(T)}{(\lambda i + v)^2} \\ \frac{\partial^2}{\partial v^2} \log L(\lambda, v, \kappa) &= - \sum_{i=n_{\min}(T)}^{n_{\max}(T)} \frac{n_i(T)}{(\lambda i + v)^2} \\ \frac{\partial^2}{\partial \lambda^2} \log L(\lambda, v, \kappa) &= - \frac{D(T)+E(T)}{\kappa^2}. \end{aligned} \quad (3.150)$$

We obtain the asymptotic covariance matrix of the mle by plugging in the mle's into (3.150) and invert the negative of the resulting matrix (with zeros at the cross-derivatives between κ and λ or v). Here $se(\hat{\lambda})=2.5 \times 10^{-4}$, $se(\hat{v})=1.9 \times 10^{-3}$, $corr(\hat{\lambda}, \hat{v})=0.19$, and $se(\hat{\kappa})=2.5 \times 10^{-4}$.

When we ignore the information about the character of each event, the process behaves like an immigration-death (ID) process. In fact, the likelihood ratio statistic for testing the null hypothesis $\lambda=0$ is zero to six decimal places. As seen in Exercise 5, the stationary distribution for an ID process is geometric with mean v/κ , which we estimated to be 35.8. Note that the entire path of this process lies above its stationary mean, with an observed mean of 41.3. Using a Taylor expansion (Exercise 8), we estimate the variance of $\hat{v}/\hat{\kappa}$ to be $(se(\hat{v})/\hat{\kappa})^2 + (se(\hat{\kappa})\hat{v}/\hat{\kappa})^2$, so the standard error of the stationary mean is 9.5. The observed mean is therefore well within range of the usual.

Since in this case we have information about the character of each event, we can use this to write the likelihood

$$L(\lambda, v, \mu, \eta) \propto \lambda^{B(T)} v^{I(T)} \mu^{D(T)} \eta^{E(T)} \exp(-vT - (\lambda + \mu + \eta)S(T)). \quad (3.151)$$

Here $B(T)=10$, $I(T)=E(T)=3$, and $D(T)=12$. The mle's are $\hat{\lambda}=B(T)/S(T)$, $\hat{v}=I(T)/T$, $\hat{\mu}=D(T)/S(T)$, and $\hat{\eta}=E(T)/S(T)$. Numerically these are $\hat{\lambda}=6.5 \times 10^{-4}$, $\hat{v}=8.0 \times 10^{-3}$, $\hat{\mu}=7.8 \times 10^{-4}$, and $\hat{\eta}=1.9 \times 10^{-4}$. The main difference from the previous estimates is the value of $\hat{\lambda}$. Since the likelihood factors into four terms, one for each parameter, the estimates are asymptotically independent. Standard errors are again estimated using the second derivative of the log likelihood, which is

$$\begin{aligned} \frac{\partial^2}{\partial \lambda^2} \log L(\lambda, v, \mu, \eta) &= -B(T)/\lambda^2 \\ \frac{\partial^2}{\partial v^2} \log L(\lambda, v, \mu, \eta) &= -I(T)/v^2 \\ \frac{\partial^2}{\partial \mu^2} \log L(\lambda, v, \mu, \eta) &= -D(T)/\mu^2 \\ \frac{\partial^2}{\partial \eta^2} \log L(\lambda, v, \mu, \eta) &= -E(T)/\eta^2. \end{aligned} \quad (3.152)$$

Again, we substitute the estimates and invert the negative of the resulting (diagonal) matrix to compute the asymptotic covariance matrix. The standard errors are $se(\hat{\lambda})=2.1 \times 10^{-4}$, $se(\hat{v})=4.6 \times 10^{-3}$, $se(\hat{\mu})=2.2 \times 10^{-4}$, and $se(\hat{\eta})=1.1 \times 10^{-4}$. One possible explanation to why the population gain parameters v and λ are so different when we do or do not include detailed information could be that only a small proportion of the population is fertile.

The stationary distribution of a BIDE process is, as we have seen earlier, negative binomial with parameters $r=v/\lambda$ and $p=1-\lambda/(\mu+\eta)$, so the mean is $r(1-p)/p$ which we estimate to be 24.8 with a standard error of 14.3. Again, the entire path lies above the stationary mean, this time by 1.2 standard errors. \square

Application (Communications center, continued) We return to the queue in the receipt center of the communications terminal, mentioned in section 3.5. Figure 3.7 depicts the queue length at 5-minute intervals for four hours.

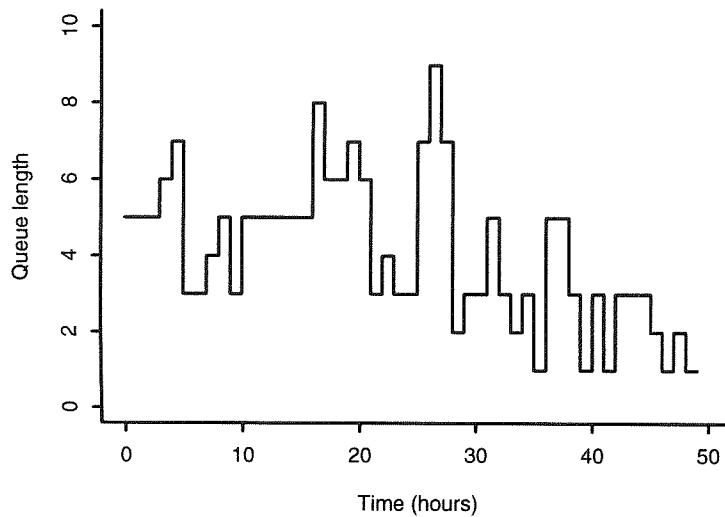


Figure 3.7. Receipt center queue size. Adapted from Hatzikostados and Howe (1967), in R. Cruon (ed.), *Queueing Theory: Recent Developments and Applications*, published by American Elsevier, 1967.

If we had continuous observations we could test the assumptions of an M/M/1 model directly, but since we only observe the discrete skeleton to scale 1/12 of an hour, this is not so straightforward. We will, however, assume an M/M/1 model, and derive the likelihood of the skeleton chain. This is a discrete time Markov chain with transition probabilities $p_{ij}=P_{ij}(1/12)$. The main point of this application is how difficult this estimation problem is, in spite of this being one of the simplest possible models for a queue.

In order to determine $P_{ij}(t)$ we first note that this is a birth and death process with rates λ and μ and no loss possible when the population size is zero. The forward equation is

$$\begin{aligned} \frac{dP_{ij}(t)}{dt} &= \lambda P_{i,j-1}(t) - (\lambda + \mu)P_{ij}(t) + \mu P_{i,j+1}(t), \quad j \geq 1, \\ \frac{dP_{i0}(t)}{dt} &= -\lambda P_{i0}(t) + \mu P_{i1}(t). \end{aligned} \quad (3.153)$$

Using the Laplace transform approach (and a lot of complicated algebra) the solution is (Bailey, 1975, equation 11.65)

$$\begin{aligned} P_{ij}(t) &= \rho^{\frac{1}{2}(j-i)} \{ \exp(-(\lambda+\mu)t) I_{i-j}(2vt) \\ &\quad + \int_0^t \exp(-(\lambda+\mu)s) (\lambda I_{i+j+2}(2vs) \\ &\quad - 2v I_{i+j+1}(2vs) + \mu I_{i+j}(2vs)) ds \} \end{aligned} \quad (3.154)$$

where $\rho = \lambda/\mu$ is the traffic intensity, $v = (\lambda\mu)^{1/2}$, and $I_n(x)$ is the modified Bessel function of the first kind, defined by

$$I_n(x) = (x/2)^n \sum_{k=0}^{\infty} \frac{(x^2/4)^k}{k!(n+k)!}. \quad (3.155)$$

For small values of x we have $I_n(x) \approx (x/2)^n / n!$. Using this approximation in (3.154) we can evaluate the likelihood explicitly in terms of incomplete gamma functions, or, equivalently, cdf's of gamma distributions. It is astonishing how complex the exact probabilities are for this the simplest of all queueing models! No wonder equilibrium approaches are important in queueing theory (and simulation studies, rather than algebraic derivations, the rule).

Evaluating the likelihood (using the approximation to I_n), Figure 3.8 obtains.

