6 $M/M/1$ type models

In this chapter we consider $M/M/1$ type models, more commonly known as *quasi birth-death processes*. We will present two methods for analyzing the equilibrium behavior of $M/M/1$ type models: the matrix-geometric method and the spectral expansion method.

6.1 Model

We consider a Markov process, the state space of which consists of two parts: the *boundary states* $(0, j)$ where j ranges from 0 to n, and a semi infinite strip of states (i, j) where i ranges from 1 to ∞ and j from 0 to m. The states are ordered lexicographically, that is, $(0, 0), (0, 1), \ldots, (0, n), (1, 0), \ldots, (1, m), (2, 0), \ldots, (2, m), \ldots$. The set of boundary states $\{(0, 0), (0, 1), \ldots, (0, m)\}$ will be called *level* 0, and the set of states $\{(i, 0), (i, 1), \ldots, (i, n)\}$, $i \geq 1$, will be called *level* i. Note that the number of states at level 0 may be different from the number of states at higher levels (and this is typically the case in many problems). A picture of the state space is given in figure 1.

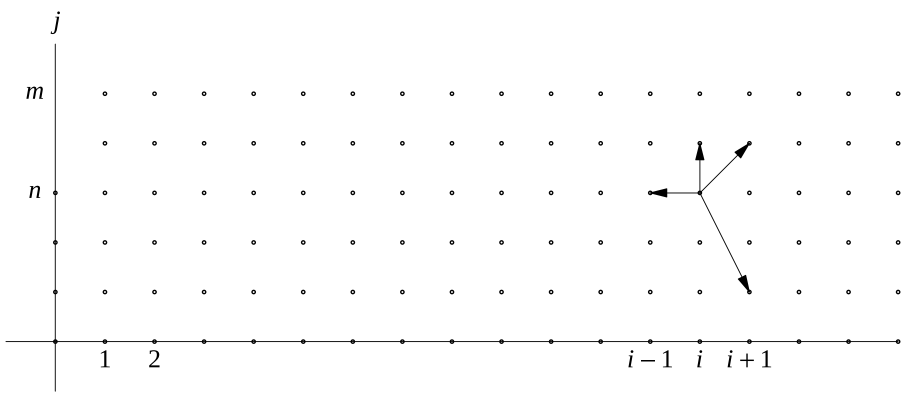

Figure 1: State space of $M/M/1$ type model

We partition the state space according to these levels, and for this partitioning we assume that the generator Q is of the form

$$Q = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & 0 & \ldots \\ B_{10} & B_{11} & A_0 & 0 & 0 & \ldots \\ 0 & A_2 & A_1 & A_0 & 0 & \ldots \\ 0 & 0 & A_2 & A_1 & A_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

where the matrix B_{00} is of dimension $(n+1) \times (n+1)$, B_{01} of dimension $(n+1) \times (m+1)$, B_{10} of dimension $(m + 1) \times (n + 1)$, and B_{11}, A_0, A_1, A_2 are square matrices of dimension $m + 1$. Note that $A_0 + A_1 + A_2$ is a generator; it describes the behavior of the Markov process Q in the (vertical) j-direction only.

Example 6.1 For the problem in section 5.1 (machine with set-up times) we have that level 0 is $\{(0,0)\}$ and level i is the pair of states $\{(i,0),(i,1)\}$; so $n = 0$ and $m = 1$. Further, the matrices of transition rates are given by

$$B_{00} = \begin{pmatrix} -\lambda \end{pmatrix}, \quad B_{01} = \begin{pmatrix} -\lambda & 0 \end{pmatrix}, \quad B_{10} = \begin{pmatrix} 0 \\ \mu \end{pmatrix},$$

$$A_0 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad B_{11} = A_1 = \begin{pmatrix} -(\lambda+\theta) & \theta \\ 0 & -(\lambda+\mu) \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}.$$

Note that

$$A_0 + A_1 + A_2 = \begin{pmatrix} -\theta & \theta \\ 0 & 0 \end{pmatrix};$$

so state 1 is an absorbing state.

Example 6.2 For the example in section 5.2 (unreliable machine) level i is the set of states $\{(i,0),(i,1)\}$, $i \geq 0$ (so $n = m = 1$). The transition matrices are given by

$$B_{00} = \begin{pmatrix} -(\lambda+\theta) & \theta \\ \eta & -(\lambda+\eta) \end{pmatrix},$$

$$A_0 = B_{01} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} -(\lambda+\theta) & \theta \\ \eta & -(\lambda+\mu+\eta) \end{pmatrix}, \quad A_2 = B_{10} = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix},$$

and the generator $A_0 + A_1 + A_2$ is equal to

$$A_0 + A_1 + A_2 = \begin{pmatrix} -\theta & \theta \\ \eta & -\eta \end{pmatrix}.$$

From here on we will assume that the Markov process Q is irreducible and that the generator $A_0 + A_1 + A_2$ has exactly one communicating class. Concerning the stability of Q we state the following result.

