that $A(x)$ is strictly increasing for $1 < x < R$. Further, using the assumption of $\lambda\mu < 1$ and the assumption (ii), it follows that $A(1) = \mu < 1/\lambda$ and $A(x) \to \infty$ as $x \to R$. Thus we can conclude that there is a unique $z_0 \in (1, R)$ such that the graph of $A(x)$ and the line $y = 1/\lambda$ intersect at $x = z_0$. Hence $1 - \lambda A(x)$ has a unique zero z_0 on the interval $(1, R)$. As a by-product of the proof, we find that

$$A(x) - 1/\lambda < 0 \quad \text{for } 1 < x < z_0$$

and

$$A(x) - 1/\lambda > 0 \quad \text{for } z_0 < x < R.$$

Using the mean-value theorem $f(c + d) = f(c) + df'(c) + \frac{1}{2}d^2 f''(c + \theta d)$ for some $0 \leq \theta \leq 1$, it follows that $f'(c) \neq 0$ when $f(x)$ is strictly convex on (a, b) and changes sign at $x = c$ with $c \in (a, b)$. The function $A(x) - 1/\lambda$ is strictly convex on $(1, R)$, since the second derivative $A''(x)$ is positive. Also $A(x) - 1/\lambda$ changes sign at $x = z_0$. Thus the derivative of $A(x) - 1/\lambda$ at $x = z_0$ is not equal to 0, implying that the zero z_0 of $1 - \lambda A(x)$ is of multiplicity 1. To verify that $1 - \lambda A(z)$ has no zero for $1 < |z| < z_0$, we use the basic result that the power series representation $A(z) = \sum_{n=0}^{\infty} a_n z^n$ extends to $|z| < R$. Since the a_n's are reals and positive, it now follows that $|A(z)| \leq A(|z|)$ for $|z| < R$. This inequality and the inequality $A(x) - 1/\lambda < 0$ for $1 < x < z_0$ imply that $|A(z)| < 1/\lambda$ for all z in the domain $1 < |z| < z_0$, showing that $1 - \lambda A(z)$ has no zero in this domain. It remains to verify that $1 - \lambda A(z)$ has z_0 as the only zero on the circle $|z| = z_0$. Thus we must prove that $x = z_0$ and $y = 0$ for any complex number $z = x + iy$ with $A(z) = 1/\lambda$ and $|z| = z_0$. To do so, note that $|z| = (x^2 + y^2)^{1/2}$ and so $(x^2 + y^2)^{1/2} = z_0$ implying that $x \leq z_0$. Next, using that $|e^{iu}| = 1$ for any real u, we obtain from (C.11) that

$$\frac{1}{\lambda} = |A(z)| \leq \int_0^\infty |e^{-\lambda(1-z)t}| \{1 - B(t)\}\, dt$$

$$= \int_0^\infty e^{-\lambda(1-z_0)t} \{1 - B(t)\}\, dt = \frac{1}{\lambda}.$$

This implies that x must be equal to z_0. Also, by $(x^2 + y^2)^{1/2} = z_0$, we find $y = 0$. This completes the proof.

APPENDIX D. NUMERICAL SOLUTION OF MARKOV CHAIN EQUATIONS

In Markov chain applications one often has to solve a finite system of linear equations of the form

$$x_i = \sum_{j=1}^{N} p_{ji} x_j, \quad i = 1, \ldots, N, \tag{D.1}$$

$$\sum_{i=1}^{N} x_i = 1. \tag{D.2}$$

Under a mild regularity condition posed on the underlying Markov chain (see Chapter 2), this system of linear equations has a unique solution. Moreover, any solution to the balance equations (D.1) is uniquely determined up to a multiplicative constant. This constant is found by using the normalizing equation (D.2).

In general there are two methods to solve the Markov chain equations: (a) direct methods; (b) iterative methods.

A convenient direct method is a Gaussian elimination method such as the Gauss–Jordan method. This reliable method is recommended as long as the dimension N of the system of linear equations does not exceed 200 (say). The computational effort of Gaussian elimination is proportional to N^3. Ready-to-use codes for Gaussian elimination methods are widely available; see e.g. the source book by Press *et al* (1986). A Gaussian elimination method requires that the whole coefficient matrix is stored, since this matrix must be updated at each step of the algorithm. This explains why a Gaussian elimination method suffers from computer memory problems when N gets large. Another direct method for solving Markov chain equations is the probabilistic method proposed in Grassmann *et al* (1985). However, this method has the same drawbacks as Gaussian elimination when N gets large. In solving (D.1) and (D.2) by a direct method one of the balance equations (D.1) is omitted in order to obtain a square system of linear equations.