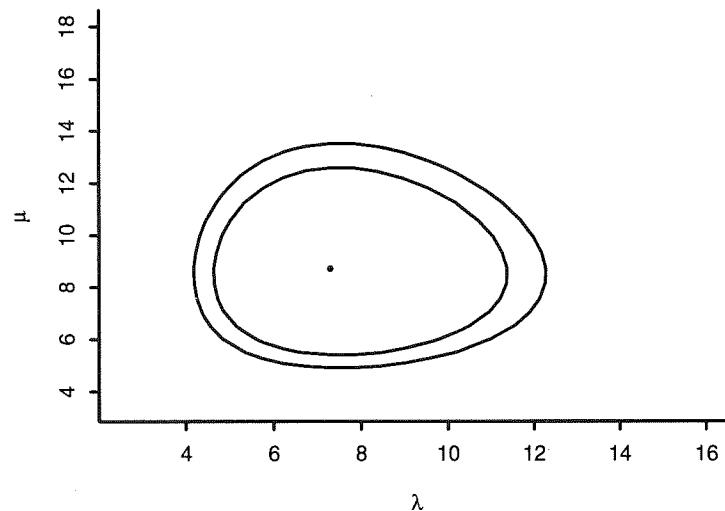


Figure 3.8. Approximate 95% (3 units of log likelihood below the maximum) and 99% (4.6 units below the maximum) contours for the queue log likelihood.

The mle is $\hat{\lambda} = 7.5$, $\hat{\mu} = 8.5$, for an estimated traffic intensity of $\hat{\rho} = \hat{\lambda}/\hat{\mu} = 0.88$. Note

that the likelihood surface is fairly flat near the maximum. The size of the 95% likelihood region is considerable.

A much simpler estimate of ρ can be computed from equilibrium considerations. The theory developed in section 3.5 indicates that the stationary distribution is geometric with mean $\rho/(1-\rho)$. Since the observed mean is 4.08, this corresponds to a ρ -value of 0.80, reasonably close to $\hat{\rho}$. This method does not, however, yield an estimate of λ . \square

3.8. Modeling neural activity

The central nervous system is a complicated system of cells interacting using electric impulses. Apparently there is substantial non-deterministic behavior, i.e., given the same input current to a nerve cell (**neuron**) it may or may not react in a sense that we shall describe below. In this section we study single neurons, and try to deduce facts about their structure (in a probabilistic sense) from measurements of currents, as was described in the last application of the previous chapter.

Figure 3.9 depicts a neuron. It has three main parts: the cell body or **soma**, containing the cell nucleus; the **dendrites** which are a series of root-like extrusions from the soma, and the **axon**, a long wire-like structure connecting the cell to other neurons. The electrical activity in nerve cells is due to the presence of internal and external fluids in the form of salts containing potassium (K^+), sodium (Na^+), or chloride (Cl^-) which are ionized in solution. The ions move from one side of the cell membrane to the other through channels that are specific to a certain ion (such as a potassium channel), and which can be open or closed, as described in the previous chapter. Different neurons have different permeabilities for the various ions of interest. If the concentration of potassium ions is much higher on the inside of the cell than on the outside, this yields a concentration gradient across the membrane, with excess potassium moving outwards. This diffusion of positive ions disturbs the charge balance and results in an electric field opposing the chemical field. When the fields are balanced, an equilibrium state is obtained. This equilibrium has an electric potential difference across the membrane, with the inside more negative, so the **membrane potential** (inside minus outside potential) is negative. This is called the **rest potential**, and is about -60 to -70 mV. The equilibrium can be disturbed either by applying an electric current across the membrane, or by changing the ionic concentration on the outside of the cell. If the change makes the membrane potential more positive it is called **depolarization**. The key feature of these cells is that when depolarization in a neuron exceeds a time-dependent threshold, the potential shoots up dramatically to about $+30$ mV, and then decreases exponentially to the rest level. This **firing** is called an **action potential**.

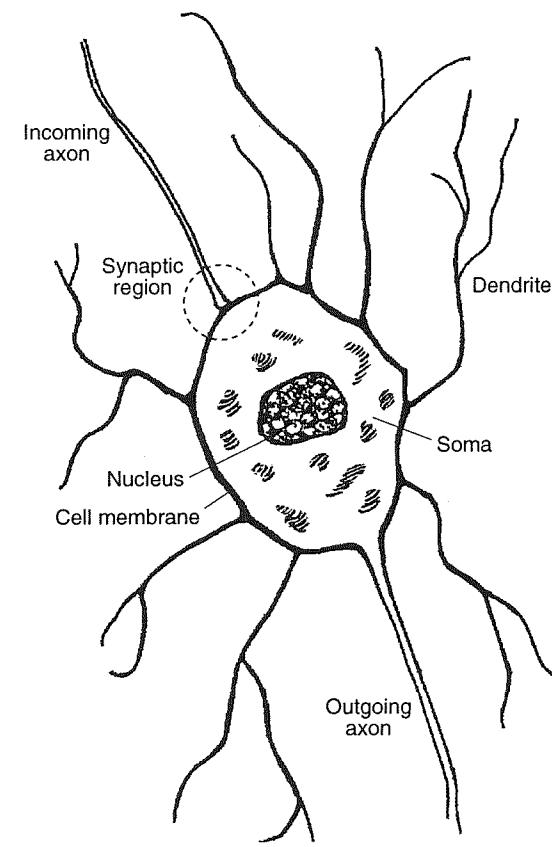


Figure 3.9. A schematic depiction of a neuron.

At the **synapse**, the place where an incoming axon connects with the next cell body, action potential from one neuron releases a chemical, called **neurotransmitter**, from the bouton of the incoming axon (see Figure 3.9). The most common neurotransmitter is acetylcholine (ACh). Binding of acetylcholine to a **receptor**, which protrudes from the protein mass of the receiving neuron, produces a depolarization of the postsynaptic membrane. The soma accumulates all the postsynaptic potentials from the incoming axons. When the integrated effect exceeds the threshold, the neuron fires, as described above. In this section we shall concentrate on the currents in a given synapse, while we return to the action potentials in Chapter 5.

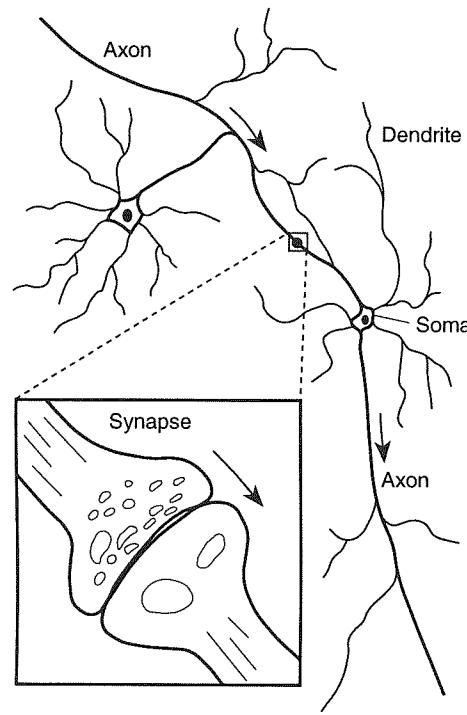


Figure 3.10. A schematic drawing of a network of neurons. The synapse is shown in the detailed blowup.

It is well established that a channel has multiple open and closed states. There are several possible models. The simplest allows for two different closed states. In the resting state, the ligand (ACh) is free. It is known that the ligand must be chemically bound to the receptor for the channel to open. In state two, the ligand is bound to the receptor, but the channel remains closed. The third state is open, with ligand again bound to the receptor. We show this schematically in Figure 3.11, where the arrows denote possible transitions, C stands for closed, O for open, and L for ligand.



Figure 3.11. A three-state model for the ACh channel.

We shall see that this model is too simple. There is evidence of at least two

kinds of open states. One possible such model allows for zero, one or two units of ligand bound to the receptor. In order for the channel to open, at least one ligand must be bound. The schematic is given in Figure 3.12.

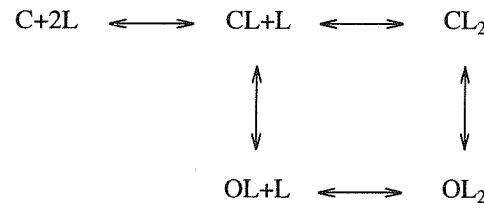


Figure 3.12. A more realistic model for the ACh channel.

We will regard the underlying process as a Markov chain in continuous time, with state space $\{1, 2, 3, 4, 5\}$ corresponding to states $\{C+2L, CL+L, CL_2, OL+L, OL_2\}$, respectively. We are not able to observe this process: all we can tell is whether the channel is open or closed. However, we can write down the intensities based on the chemistry proposed in Figure 3.12. Suppose that we have a concentration c of ligand present. We assume that the binding rate is proportional to the concentration of ligand, so that, for example, $Q_{12} = ck_{12}$. It is useful to group the Q-matrix into transitions between closed states (1, 2, 3), between open states (4, 5), and actual opening or closing. We write the intensity matrix

$$Q = \begin{bmatrix} Q_{cc} & Q_{co} \\ Q_{oc} & Q_{oo} \end{bmatrix} \quad (3.156)$$

where

$$Q_{cc} = \begin{bmatrix} -ck_{12} & ck_{12} & 0 \\ k_{21} & -(k_{21} + ck_{23} + k_{24}) & ck_{23} \\ 0 & k_{32} & -(k_{32} + k_{35}) \end{bmatrix} \quad (3.157)$$

covers transitions between closed states, and c is the concentration of ligand present;

$$Q_{oc} = \begin{bmatrix} 0 & k_{42} & 0 \\ 0 & 0 & k_{53} \end{bmatrix} \quad (3.158)$$

and

$$Q_{co} = \begin{bmatrix} 0 & 0 \\ k_{24} & 0 \\ 0 & k_{35} \end{bmatrix} \quad (3.159)$$

correspond to transitions between the open and closed states, and

$$\mathbf{Q}_{oo} = \begin{pmatrix} -(k_{42} + ck_{45}) & ck_{45} \\ k_{54} & -(k_{53} + k_{54}) \end{pmatrix} \quad (3.160)$$

describes transitions within the open states.