Theorem 6.3 *The Markov process Q is ergodic (stable) if and only if*

$$\pi A_0 e < \pi A_2 e, \tag{1}$$

where e is the column vector of ones and $\pi = (\pi_0, \pi_1, \ldots, \pi_m)$ is the equilibrium distribution of the Markov process with generator $A_0 + A_1 + A_2$; so

$$\pi(A_0 + A_1 + A_2) = 0, \qquad \pi e = 1.$$

Condition (1) has an appealing intuitive interpretation. The term $\pi A_0 e$ is the *mean drift* from level i to level $i+1$, and $\pi A_2 e$ is the mean drift from level $i+1$ to level i; clearly the process is stable if the drift to the left is greater than the drift to the right (cf. the $M/M/1$ model where the drift to the right is λ and the drift to the left μ). Condition (1) is known as Neuts' *mean drift condition*. For a rigorous proof of theorem 6.3 we refer the reader to [4].

Example 6.4 For the example in section 5.2 (unreliable machine) condition (1) reduces to

$$(\pi_0 + \pi_1)\lambda < \pi_1\mu,$$

where $\pi = (\pi_0, \pi_1)$ is the equilibrium distribution of

$$A_0 + A_1 + A_2 = \begin{pmatrix} -\theta & \theta \\ \eta & -\eta \end{pmatrix}.$$

Hence,

$$\pi_0 = \frac{\eta}{\theta + \eta}, \qquad \pi_1 = \frac{\theta}{\theta + \eta},$$

and thus the stability condition becomes (cf. (15) in section 5.2)

$$\frac{\lambda}{\mu} < \pi_1 = \frac{\theta}{\theta + \eta} = \rho_U.$$

In the sequel we will assume that the Markov process Q is ergodic. Thus the equilibrium probabilities $p(i, j)$ exist. In the following sections we will present methods to find these probabilities.

6.2 The matrix-geometric method

For an elegant treatment of matrix-geometric solutions the reader is referred to [4, 2]. In this section we just state some of the main results.

Provided the Markov process Q is ergodic, the equilibrium probability vectors p_i are given by the matrix-geometric form

$$p_i = (p(i, 0), p(i, 1), \ldots, p(i, m)) = p_1 R^{i-1}, \qquad i = 1, 2, \ldots, \tag{2}$$

where the matrix R is the *minimal nonnegative solution* of the matrix-quadratic equation

$$A_0 + RA_1 + R^2 A_2 = 0. \tag{3}$$

That is, any other nonnegative solution $\tilde{R}$ of the above matrix equation satisfies $R \leq \tilde{R}$. The matrix R, usually called the rate matrix of the markov process Q, has spectral radius less than one (so $I - R$ is invertable). Note that, if p_i is of the matrix-geometric form (2), then substitution of (2) into the equilibrium equations

$$p_{i-1}A_0 + p_i A_1 + p_{i+1}A_2 = 0,$$

yields

$$p_1 R^{i-2}(A_0 + RA_1 + R^2 A_2) = 0.$$

So indeed, if R is a solution to (3), then the form (2) satisfies the equilibrium equations.

The equilibrium equations for the probability vectors p_0 and p_1 are given by

$$p_0 B_{00} + p_1 B_{10} = 0, \tag{4}$$
$$p_0 B_{01} + p_1 B_{11} + p_2 A_2 = 0. \tag{5}$$

Hence, substituting $p_2 = p_1 R$ we get the following boundary equations for p_0 and p_1,

$$\begin{aligned} p_0 B_{00} + p_1 B_{10} &= 0, \\ p_0 B_{01} + p_1 B_{11} + p_1 R A_2 &= 0. \end{aligned}$$

To uniquely determine p_0 and p_1 we further need the normalization equation

$$1 = \sum_{i=0}^{\infty} p_i e = p_0 e + p_1 (I + R + R^2 + \cdots) e = p_0 e + p_1 (I - R)^{-1}.$$

For the computation of the matrix R we may rewrite (3) in the form

$$R = -(A_0 + R^2 A_2) A_1^{-1};$$

note that A_1 is indeed invertable, since A_1 is a transient generator. The above (fixed point) equation may be solved by successive substitutions, so

$$R_{k+1} = -(A_0 + R_k^2 A_2) A_1^{-1}, \qquad k = 0, 1, 2, \ldots \tag{6}$$

starting with $R_0 = 0$. It can be shown that, as k tends to infinity,

$$R_k \uparrow R.$$

This is a very simple scheme for the computation of R; in the literature more sophisticated and efficient schemes have been developed, see, e.g., [1, 2].