Iterative method of successive overrelaxation

Iterative methods have to be used when the size of the system of linear equations gets large. In specific applications an iterative method can usually avoid computer memory problems by exploiting the structure of the application. An iterative method works with the original matrix of coefficients. In applications these coefficients are usually composed from a few constants. Then only these constants have to be stored in memory.

The iterative method of successive overrelaxation is a suitable method for solving the linear equations of large Markov chains. The well-known Gauss–Seidel method is a special case of the method of successive overrelaxation. The iterative methods generate a sequence of vectors $x^{(0)} \to x^{(1)} \to x^{(2)} \to \ldots$ converging towards a solution of the balance equation (D.1). The normalization is done at the end of the calculations. To apply successive overrelaxation, we first rewrite the balance equations (D.1) into the form

$$x_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} x_{ij}, \quad i = 1, \ldots, N,$$

where

$$a_{ij} = \frac{p_{ji}}{1 - p_{ii}}, \quad i, j = 1, \ldots, N, \ j \neq i.$$

Here $B(x)$ is the probability distribution function of a positive random variable with probability density $b(x)$ and finite mean m, while λ and σ are positive constants such that $\lambda m/\sigma < 1$. The integro-differential equation appears in Section 1.8 and the function $q(x)$ represents amongst others the complementary waiting-time distribution in the standard single-server queueing system with Poisson input. From physical considerations we know that $q(x)$ is decreasing in x and has limit 0 as $x \to \infty$.

Denoting by $q^*(s)$ and $b^*(s)$ the Laplace transforms of $q(x)$ and $b(x)$, it follows by using (C.1)–(C.4) that for all s with $Re(s) > 0$,

$$sq^*(s) - q(0) = -\frac{\lambda}{\sigma}\left\{\frac{1}{s} - \frac{b^*(s)}{s}\right\} + \frac{\lambda}{\sigma}q^*(s) - \frac{\lambda}{\sigma}q^*(s)b^*(s).$$

The unknown $q(0)$ follows by letting $s \to 0$ in both sides of this equation and applying (C.6) with f replaced by q. Noting that $\lim_{s\to 0}\{1 - b^*(s)\}/s = m$ (use L'Hospital's rule), we obtain $q(\infty) - q(0) = -\lambda m/\sigma$. Thus, by $q(\infty) = 0$,

$$q(0) = \frac{\lambda m}{\sigma}.$$

Suppose now that $b(x)$ is the probability density of a Coxian-2 distributed random variable X. In other words, $X = X_1 + X_2$ with probability b and $X = X_1$ with probability $1 - b$ for two independent exponentials X_1 and X_2 with respective means $1/\mu_1$ and $1/\mu_2$. Using (C.5) and noting that

$$E(e^{-sX}) = bE\left(e^{-s(X_1+X_2)}\right) + (1-b)E\left(e^{-sX_1}\right),$$

it follows that

$$b^*(s) = \frac{b\mu_1\mu_2}{(s+\mu_1)(s+\mu_2)} + \frac{(1-b)\mu_1}{s+\mu_1}.$$

Next, using that $m = 1/\mu_1 + b/\mu_2$, we obtain after some algebra

$$q^*(s) = \frac{\lambda(\sigma\mu_1\mu_2)^{-1}\left[(b\mu_1+\mu_2)s + b\mu_1(\mu_1+\mu_2) + \mu_2^2\right]}{(s+\mu_1)(s+\mu_2) - (\lambda/\sigma)(s+b\mu_1+\mu_2)}.$$

Thus, using the inversion formula (C.7), we obtain

$$q(x) = A_1e^{-b_1x} + A_2e^{-b_2x}, \quad x \geq 0.$$

Explicit expressions for the coefficients A_i and b_i are easily derived. However, they are omitted since our only purpose is to establish that $q(x)$ is the sum of two exponential functions when $b(x)$ is a Coxian-2 density.