Since we cannot observe the actual process, but only whether the channel is open or closed, we need to use indirect methods to assess the underlying model. Consider a general model with N_o open and N_c closed states. We first try to determine the distribution of an observable quantity such as T , the **open dwell time**. Let

$$F_{ij}(t) = \mathbb{P}(T \leq t, \text{exit to } j \mid \text{in open state at } t=0). \quad (3.161)$$

Define a new process where the closed states are absorbing:

$$\bar{\mathbf{Q}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{oo} & \mathbf{Q}_{oc} \end{pmatrix}. \quad (3.162)$$

Then $F_{ij}(t) = (\bar{P}_{oc}(t))_{ij}$. From the forward equation

$$\begin{aligned} \bar{P}'(t) &= \bar{P}(t)\bar{\mathbf{Q}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \bar{P}_{oc}(t) & \bar{P}_{oo}(t) \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Q}_{oc} & \mathbf{Q}_{oo} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \bar{P}_{oo}(t)\mathbf{Q}_{oo} & \bar{P}_{oo}(t)\mathbf{Q}_{oc} \end{pmatrix} \end{aligned} \quad (3.163)$$

and, since $\bar{P}_{oo}(t) = \exp(\mathbf{Q}_{oo}t)$,

$$f_{ij}(t) = F_{ij}(t) = (\exp(\mathbf{Q}_{oo}t)\mathbf{Q}_{oc})_{ij}. \quad (3.164)$$

Let $\eta_i = \mathbb{P}(\text{open aggregate entered via open state } i)$. Then the density of T , called f_o , is given by

$$f_o(t) = \sum_{i \in O} \sum_{j \in C} \eta_i f_{ij}(t) = \boldsymbol{\eta} \exp(\mathbf{Q}_{oo}t) \mathbf{Q}_{oc} \mathbf{1}^T \quad (3.165)$$

where $\boldsymbol{\eta}$ is the vector of η_i and $\mathbf{1}$ is a vector of N_c ones. In order to compute f_o explicitly we need to do some linear algebra. Every symmetric matrix A (so $A^T = A$) can be diagonalized by an orthogonal matrix B (so $B^T B = \mathbb{I}$ and B is invertible with $B^{-1} = B^T$) in the sense that

$$B^{-1}AB = D \quad (3.166)$$

where D is a diagonal matrix with the eigenvalues of A on the diagonal. Equivalently we can write A in the form

$$A = BDB^{-1}. \quad (3.167)$$

An advantage with diagonalization is that matrix powers become very easy to compute:

$$A^n = BDB^{-1} \times BDB^{-1} \times \cdots \times BDB^{-1} = BD^n B^{-1} \quad (3.168)$$

where D^n is diagonal with diagonal elements d_i^n . In particular, matrix exponentials can be computed explicitly:

$$\begin{aligned} \exp(At) &= \sum_{k=0}^{\infty} (At)^k / k! = \sum_{k=0}^{\infty} B(Dt)^k B^{-1} / k! \\ &= B \sum_{k=1}^{\infty} (Dt)^k / k! B^{-1} = BE(t)B^{-1} \end{aligned} \quad (3.169)$$

where $E(t)$ is a diagonal matrix with diagonal elements $\exp(d_i t)$. Hence if \mathbf{Q}_{oo} were symmetric, it would be straightforward to compute f_o . Unfortunately, \mathbf{Q}_{oo} is not, in general, symmetric. But we are considering a closed chemical system, so the law of detailed balance holds (cf. the discussion in section 2.4). Thus there is a stationary distribution $\boldsymbol{\pi}$ satisfying

$$\pi_i Q_{ij} = \pi_j Q_{ji} \quad (3.170)$$

for all i and j . Let $\Pi = \text{diag}(\pi_i)$. Then the equilibrium condition (3.170) can be written

$$\Pi \mathbf{Q} = \mathbf{Q}^T \Pi. \quad (3.171)$$

Let $\Pi^{1/2} = \text{diag}(\pi_i^{1/2})$. Pre- and post-multiplying (3.171) by $\Pi^{-1/2}$ we see that

$$\Pi^{1/2} \mathbf{Q} \Pi^{-1/2} = \Pi^{-1/2} \mathbf{Q}^T \Pi^{1/2} = (\Pi^{1/2} \mathbf{Q} \Pi^{-1/2})^T \quad (3.172)$$

so $\Pi^{1/2} \mathbf{Q} \Pi^{-1/2}$ is symmetric, and therefore diagonalizable. In particular, writing

$$\Pi = \begin{pmatrix} \Pi_c & \mathbf{0} \\ \mathbf{0} & \Pi_o \end{pmatrix}, \quad (3.173)$$

we see that

$$\begin{pmatrix} \Pi_c^{1/2} & \mathbf{0} \\ \mathbf{0} & \Pi_o^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{Q}_{cc} & \mathbf{Q}_{co} \\ \mathbf{Q}_{oc} & \mathbf{Q}_{oo} \end{pmatrix} \begin{pmatrix} \Pi_c^{-1/2} & \mathbf{0} \\ \mathbf{0} & \Pi_o^{-1/2} \end{pmatrix} = \begin{pmatrix} \Pi_c^{1/2} \mathbf{Q}_{cc} \Pi_c^{-1/2} & \Pi_c^{1/2} \mathbf{Q}_{co} \Pi_o^{-1/2} \\ \Pi_o^{1/2} \mathbf{Q}_{oc} \Pi_c^{-1/2} & \Pi_o^{1/2} \mathbf{Q}_{oo} \Pi_o^{-1/2} \end{pmatrix}. \quad (3.174)$$

Since $\Pi_o^{1/2} \mathbf{Q}_{oo} \Pi_o^{-1/2}$, in particular, is symmetric, it is diagonalizable by an orthogonal matrix B . Consequently \mathbf{Q}_{oo} is diagonalizable using the matrix $\Pi_o^{-1/2} B$, so we can write

$$\mathbf{Q}_{oo} = \Pi_o^{-1/2} B D (\Pi_o^{-1/2} B)^{-1} = \Pi_o^{-1/2} B D B^{-1} \Pi_o^{1/2}. \quad (3.175)$$

Inserting this into equation (3.169) we see that

$$\exp(\mathbf{Q}_{oo}t) = \Pi_o^{-1/2} B E(t) B^{-1} \Pi_o^{1/2}. \quad (3.176)$$

Finally returning to (3.165) we can write

$$f_o(t) = \boldsymbol{\eta} \Pi_o^{-1/2} B E(t) B^{-1} \Pi_o^{1/2} \mathbf{Q}_{oc} \mathbf{1}^T. \quad (3.177)$$

The important thing about this formula is that it is a certain linear combination

of the elements of $E(t)$, i.e.,

$$f_o(t) = \sum_{i=1}^{N_o} \alpha_i \exp(-\beta_i t) \quad (3.178)$$

for some α_i, β_i . In other words, the dwell time in the open state is distributed as a mixture of exponential distributions. If we knew N_o (as in the model depicted in Figure 3.12) it would be relatively straightforward to estimate the parameters α and β using maximum likelihood. If, on the other hand, we do not know N_o , we can perform the estimation for a sequence of possible values of it. We then see which is the smallest N_o that adequately describes the data, e.g., by computing a χ^2 -statistic for goodness of fit. This way of estimating N_o gives some information about the number of open states in the system. Since some of the α_i may be zero, this estimate is actually only a lower bound to the number of open states.

Example (The states of an ACh receptor) Using 163 observed dwell times in open states, obtained from applying a voltage of +100 mV to a planar lipid bilayer containing purified ACh receptor, activated by ACh at the concentration of 50 μM , we get the results in Table 3.3.