The rate matrix R also has an interesting (and useful) probabilistic interpretation. The element R_{jk} is the expected time spent in state $(i+1, k)$ before the first return to level i, expressed in time unit $-1/(A_1)_{jj}$, given the initial state (i, j). Note that $-1/(A_1)_{jj}$ is the expected sojourn time in state (i, j) with $i > 1$. From the interpretation of R we may directly conclude that zero rows in A_0 correspond to zero rows in R.

Remark 6.5 The matrix-geometric analysis for (discrete-time) Markov chains is very similar. Suppose the transition probability matrix P is of the form

$$P = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & 0 & \ldots \\ B_{10} & B_{11} & A_0 & 0 & 0 & \ldots \\ 0 & A_2 & A_1 & A_0 & 0 & \ldots \\ 0 & 0 & A_2 & A_1 & A_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

where B_{00}, B_{01}, B_{10}, B_{11}, A_0, A_1 and A_2 are now matrices with transition probabilities (instead of transition rates). The Markov chain P is ergodic if and only if drift condition (1) holds, where π is the equilibrium distribution of the Markov chain with transition probability matrix $A_0 + A_1 + A_2$. Further, if P is ergodic, then the probability vectors p_i have the form (2), where R is the minimal nonnegative solution of

$$A_0 + RA_1 + R^2A_2 = R.$$

The matrix element R_{jk} can now be interpreted as the expected number of visits to state $(i+1, k)$ before the first return to level i, given the initial state (i, j).

In fact, the result for discrete-time Markov chains P can be directly translated to the result for continuous-time Markov chains Q by observing that $I + \Delta Q$ is a probability transition matrix for sufficiently small $\Delta \geq 0$.

6.3 Explicit solutions for the rate matrix

In this section we briefly describe two cases in which the rate matrix R can be determined explicitly; for more details the reader is referred to [5]. Let us first assume that A_2 is of the form

$$A_2 = v \cdot \alpha, \tag{7}$$

where v is column vector and α a row vector of dimension $m+1$, with $\alpha e = 1$, so

$$v = \begin{pmatrix} v_0 \\ \vdots \\ v_m \end{pmatrix}, \qquad \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_m), \qquad \alpha e = 1.$$

This means that all rows of A_2 are the same, except for scaling. Thus, when the process Q jumps from level i to level $i-1$, the probability of jumping to state $(i-1, j)$ is *independent* of the starting state at level i. We will investigate its consequences for R.

Substitution of (7) into the equilibrium equations for p_i, $i > 1$, yields

$$p_{i-1}A_0 + p_iA_1 + p_{i+1}v\alpha = 0. \tag{8}$$

To eliminate p_{i+1} from this equation we derive a relation between p_i and p_{i+1} by equating the flow between level i and level $i+1$, i.e.,

$$p_iA_0e = p_{i+1}A_2e = p_{i+1}v\alpha e = p_{i+1}v, \tag{9}$$

Hence, by substituting (9) into (8), we obtain

$$p_{i-1}A_0 + p_iA_1 + p_iA_0e\alpha = 0,$$

which can be rewritten as

$$p_i = p_{i-1}R, \qquad i > 1,$$

where

$$R = -A_0(A_1 + A_0 e\alpha)^{-1}.$$

Note that the matrix $A_1 + A_0 e\alpha$ is invertable, since it is a transient generator.

The other case in which we can solve R explicitly is when A_0 is of the form

$$A_0 = w \cdot \beta,$$

where w is column vector and β a row vector of dimension $m+1$, with $\beta e = 1$, so

$$w = \begin{pmatrix} w_0 \\ \vdots \\ w_m \end{pmatrix}, \qquad \beta = (\beta_0, \beta_1, \ldots, \beta_m), \qquad \beta e = 1.$$

This means that all rows of A_0 are the same, except for scaling. Thus, when the process Q jumps from level i to level $i+1$, the probability of jumping to state $(i+1, j)$ is *independent* of the starting state at level i. Below we investigate the implications of this special form of A_0.

From the recursive scheme (6) we obtain

$$R_0 = 0, \quad R_1 = -A_0 A_1^{-1} = -w\beta A_1^{-1} = w \cdot a_1,$$

with $a_1 = -\beta A_1^{-1}$. Repeating the iteration shows that all R_k's are of the form

$$R_k = w \cdot a_k,$$

where a_k is a row vector of dimension $m+1$. Since $R_k \uparrow R$ as $k \to \infty$, we can conclude that also R is of the form

$$R = w \cdot a,$$

for some row vector $a \geq 0$. Hence,

$$R^i = (aw)^{i-1} R = \eta^{i-1} R,$$

where $\eta = aw$. Clearly η is the spectral radius of R. Now the matrix-geometric form of the probability vectors p_i reduces to

$$p_i = p_1 R^{i-1} = \eta^{i-2} p_1 R = \eta^{i-2} p_2, \qquad i > 1.$$

Note that η can be characterized as the unique root in $(0,1)$ of the determinantal equation

$$\det(A_0 + \eta A_1 + \eta^2 A_2) = 0,$$

which may be computed by straightforward bisection.