Asymptotic expansions from the generating function

The generating function is in fact a special case of the Laplace transform. Let $\{p_n, n = 0, 1, \ldots\}$ be a discrete probability distribution. The generating function

(or z-transform) of this distribution is defined by

$$P(z) = \sum_{n=0}^{\infty} p_n z^n, \quad |z| \leq 1,$$

with z being a complex variable. Note that $P(z)$ can be interpreted as a Laplace transform by taking $z = e^{-s}$. In many applications the p_n's are unknown probabilities, but an explicit expression for the generating function $P(z)$ is obtained by some reasoning. Under rather weak regularity conditions an asymptotic estimate for the probability p_n with n large can be derived from the generating function $P(z)$. As in the Fast Fourier Transform method which provides a numerical tool for recovering all the p_j's from $P(z)$, we need help from complex numbers. To be specific, let us assume that $P(z)$ can be represented as

$$P(z) = \frac{N(z)}{D(z)},$$

where $N(z)$ and $D(z)$ are analytic functions whose domains of definition can be extended to a region $|z| < R$ in the complex plane for some $R > 1$. It is no restriction to assume that $N(z)$ and $D(z)$ have no common zeros; otherwise, cancel out common zeros. Let us further assume that the following regularity conditions are satisfied:

C1 The equation $D(z) = 0$ has a *real* root z_0 on the interval $(1, R)$.

C2 The function $D(z)$ has no zeros in the domain $1 < |z| < z_0$ of the complex plane.

C3 The zero $z = z_0$ of $D(z)$ is of multiplicity 1 and is the only zero of $D(z)$ on the circle $|z| = z_0$.

Theorem C.1 *Under the conditions C1–C3,*

$$p_j \approx \gamma_0 z_0^{-j} \quad \textit{for } j \textit{ large enough,} \tag{C.8}$$

where the constant γ_0 is given by

$$\gamma_0 = -\frac{1}{z_0}\frac{N(z_0)}{D'(z_0)}. \tag{C.9}$$

Here $D'(z_0)$ denotes the derivative of $D(x)$ at $x = z_0$.

Proof We first mention the following basic facts from complex function theory. The most important fact is that a function $f(z)$ is analytic at a point $z = a$ if and only if $f(z)$ can be expanded in a power series $f(z) = \Sigma_{n=0}^{\infty} a_n(z-a)^n$ in $|z-a| < \rho$ for some $\rho > 0$; see e.g. Courant (1964). The coefficient a_n of the Taylor series is the nth derivative of $f(z)$ at $z = a$ divided by $n!$. The analytic function $f(z)$ is said to have a zero of multiplicity k in $z = a$ if $a_0 = \ldots = a_{k-1} = 0$ and $a_k \neq 0$. Another basic result is the following. The Taylor series $\Sigma_{n=0}^{\infty} a_n(z-a)^n$ of a function

$f(z)$ at the point $z = a$ coincides with the function $f(z)$ in the interior of the largest circle whose interior lies wholly within the domain where $f(z)$ is analytic.

The proof of (C.8) now proceeds as follows. The conditions C1–C3 imply that there is a circle around $z = 0$ with radius R_0 larger than z_0 such that $P(z)$ is analytic in $|z| < R_0$ except for the isolated point $z = z_0$. Since $D(z)$ has a zero of multiplicity 1 at $z = z_0$, it follows from the Taylor series that $D(z) = (z - z_0)\phi(z)$ in $|z| < R_0$, where $\phi(z)$ is an analytic function with $\phi(z_0) \neq 0$. Thus we can write $P(z)$ as $P(z) = H(z)/(z - z_0)$ for some analytic function $H(z)$ in $|z| < R_0$ with $H(z_0) \neq 0$. Using a Taylor expansion $H(z) = H(z_0) + (z - z_0)U(z)$, we next find that $P(z)$ can be represented as

$$P(z) = \frac{r_0}{z - z_0} + U(z) \tag{C.10}$$

in $|z| < R_0$, $z \neq z_0$. Here $U(z)$ is an analytic function in the domain $|z| < R_0$ and the residue r_0 is given by