Table 3.3 Goodness of fit for different values of N_o ,

N_o	df	χ^2	P-value	i	α	$-1/\lambda$
1	161	393.8	0	1	.0030	48.7
2	159	201.9	.01	2	.0977	5.83
3	157	138.8	.88	3	.8408	.65

We conclude that there are at least 3 open states. \square

The same type of argument shows that the two-dimensional densities are

$$\begin{aligned} f_{oc}(t,s) &= \pi_o \exp(\mathbf{Q}_{oo} t) \mathbf{Q}_{oc} \exp(\mathbf{Q}_{cc} s) \mathbf{Q}_{co} u_o \\ &= \sum_{i=1}^{N_o} \sum_{j=1}^{N_c} \alpha_{ij} \exp(\lambda_i t + \omega_j s) \end{aligned} \quad (3.179)$$

and

$$f_{co}(s,t) = \sum_{i=1}^{N_c} \sum_{j=1}^{N_o} \beta_{ji} \exp(\lambda_i t + \omega_j s). \quad (3.180)$$

We can check the Markov assumption by estimating parameters in the one-dimensional distributions, and then compare the observed and predicted two-dimensional distributions.

In some cases the successive dwell times are independent. Two examples are shown in Figure 3.13. The basic idea is that there is only one closed state that connects directly to any of the open states, so an open dwell time is independent of the preceding closed one (since we know which is the gateway between the two classes of states).

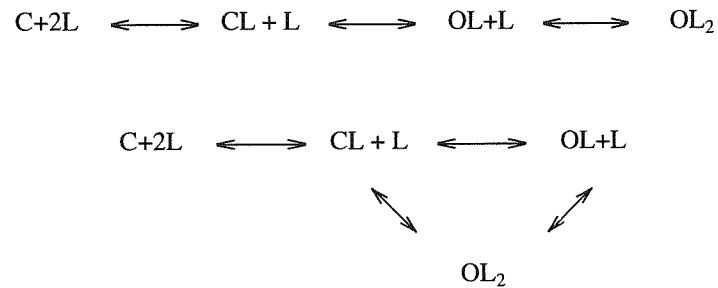


Figure 3.13. Two models with independent dwell times in open and closed states.

The following definitions will be useful. A **graph** is a collection of **nodes**, with certain pairs of nodes connected by **edges**. The graph is **connected** if there is an edge-path between any two nodes. If we think of the state space as a graph with states as nodes, and edges between nodes for which a transition is possible, a **gateway state** is a state such that if it is removed, the graph is disconnected into two components, one with open states and one with closed states. Then the successive dwell times are independent, since the resulting joint density factors: only one ω_j combines with the λ_i . Thus, looking at the covariance between successive dwell times, we may be able to rule out all models with a gateway state. In fact, a tedious computation shows that if S_i are closed and T_i open dwell times we have that

$$\Gamma_c(k) = \mathbf{Cov}(S_i, S_{i+k}) = \sum_{i=1}^{M-1} u_i \kappa_i^{|k|} \quad (3.181)$$

and

$$\Gamma_o(k) = \mathbf{Cov}(T_i, T_{i+k}) = \sum_{i=1}^{M-1} v_i \sigma_i^{|k|} \quad (3.182)$$

where M is the minimum of the number of open entry states, open exit states, closed entry states, and closed exit states.

Example (Autocorrelation function for an ACh receptor) The autocorrelation function, i.e., the correlations $\Gamma_o(k)/\Gamma_o(0)$ as a function of k , for the open aggregate of the ACh receptor channel, based on 1600 openings with ACh concentration 50 μM and voltage +100 mV, is shown in Figure 3.14.

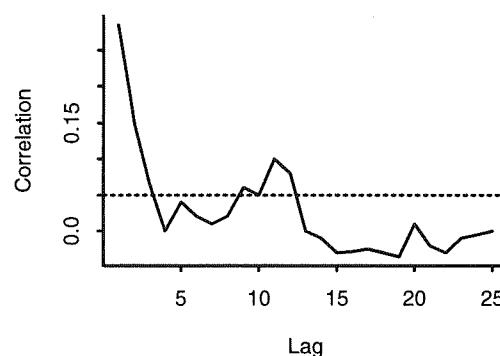


Figure 3.14. Autocorrelation function for the open aggregate of the ACh receptor. Adapted from Labarca et al. (1984); with permission from the Biophysics Society.

The dashed line in the figure corresponds to twice the standard error under the assumption of independence (white noise). The estimate is clearly significantly different from zero, and therefore there must be at least two entry/exit states through which the open and closed aggregates communicate. \square

3.9. Blood formation in cats

The bone marrow of humans and other vertebrates contains a relatively small number of very remarkable cells, the **hematopoietic stem cells**. These cells produce by replication and differentiation all the different kinds of blood cells: red cells which transport oxygen throughout the body; white cells which form the immune defense; and platelets which initiate clotting. It has proven very hard to isolate stem cells from the bone marrow. To maintain a reserve for the lifetime of the animal, the stem cell must be able to self-replicate. Also, to supply cells for development into mature blood cells, stem cells must be able to produce differentiated cells. When an uncommitted cell divides it may replicate itself and/or differentiate and become the head of a sequence of cell divisions leading to mature blood cells. We assume that the regulation of the stem cell is similar. As differentiation proceeds, cells become committed to particular developmental pathways. This complex process of specialization of stem cells into mature blood cells is called **hematopoiesis**. Near the beginning of these pathways we find totipotent cells, i.e., cells that can produce all the different types of mature blood cells. Further along the path a partially committed cell

(called a **progenitor** cell) may still be able to produce several different types of mature cells. Apparently the capacity to produce different types of cells is lost randomly along the pathway until only one type is produced in successive cell divisions. It is impossible to distinguish morphologically stem cells from progenitor cells.

There are two different theories regarding the kinetics of early hematopoiesis. In both, a large supply of stem cells is postulated. These stem cells represent the only source for mature blood cells, and must last the lifetime of the animal. One theory, which we may call the standard theory, is that the entire supply of stem cells is proliferating (actively dividing), perhaps with a very slow rate of division. Thus, at any time all stem cells contribute to hematopoiesis. The other, Kay's (1965) theory of **clonal succession**, hypothesizes that most stem cells are inactive, and that at any time a small number are proliferating. This theory postulates that the stem cells have a finite lifespan, at the end of which they are replaced by previously dormant stem cells (see Brecher et al. (1986) for a discussion). Evidence of clonal succession may have important implications for research on cancer treatments, bone marrow transplantation, and gene transfer methods.

Much of the work on stem cells, particularly the development of the standard theory of hematopoiesis, has been done using experiments on mice. Since the lifetime of mice is relatively short, it is possible that this evidence may be spurious. Single cells (stem cells) from one mouse can maintain hematopoiesis in another throughout its life time. Retrovirally marked stem cells can contribute to hematopoiesis through 2–3 serial transplants. However, this does not by itself prove that stem cells have very long life times, only that these life times are long compared to the life time of a mouse. For this reason, Abkowitz et al. (1988, 1990, 1993) embarked on a series of experiments using female Safari cats: a cross between the domestic cat (*Felis catus*) and the South American Geoffroy wild cat (*Leopardus geoffroyi*).

Although stem cells cannot be isolated in the laboratory, it is possible to analyze different types of progenitor cells. In female cats only one of the X-chromosomes, chosen at random early in embryogenesis, remains active. Hence each cell can be classified as domestic type or Geoffroy type according to which chromosome remains active. The X-linked enzyme glucose phosphate dehydrogenase (G6PD) forms a neutral genetic marker for each cell: the G6PD in each cell is determined from the active X-chromosome. G6PD types are determinable by electrophoresis. The label is conserved through replication and differentiation, so that the G6PD type is the same for a stem cell and all its progeny (the **clone** of the stem cell). Measurements of the proportion of domestic cells were done among colonies grown from progenitor cells sampled from the marrow of the experimental animals. Typically there were 50–100 such colonies grown, so that the G6PD type of 50–100 progenitors could be determined.

Normal cats provided no evidence for or against the clonal succession hypothesis. The model given below indicated that a relatively large number (more than 40) of stem cells were operating, so that small changes in the number of active domestic stem cells would have little influence on the observed proportions among progenitor cells (Figure 3.15).

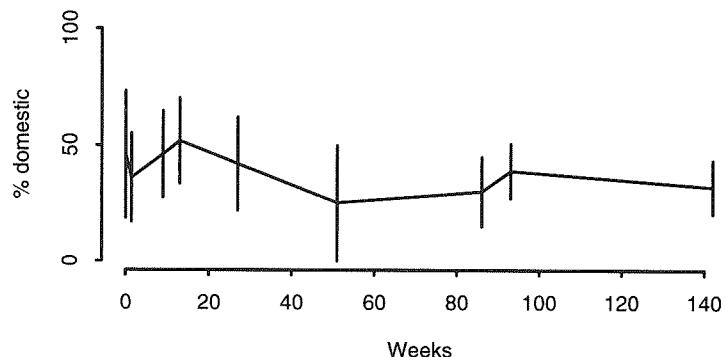


Figure 3.15. Observed proportions of domestic progenitor cells for a normal cat. The vertical lines are pointwise binomial confidence bands. Adapted from Guttrop et al. (1990), published by the Institute of Mathematics and Its Applications.