6.4 Spectral expansion method

In this section we describe the spectral expansion method; for more details the reader is referred to [3].

The basic idea of the method is to first try to find *basis solutions* of the form

$$p_i = yx^{i-1}, \qquad i = 1, 2, \ldots, \tag{10}$$

where $y = (y(0), y(1), \ldots, y(m)) \neq 0$ and $|x| < 1$, satisfying the equilibrium equations for the levels $i > 1$, i.e.,

$$p_{i-1}A_0 + p_iA_1 + p_{i+1}A_2 = 0. \tag{11}$$

We require that $|x| < 1$, since we want to be able to normalize the solution of (11). Then the basis solutions will be linearly combined so as to also satisfy the equilibrium equations for the boundary states (i.e., levels 0 and 1).

Substitution of (10) into (11) and dividing by common powers of x yields

$$y(A_0 + xA_1 + x^2A_2) = 0. \tag{12}$$

These equations have a non-null solution for y if

$$\det(A_0 + xA_1 + x^2A_2) = 0. \tag{13}$$

Hence, the desired values of x are the roots x with $|x| < 1$ of the determinantal equation (13). Equation (13) is a polynomial equation of degree $2(m+1)$. Hence it has $2(m+1)$ (possibly complex) roots. Provided the Markov process Q is ergodic, there exist exactly $m+1$ roots x with $|x| < 1$ (where each root is counted according to its multiplicity); this number will appear to be exactly enough to satisfy the boundary equations. Let us assume that these $m+1$ roots are different, and label them as $x_0, x_1, \ldots, x_m$. Let y_j be a non-null solution of (12) with $x = x_j$, $j = 0, 1, \ldots, m$. We then set

$$p_i = \sum_{j=0}^{m} c_j y_j x^{i-1}, \qquad i = 1, 2, \ldots \tag{14}$$

Expression (14) is usually referred to as *the spectral expansion of the equilibrium probability vectors* p_i. The coefficients c_j of this expansion have to be determined yet. Note that, since the equilibrium equations are linear, the expansion (14) satisfies the equilibrium equations for the levels $i > 1$ for any choice of the coefficients $c_0, c_1, \ldots, c_m$.

Substituting the spectral expansion for p_1 and p_2 into (4)-(5) we get the following set of equations for the coefficients $c_0, \ldots, c_m$ and the vector p_0,

$$\begin{aligned} p_0B_{00} + \sum_{j=0}^{m} c_j y_j B_{10} &= 0, \\ p_0B_{01} + \sum_{j=0}^{m} c_j y_j B_{11} + \sum_{j=0}^{m} c_j y_j x_j A_2 &= 0. \end{aligned}$$

Together with the normalization equation

$$1 = p_0 e + \sum_{j=0}^{m} c_j y_j e \frac{1}{1 - x_j},$$

this set of equations uniquely determines p_0 and $c_0, \ldots, c_m$.

Remark 6.6 The roots $x_0, x_1, \ldots, x_m$ do not have to be different. If we assume that, when a root x occurs k times, it is possible to find k linearly independent solutions of (12), then the analysis proceeds in exactly the same way. In case there are less than k independent solutions, we would also have to consider more complicated basis solutions of the form yix^{i-1} (or even with higher powers of i).

Remark 6.7 The relation between the matrix-geometric representation (2) and the spectral expansion (14) for the equilibrium probability vectors p_i is clear: the roots $x_0, x_1, \ldots, x_m$ are the eigenvalues of R with corresponding left eigenvectors $y_0, y_1, \ldots, y_m$.

References

[1] G. Latouche and V. Ramaswami, *A logarithmic reduction algorithm for quasi-birth-death processes*. J. Appl. Prob., 30 (1993), pp. 650–674.

[2] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM, 1999

[3] I. Mitrani and D. Mitra, A spectral expansion method for random walks on semi-infinite strips, in: R. Beauwens and P. de Groen (eds.), *Iterative methods in linear algebra*, North-Holland, Amsterdam (1992), 141–149.

[4] M.F. Neuts, *Matrix-geometric solutions in stochastic models*. The John Hopkins University Press, Baltimore, 1981.

[5] V. Ramaswami and G. Latouche, *A general class of Markov processes with explicit matrix-geometric solutions*. OR Spektrum, 8 (1986), pp. 209–218.