$$r_0 = \lim_{z \to z_0} (z - z_0)P(z) = N(z_0)/D'(z_0).$$

The remainder of the proof is simple. Since $U(z)$ is analytic for $|z| < R_0$ we have the power series representation $U(z) = \sum_{j=0}^{\infty} u_j z^j$ for $|z| < R_0$. Let R_1 be any number with $z_0 < R_1 < R_0$. Then, for some constant b, $|u_j| \leq bR_1^{-j}$ for all $j \geq 0$. This follows from the fact that the series $\sum_{j=0}^{\infty} u_j z^j$ is convergent for $z = R_1$ and so the sequence $\{u_j R_1^j\}$ is bounded. Using the power series representation of $U(z)$ and the fact that the power series representation $P(z) = \sum_{j=0}^{\infty} p_j z^j$ extends to $|z| < z_0$, it follows from (C.10) that

$$\sum_{j=0}^{\infty} p_j z^j = \frac{-r_0}{z_0} \sum_{j=0}^{\infty} (z/z_0)^j + \sum_{j=0}^{\infty} u_j z^j, \quad |z| < z_0.$$

Equating coefficients yields

$$p_j = -r_0 z_0^{-j-1} + u_j, \quad j \geq 0.$$

Since $|u_j| \leq bR_1^{-j}$, $j \geq 0$, for some constant b and since $R_1 > z_0$, the coefficient u_j tends faster to zero than z_0^{-j}. This completes the verification of the asymptotic expansion (C.8).

The asymptotic expansion (C.8) is very useful for both theoretical and computational purposes. It appears that in many applications the asymptotic expansion for p_j can already be used for relatively small values of j.

Illustration

To illustrate the asymptotic expansion (C.8), consider the *M/G/1* queue in which customers arrive according to a Poisson process with rate λ and the service time

of a customer has a probability distribution function $B(x)$. It is assumed that the offered load $\rho = \lambda\mu$ is smaller than 1, where μ is the mean service time. Denote by $\{p_j,\ j = 0, 1, \ldots\}$ the limiting distribution of the number of customers present at an arbitrary epoch. In Section 4.2 of Chapter 4 it was shown that the generating function $P(z) = \sum_{n=0}^{\infty} p_n z^n$, $|z| \leq 1$, is given by

$$P(z) = (1 - \rho)\frac{1 - \lambda(1 - z)A(z)}{1 - \lambda A(z)}$$

where $A(z) = \sum_{n=0}^{\infty} a_n z^n$, $|z| \leq 1$, with $a_n = (1/n!) \int_0^{\infty} \{1 - B(t)\} e^{-\lambda t} (\lambda t)^n \, dt$. Note that $A(z)$ can be written as

$$A(z) = \int_0^{\infty} e^{-\lambda(1-z)t} \{1 - B(t)\} \, dt. \tag{C.11}$$

The generating function $P(z)$ can indeed be represented by the ratio of two functions $N(z)$ and $D(z)$ with $N(z) = (1 - \rho)[1 - \lambda(1 - z)A(z)]$ and $D(z) = 1 - \lambda A(z)$. The domain of definition of $N(z)$ and $D(z)$ can be extended outside the unit circle when the following assumptions are made:

(i) $\int_0^{\infty} e^{st} \{1 - B(t)\} \, dt < \infty$ for some $s > 0$,

(ii) $\lim_{s \to B} \int_0^{\infty} e^{st} \{1 - B(t)\} \, dt = \infty$, where $B = \sup\{s | \int_0^{\infty} e^{st} \{1 - B(t)\} \, dt < \infty\}$.

The assumption (i) requires that the service-time distribution has no extremely long tail. This assumption is satisfied in most applications. The other assumption is a technical one. Let $R > 1$ be defined by

$$R = 1 + \frac{B}{\lambda}.$$

It follows from the representation (C.11) that the function $A(z)$ has an analytic continuation to the region $|z| < R$. This implies that $N(z)$ and $D(z)$ are analytic functions for $|z| < R$. It now holds that

$$p_j \approx \frac{(1 - \rho)}{\lambda^2} \left[\int_0^{\infty} t e^{-\lambda(1 - z_0)t} \{1 - B(t)\} \, dt \right]^{-1} z_0^{-j} \quad \text{for } j \text{ large enough} \tag{C.12}$$

where z_0 is the unique root of the equation

$$\int_0^{\infty} e^{-\lambda(1-z)t} \{1 - B(t)\} \, dt = \frac{1}{\lambda}$$

on the interval $(1, R)$.

For completeness we give a proof of (C.12). This proof is representative for other applications as well. The reader who is only interested in the result (C.12) may skip the proof. To verify that the conditions C1–C3 are satisfied, we first note