In order to obtain a more focused view of the system in operation, autologous bone marrow transplants were done. Some bone marrow cells were extracted, the subject was irradiated to kill all remaining marrow, and only very few of the reserved cells were returned. As before the transplants, percentages of domestic type committed progenitor cells were recorded every 2–3 weeks following the recovery of normal blood cell counts. Hematopoiesis was operating as before, but G6PD analysis suggested that much fewer cells were responsible for maintaining normal blood production. Figure 3.16 shows the data for one of the treated cats.

The complex processes which control the differentiation and amplification of hematopoietic cells cannot be determined from the G6PD data alone. Any model designed to learn something from these data must necessarily ignore some details, while at the same time be rich enough to account for most of the structure of the cell kinetics of early hematopoiesis. The first simplifying assumption is to consider the cells of interest as falling into one of two compartments: a stem-cell compartment, denoted C_1 , and a progenitor-cell compartment, denoted C_2 . The essential modeling problems are well illustrated in Figure 3.17 which is an idealized view of the progenitor-cell compartment at two time points. In a real system, many thousands of cells inhabit C_2 , although only about 30 or so are shown in this idealization. At time 1, C_2 is composed of three clones; that is the offspring from three stem cells released into C_2 from

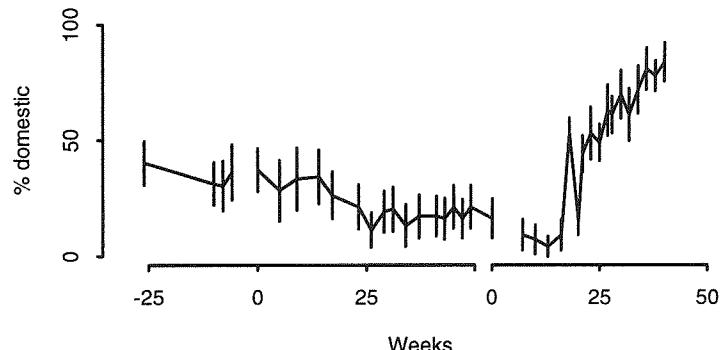


Figure 3.16. Observed proportions of domestic progenitor cells for a cat with two autologous bone marrow transplants. Negative time is before the first transplant. Time starts over at the time of the second transplant. Adapted from Guttrop et al. (1990), published by the Institute of Mathematics and Its Applications.

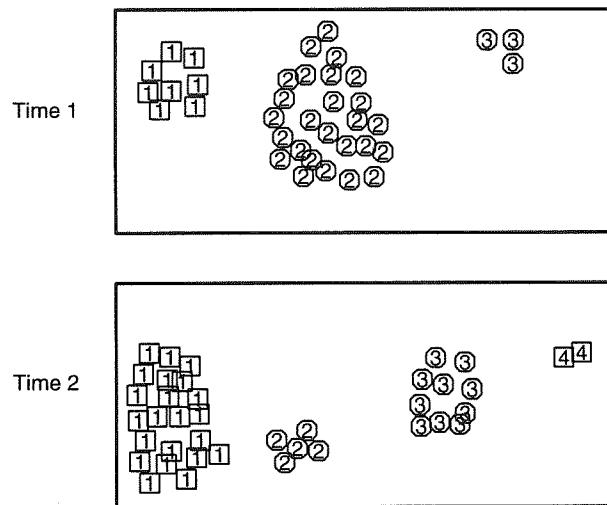


Figure 3.17. An idealization of the progenitor-cell compartment: each number corresponds to a progenitor cell. Cells with the same number are in the same clone. Boxes represent cells having the domestic type G6PD, and circles cells having Geoffroy type G6PD.

C_1 . In this schematic, clone 2 dominates the pool, having more cells than clones 1 and 3 combined. For the cells sampled at time 1, the G6PD type (square denoting domestic and circle Geoffroy) is recorded. Later, at time 2, the composition of C_2 has changed—this evolution being driven by several factors:

- expansion of a clone (e.g., clones 1 and 3)
- terminal differentiation (e.g., most of clone 2)
- release of new stem cells (e.g., clone 4)

We refer to the lifetime of a clone as the time from release of its stem-cell ancestor until terminal differentiation of its constituent cells.

One limitation of the data is that the clone of a cell cannot be identified. For example, cells from clone 2 cannot be distinguished from cells of clone 3, given the binary nature of the marker. If clonal succession explains cellular development, then we would expect the clones to be relatively few in number, and to last a relatively short amount of time. If the standard theory is more appropriate, then clones would have a relatively longer lifetime.

If the first compartment, C_1 , has a large number of dormant or self-replicating cells, it is reasonable to assume that the proportion of domestic type cells in this compartment is constant over time. We denote this proportion by p_d .

As a dynamic process, the number of clones composing C_2 may fluctuate because of terminal differentiation and new stem-cell release. A balance is expected if the process is stable. We assume (at least as a first approximation) that the total number of clones, denoted N , stays constant over time. A number $X(t)$ of these N clones are of the domestic type at time t . The fraction $X(t)/N$ fluctuates between 0 and 1 because a depleted clone may be replaced by a clone having a different G6PD type. Let $X(t)$ be a continuous-time, finite-state, birth and death process. The transition intensities are

$$\begin{aligned} x \rightarrow x+1 &\text{ with intensity } (N-x)\lambda p_d \\ x \rightarrow x-1 &\text{ with intensity } x\lambda(1-p_d). \end{aligned} \quad (3.183)$$

This continuous-time process induces a finite Markov chain (X_1, X_2, \dots, X_m) by restriction to the sampling times t_1, \dots, t_m . Because sampling times are unequal, the transition probabilities for this chain are not stationary. They can be written

$$\begin{aligned} P(X_i=k \mid X_{i-1}=j) &= \sum_{l=\max(0,j+k-N)}^{\min(j,k)} \binom{j}{l} \binom{N-j}{k-l} \\ &\times p_{00}^{N-j-k+l} p_{01}^{k-l} p_{10}^{j-l} p_{11}^l \end{aligned} \quad (3.184)$$

where, writing $\mu = \exp(-\lambda)$ and $\Delta t_i = t_i - t_{i-1}$, we have

$$\begin{aligned} p_{00} &= \mu^{\Delta t_i} + (1-\mu^{\Delta t_i})(1-p_d), \\ p_{01} &= (1-\mu^{\Delta t_i})p_d, \\ p_{10} &= (1-\mu^{\Delta t_i})(1-p_d), \\ p_{11} &= \mu^{\Delta t_i} + (1-\mu^{\Delta t_i})p_d. \end{aligned} \quad (3.185)$$

The number $X(t)$ of domestic type clones in C_2 influences the number of domestic type cells in this progenitor pool. The proportion p_t of domestic type cells in C_2 differs from $X(t)/N$ because all the clones do not have exactly the same number of cells at every point in time. Intuitively, we expect p_t to equal $X(t)/N$ on average, and in fact for certain models the fluctuations of p_t can be quantified. A first approximation is to assume that these fluctuations are negligible, and under this assumption the observation distribution is

$$(Y_i \mid X(t_i), N) \sim \text{Bin}(n_i, \frac{X(t_i)}{N}) \quad (3.186)$$

where Y_i , n_i , and t_i are as defined above. This model is a continuous time **hidden Markov chain**, very much like the discrete time hidden Markov chain in the previous chapter.

Evaluation of the likelihood function L at a given parameter triple $\theta = (N, \lambda, p)$ is a nontrivial exercise because the model probabilities are specified in terms of an unobservable process. The likelihood function is

$$L(\theta) = \mathbf{P}_\theta(Y_1=y_1, Y_2=y_2, \dots, Y_m=y_m) = \mathbf{P}_\theta(Y=y). \quad (3.187)$$

where $y = (y_1, y_2, \dots, y_m)$ are the observed counts. Because the model is specified in two stages, it is natural to rewrite the likelihood

$$\begin{aligned} L(\theta) &= \sum_{x \in S} \mathbf{P}_\theta(Y=y \mid X=x) \mathbf{P}_\theta(X=x) \\ &= \sum_{x \in S} \prod_{i=1}^m \left[\mathbf{P}_\theta(Y_i=y_i \mid X_i=x_i) \mathbf{P}_\theta(X_i=x_i \mid X_{i-1}=x_{i-1}) \right] \end{aligned} \quad (3.188)$$

where the summation is over the set S of possible X -values at the sampling times

$$S = \{x = (x_1, x_2, \dots, x_m) : 0 \leq x_i \leq N\}. \quad (3.189)$$

The cardinality of S is $(N+1)^m$, making an algorithm based on equation (3.188) unviable. A recursive algorithm can be developed by writing the likelihood

$$L(\theta) = \mathbf{P}_\theta(Y_1=y_1) \prod_{i=2}^m \mathbf{P}_\theta(Y_i=y_i \mid Y_1=y_1, \dots, Y_{i-1}=y_{i-1}) \quad (3.190)$$

and noting that each factor in this product can be expanded into a sum over the $N+1$ possible levels of the state at that time. Expansion of the first factor uses

the stationary marginal binomial distribution of X_1 . The latter factors can be expressed as

$$\begin{aligned} \mathbf{P}_\theta(Y_i=y_i \mid Y_1^{i-1}=y_1^{i-1}) &= \sum_{j=0}^N \mathbf{P}_\theta(Y_i=y_i \mid X_i=j) \mathbf{P}_\theta(X_i=j \mid Y_1^{i-1}=y_1^{i-1}) \\ &\equiv \sum_{j=0}^N u_i(j; \theta) v_i(j; \theta). \end{aligned} \quad (3.191)$$

Now compute

$$\begin{aligned} \mathbf{P}_\theta(X_i=j \mid Y_1^{i-1}=y_1^{i-1}) &= \sum_k \mathbf{P}_\theta(X_i=j, X_{i-1}=k \mid Y_1^{i-1}=y_1^{i-1}) \\ &= \mathbf{P}_\theta(X_i=j \mid X_{i-1}=k) \mathbf{P}(X_{i-1}=k \mid Y_1^{i-1}=y_1^{i-1}) \end{aligned} \quad (3.192)$$

using that the hidden state is Markov. Furthermore, using Bayes' theorem,

$$\begin{aligned} \mathbf{P}_\theta(X_{i-1}=k \mid Y_1^{i-1}=y_1^{i-1}) &= \frac{\mathbf{P}_\theta(Y_{i-1}=y_{i-1} \mid X_{i-1}=k) \mathbf{P}_\theta(X_{i-1}=k \mid Y_1^{i-2}=y_1^{i-2})}{\sum_{l=0}^N \mathbf{P}_\theta(Y_{i-1}=y_{i-1} \mid X_{i-1}=l) \mathbf{P}_\theta(X_{i-1}=l \mid Y_1^{i-2}=y_1^{i-2})}. \end{aligned} \quad (3.193)$$

Consequently,

$$v_i(j; \theta) = \sum_{k=0}^N \mathbf{P}_\theta(X_i=j \mid X_{i-1}=k) \frac{u_{i-1}(k; \theta) v_{i-1}(k; \theta)}{\sum_{l=0}^N u_{i-1}(l; \theta) v_{i-1}(l; \theta)}. \quad (3.194)$$

According to the observation distribution (3.186) the u_i are binomial probabilities. Importantly, equation (3.194) shows that the v_i can be computed recursively for increasing i (and known value of θ) using the past u_j 's and v_j 's. Thus the likelihood surface can be evaluated recursively. This recursion allows computation of the likelihood at a given parameter value. The entire likelihood function can be approximated by evaluating L on a large grid. Figure 3.18 shows the contours of L for the parameters N and $1/\lambda$ from one experimental cat. The third parameter, p_d , is replaced by an estimate from pre-transplant data. The likelihood is multimodal, and has a fairly unusual shape. Since one of the parameters is integer-valued, the usual likelihood theory does not apply. Instead the contours in the figure are associated with approximate coverage probabilities determined from a **bootstrap calibration** described below.

Having a way to compute the likelihood on a grid, the following question arises. How far down from the mle do you go in likelihood units to have an approximate $100(1-\alpha)\%$ confidence set, for some small number α ? This is a question of frequency calibration of the likelihood-based confidence set, which we now address. The set of parameters θ whose likelihood is at least $100(1-r)\%$ of the maximum $L(\hat{\theta})$ is called a $100(1-r)\%$ likelihood region for θ .

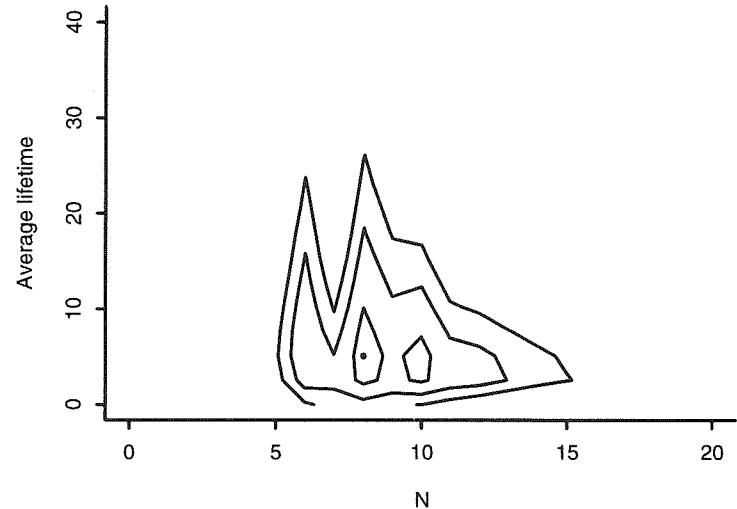


Figure 3.18. Calibrated contours of the joint likelihood of N and $1/\lambda$ for the data from an experimental cat. The contours correspond to 50%, 95%, and 99%, and the maximum is at the dot. Adapted from Guttorp et al. (1990), published by the Institute of Mathematics and Its Applications.

Such a set is

$$\text{LR} = \{\theta : R = \frac{L(\theta)}{L(\hat{\theta})} \geq r\}. \quad (3.195)$$

The cutoff r of the likelihood region determines that region's coverage probability, and so it is important to know how to choose r to achieve approximate 95% coverage, say. In general, the coverage also depends on θ_0 , and is not knowable for any particular procedure for choosing r . Hence we aim for approximate coverage. Standard theory states that asymptotically $2 \log R \sim \chi^2_{\dim(\theta)}$, but, as we have noted, it does not apply here. Without theoretical results on the distribution of R , we perform approximate calibration using the bootstrap.

To apply the bootstrap, it is convenient to modify our notation slightly. Since the likelihood depends on both data and parameters, we use the notation $L(\theta; y)$ for the likelihood determined by an observed time series $y = (y_1, \dots, y_m)$. Introduce

$$T(y; \theta) = 2 \log \frac{L(\theta; y)}{L(\hat{\theta}(y); y)} \quad (3.196)$$

as the random variable whose distribution we want to determine using the bootstrap. The bootstrap algorithm starts with a model fit, that is θ is estimated

from the data y by the maximum likelihood estimate $\hat{\theta}(y)$. (This involves the recursive updating procedure and a grid search.) The next step is to mimic the sampling process via simulation. On the computer, generate n_{boot} time series $z^1, z^2, \dots, z^{n_{boot}}$; each z^j is a series like y . Generation of each z^j takes two steps. First a hidden state X^* is generated by running continuous time Markov process with parameters determined by the fitted model. Then, restriction to the fixed sampling times gives us the binomial success probabilities of the observation distribution which are used to get z^j . For each bootstrapped series z^j , the log likelihood ratio statistic is computed;

$$T_j = T(z^j; \hat{\theta}(z^j)). \quad (3.197)$$

Note that the entire maximum likelihood computation has to be redone for every bootstrapped series z^j . The empirical distribution of the T_j 's converges to the bootstrap distribution of T as n_{boot} increases. This bootstrap distribution is used to determine the cutoff r for the likelihood-based confidence set.

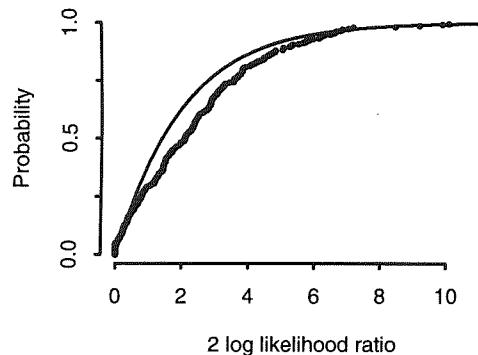


Figure 3.19. Estimated distributions of the log likelihood ratio statistic: The solid curve is the cdf of a χ_2^2 random variable, and the dots determine the empirical distribution function of 240 bootstrapped statistics.

Figure 3.19 compares the cdf of a χ_2^2 random variable to the empirical distribution function of 240 T_j 's computed using the bootstrap algorithm described above. The bootstrapped T_j are stochastically larger than a χ_2^2 random variable, which means that confidence sets produced by bootstrap calibration are larger than those produced appealing to standard theory.

The scientific conclusion (Abkowitz et al., 1990) from this analysis (as well as the analysis of two other treated cats) was that clonal succession is the mechanism of blood production, as evidenced by the short stem cell lifetimes of the treated animals.

3.10. Bibliographic remarks

The description of the various models for cosmic radiation follows Arley (1943). Hooper and Scharff (1958) is an introduction to the physics of cosmic radiation. My main sources for general theory have been Freedman (1983) and Bhattacharaya and Waymire (1990). Asmussen (1987) is a good source for queueing material.

The material on inference has Adke and Manjunath (1984) as source, while the application to baboon data was inspired by Keiding (1977). The work by John Rice and David Brillinger on neurophysiology has been of interest to me for two decades by now. Most of the material in section 3.8 is from Labarca et al. (1984) and Fredkin et al. (1985).

Janis Abkowitz got me interested in hematopoiesis, and section 3.9 builds largely on joint work with her and Michael Newton. A general review of what is known about stem cells is Golde (1991).

3.11. Exercises

Theoretical exercises

1. Prove Lemma 1.
2. For the Pólya birth process, verify equation (3.14).
3. Prove Lemma 3.3.
4. Let $X(t)$ be a continuous time Markov chain with stationary transition probabilities. Determine the stationary distribution of the jump chain, and compare it to the stationary distribution of $X(t)$.
5. Find the stationary distribution of a linear immigration-death process.
6. Derive the stationary distribution of an M/M/1 queue.
7. Consider a G/M/1 queue, and let Y_n be the number of individuals in the queue at the time of the n th arrival. Let D_n be the number of departures between the n th and $(n+1)$ th arrivals.
 - (a) Show that $Y_{n+1} = Y_n + 1 - D_n$.
 - (b) Show that Y_n is a Markov chain, and determine its transition matrix.
8. For a linear immigration-death process, find the asymptotic distribution of $\hat{V}\hat{\kappa}$.

Hint: Do a bivariate Taylor expansion of the function $f(x,y)=x/y$, and use the asymptotic distribution of $(\hat{V}, \hat{\kappa})$.
9. Consider a queue with $\lambda_n = \lambda$ and $\mu_n = \mu$, but in which the queue can become negative. An example is a taxi queue, where taxis waiting with no passengers has a positive queue length, while passengers waiting with no taxis corresponds to a negative queue length. Both passengers and taxis arrive according to a Poisson process. Loading time is negligible. Derive the stationary distribution of queue length.

10. Show that the discrete skeleton of a linear BID-process is a Bienaymé-Galton-Watson process with immigration.

11. Show that the jump chain of a linear birth-death process is a random walk.

12. Elephants wander around in herds of different sizes. It can happen that when two herds meet, they amalgamate to form a single herd. It is of course also possible that a herd may break in two. Holgate's (1967) model for this is the following. Let $X(t)$ be the number of herds at time t , and assume that the total number K of elephants is constant. If there are j herds, the probability that two merge within the next h time units is $(j-1)\mu h + o(h)$, and the probability that one herd will split off is $(K-j)\lambda h + o(h)$. Thus $Q_{i,i+1} = (K-i)\lambda$, $Q_{i,i-1} = (i-1)\mu$, and $Q_{ii} = -(K-i)\lambda - (i-1)\mu$. We are interested in estimating the unknown parameter $\theta = (\lambda, \mu)$.

(a) Show that the log likelihood is

$$\begin{aligned} l_T(\lambda, \mu) &= \sum_1^{K-1} N_T(i, i+1) \log((K-i)\lambda) + \sum_2^K N_T(i, i-1) \log((i-1)\mu) \\ &\quad - \sum_1^K A_T(i)((K-i)\lambda + (i-1)\mu) \end{aligned}$$

where $A_T(i) = \int_0^T \mathbb{1}(X(t)=i) dt$.

(b) Let U_t be the number of jumps up before time t , and D_t the number of jumps down. Show that

$$\hat{\mu}_T = \frac{U_T}{\sum_1^{K-1} (K-i)A_T(i)} \quad \text{and} \quad \hat{\lambda}_T = \frac{D_T}{\sum_2^K (i-1)A_T(i)}.$$

(c) Find the asymptotic covariance matrix for the mle.

13. Let $X(t)$ be a Poisson process with rate λ .

(a) Show that, given $X(t)=n$ the jump times τ_1, \dots, τ_n have the same distribution as an ordered sample of size n from the uniform distribution on $(0, t)$.

(b) Now assume that the rate depends on time, so $\lambda=\lambda(t)$. Show that the jump times now are distributed as an ordered sample of size n from a distribution with cdf $F(x)=\Lambda(x)/\Lambda(t)$ where $\Lambda(x)=\int_0^x \lambda(u) du$.

(c) Let $X(t)$ be as in part (b). Show that $Y(t)=X(\Lambda^{-1}(t))$ is a unit rate Poisson process, so a nonhomogeneous Poisson process is obtained from a homogeneous one by letting the speed with which time goes by vary.

14. Using the results from the previous question, suppose we observe a Poisson process and wish to test the hypothesis that the rate is constant against the hypothesis that it is monotonically increasing with time. A test statistic suggested in the reliability literature (e.g., Barlow et al., 1972, Chapter 6) is the **total time on test** statistic, given by

$$V_n = \sum_{i=1}^n \tau_i / \tau_n.$$

Show that the conditional distribution of V_n , given that $\tau_n=t$, is that of the sum of $n-1$ iid uniform random variables, and deduce that V_n under the null hypothesis has approximately a normal distribution. Would you reject for small or large values of V_n ?

15. Consider a queue with input $X(t)$ following a nonhomogeneous Poisson process with rate $\lambda(t)$, an infinite number of servers, and service time distribution with density $f(t)$.

(a) Show that the output $Y(t)$ from this queue is a nonhomogeneous Poisson process with rate $\gamma(t) = \int_0^t \lambda(t-u)f(u)du$.

(b) Now suppose that $\lambda(t) \equiv 1$. Show that the mean of the output process is $m(t) = \mathbb{E}Y(t) = \int_0^t F(x)dx$, where F is the cdf of the service time distribution.

(c) Let $S_i = m^{-1}(\tau_i)$. Show that the events S_i are a Poisson process with rate

$$\gamma^T(t) = \frac{\gamma(m^{-1}(t))}{F(m^{-1}(t))}.$$

16. In the context of Exercise 15, suppose that we wish to test the hypothesis that $\lambda(t)=\lambda$ against the hypothesis that $\lambda(t)$ is an increasing function of t , based on observations of the output process for a known service time distribution f . Show that this testing problem is equivalent to testing $\gamma^T(t)=\lambda$ against $\gamma^T(t)$ being an increasing function. In other words, the test based on the output times of one process is equivalent to a test based on the input times of a related process, such as that in Exercise 14.

17. Let $X(t)$ be a Poisson process with rate

$$\lambda(t) = \begin{cases} 1 + \theta t^{-\frac{1}{2}} & t \geq 1 \\ 0 & 0 \leq t \leq 1 \end{cases}$$

This is called a **weak trend** (Brown, 1972).

(a) Derive the likelihood equation for estimating θ based on observing $X(t), t \leq T$, and comment on its solution.

(b) A simpler estimate is the moment type estimator given by the solution to the equation $X(T) = \mathbb{E}_\theta X(T)$. Show that it is unbiased but inconsistent.

18. A system contains j energy quanta. It can lose one (if $j \geq 0$) with intensity μ , or gain one with intensity λ .

(a) Show that the equilibrium distribution is a **Gibbs distribution** of the form $\pi_j \propto \exp(-\beta j)$.

(b) Now suppose that at energy level j there are several states between which the process moves micro-reversibly, so that $Q(x, x') = Q(x', x)$ if x and x' have the same energy, and $\lambda Q(x, x') = \mu Q(x', x)$ if the energy $E(x') = E(x) + 1$. Show

that this system also has equilibrium distribution of the Gibbs form

$$\pi(x) \propto \exp(-\beta E(x)).$$

19. Consider a telephone exchange with k lines. When all lines are busy, any incoming telephone call is lost. Assume that calls arrive according to a homogeneous Poisson process, and that their lengths are exponentially distributed.

(a) Find the stationary distribution of the chain.

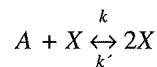
(b) Find the probability that an incoming call is lost (Erlang's loss formula).

20. Let

$$Q_1 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad Q_2 = \begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix}$$

be the intensity matrices of two Markov processes. Show that their discrete skeletons to scale $4\pi/\sqrt{3}$ are identical.

21. Consider the autocatalytic reaction



where the catalyst A is abundant, so that we can assume that it has constant concentration c_A . The number of X -molecules can be viewed as a birth-death process, where the birth rate is $\lambda_n = kc_A n$ and the death rate is $\mu_n = k'n(n-1)$. Find the stationary distribution for this reaction.

22. Consider a Markov chain with nonstationary intensity of the form $q_n(t)$. Let $\lambda(t)$ be a continuous increasing function of time, and consider changing the time scale using $\lambda(t)$, so that we observe $Y(t)=X(\lambda(t))$.

(a) Show that Y has intensity $q_n(\lambda(t))\lambda'(t)$.

(b) Consider a Pólya process observed using the time change $\lambda(t)=-\log P_0(t)$. Show that the resulting process is a (nonhomogeneous) Poisson process.

23. Let $X(t)$ be a linear birth and death process starting from 1 individual.

(a) Show that $P(X(t)=i)=(1-\alpha)(1-\beta)\beta^{i-1}$, $i \geq 1$, where

$$\alpha=\mu \frac{1-\exp((\mu-\lambda)t)}{\lambda-\mu\exp((\mu-\lambda)t)} \quad \text{and} \quad \beta=\lambda \frac{1-\exp((\mu-\lambda)t)}{\lambda-\mu\exp((\mu-\lambda)t)}.$$

(b) Deduce that if $\mu < \lambda$ the probability of extinction is μ/λ .

(c) Let T_n be the first time to hit n . Show that

$$P(T_n \leq t) \approx \frac{(1-\alpha)\beta^{n-1}}{1-\mu/\lambda}.$$

24. A simple model for the folding of protein molecules into its native structure is given by a chain of $N+1$ amino acids, with N bonds between them. The connecting bond between two neighboring amino acids can either be characterized as correct (native) or incorrect (non-native). These bonds change over time. An incorrect bond becomes correct at rate k_0 , while a correct bond becomes incorrect at rate k_1 . Let $X(T)$ denote the number of correct bonds at time t . We are interested in the mean time τ_i until $X(t)=0$, starting from i .

(a) Show that $X(t)$ is a finite birth and death process, and write down its intensity matrix.

(b) Show that $\tau=(0, \tau_1, \dots, \tau_N)$ satisfies $\tau Q = -1$.

(c) Verify that

$$\tau_i = \frac{1}{Nk_0} \sum_{n=0}^{i-1} \binom{N-1}{n}^{-1} \sum_{m=n+1}^N \binom{N}{m} K^{m-n}$$

where $K=k_0/k_1$.

(d) Statistical thermodynamics suggests that $K=v \exp(-U/kT)$ where $v+1$ is the number of possible bonds and U is an energy penalty for making an incorrect bond. Here k is Boltzmann's constant and T the absolute temperature. Show that for large N and k_0 not too small we have, approximately, that

$$\tau_i \approx \frac{1}{Nk_0} (1+v \exp(-U/kT))^N$$

regardless of the initial state i .

Remark: The case $U=0$, or a completely random bond replacement strategy, yields an enormously long mean folding time, while a search strategy biased towards correct bonds yields much more realistic folding times (Zwanzig et al., 1992).

Computing exercises

C1. Given an exponential random number generator, and a function to simulate a discrete time Markov chain, show how to generate a continuous time Markov chain.

C2. How would you solve the stationary distribution equation $Q\pi=0$?

C3. Given a Q -matrix of your choice, how big a sample do you need to estimate it accurately? How does the sample size depend on the dimensionality of Q ?

C4. Generate a linear BIDE-process with parameters given by the fit to the baboon data from page 157 and look at successive stretches of 373 days. Assess the probability that any such stretch lies entirely above its stationary mean.

C5. (a) Show that equation (3.100) is monotone as a function of m .

(b) The design of a service station is a compromise between the cost of decreasing the queue length (by adding servers) and the cost of additional idle time for the servers. Study the relationship between these costs.

Data exercises

D1. The data in Table 3.4 are counts of the number of alpha-particles emitted from a piece of polonium in successive 7.5-second intervals (Rutherford and Geiger, 1910).

Table 3.4 Alpha-particle counts

3	7	4	4	2	3	2	0	5	2
5	4	3	5	4	2	5	4	1	3
3	1	5	2	8	2	2	2	3	4
2	6	7	4	2	6	4	5	10	4

These data are part of a larger experiment, from which the average number of particles per minute was found to be 31.0. Investigate the hypothesis that the data come from a homogeneous Poisson process with this rate.

D2. A controversial issue in the study of multiple sclerosis (MS) is whether or not the disease is contagious. Kurtzke and Hellested (1986) noted the apparently sudden occurrence of MS in the Faroe Islands shortly after the arrival of British troops in 1941. As there had been no cases earlier in the century, they argued that this was an indication of an infectious agent at work. Supposing that MS cases occur following a Poisson process, one would expect that the intensity $\lambda(t)$ would be increasing in time if the disease is infectious, while if the British troops introduced some non-infectious agent precipitating MS, the intensity would be constant.

The data for disease onset in years from 1941 are given in Table 3.5 (actually, the data were only given to nearest year, and have been uniformly redistributed between $j - \frac{1}{2}$ and $j + \frac{1}{2}$; cf. Joseph et al., 1990).

Table 3.5 MS onset for Faroe Island population

2.43	2.62	2.94	3.28	3.92	4.21	4.31	4.66
4.99	5.40	5.60	6.24	6.79	7.01	7.79	8.50
11.19	11.59	12.95	13.85	14.82	15.54	16.90	18.10
18.13	19.32	20.01	23.68	26.79	27.89	29.02	31.92

Assume that the incubation time for the disease (corresponding to the service time in the queueing description) has a Weibull distribution with density

$$f(x) = cx^{c-1} \exp(-x^c)$$

and $c = 1/3$, corresponding to a mean incubation time of 6 years. Using the results in Exercises 14–16, assess the contagion theory for MS based on these data.

D3. Using the baboon data from page 157, use the jump chain of births and deaths/emigrations (ignoring the few immigrations) to test the hypothesis that these form a random walk (cf. Exercise 11).

D4. The data in data set 4 are successive counts of the number of gold particles observed by Westgren (1916, Experiment C) in a fixed small volume of water every 1.39 seconds. Physical theory suggests that these counts should be described by a linear immigration-death process.

(a) Determine a confidence set for the rates in this process.

(b) Test the goodness of fit of the model using the autocorrelation approach of the application on page 100.

D5. Table 3.6 contains data from Sartwell (1950) on the incubation period of streptococcal infection.

Table 3.6 Streptococcal incubation time in 12-hr periods

Incubation period	0–1	1–2	2–3	3–4	4–5	5–6
Number	0	1	7	11	11	7
Incubation period	6–7	7–8	8–9	9–10	10–11	11–12
Number	5	4	2	2	0	1

Assuming, following Morgan and Watts (1980), that the infections started with one bacterium, use the theory developed in Exercise 23 to estimate μ , λ , and the incubation period n (we assume that the incubation period is the time needed to reach a threshold n of bacteria in the organism). Assess the precision of your estimates. The form of the parameters α and β suggest that $\lambda - \mu$ is an important parameter. How well can it be estimated from these data?

D6. The whooping crane is a very rare migratory bird with breeding grounds in Canada's Northwest Territories and wintering grounds in Texas. Miller et al. (1974) give annual counts from 1938–1972 of whooping cranes arriving in Texas in the fall. Using a linear BID model to fit these data, try to distinguish between the hypothesis of a constant breeding population and that of a breeding population proportional to the population size. How would you estimate the extinction probability at the beginning of the observation period? At the end?

Table 3.7 Texas whooping crane counts

14	22	26	15	19	21	18	17	25	31
30	34	31	25	21	24	21	28	24	26
32	33	36	38	32	33	42	44	43	48
50	56	57	56	51					

CHAPTER 4

Markov random fields

When “time” is a spatial index, the random process is called a random field. We generalize the Markov assumption to deal with this situation, and discuss some classes of processes that originated in statistical physics, and have proved quite useful in a variety of applications. The Ising model of ferromagnetism is applied to an agricultural problem. A simple autoregressive model is useful for improving some blurred astronomical pictures, while a hidden Markov random field is applied to the statistical analysis of pedigrees in population genetics.

4.1. The Ising model of ferromagnetism

The magnetization of a permanent magnet diminishes in strength as the magnet is heated. Above a certain critical temperature the magnet stops being a magnet. At the other extreme, for very low temperatures non-magnets may exhibit spontaneous magnetization. At the atomic level, each atom is by itself a small magnet. A material is magnetized when all (or most) of the atoms align magnetically. The interaction between these tiny magnets explains spontaneous magnetization.

Consider N sites (or atoms), ordered linearly, and associate with each site a magnetic dipole (spin) which can be either positive or negative (Figure 4.1).

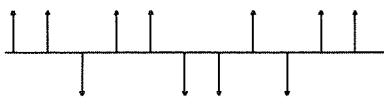


Figure 4.1. Schematic picture of a linear magnet. Adapted from *Markov Random Fields and their Applications*, by Ross Kindermann and J. Laurie Snell, Contemporary Mathematics, Volume 1, by permission of the American Mathematical Society.