Box 31.2 M/M/m Queue

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
   m = number of servers
2. Traffic intensity: $\rho = \lambda/(m\mu)$
3. The system is stable if the traffic intensity ρ is less than 1.
4. Probability of zero jobs in the system:

$$p_0 = \left[1 + \frac{(m\rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!}\right]^{-1}$$

5. Probability of n jobs in the system:

$$p_n = \begin{cases} p_0 \dfrac{(m\rho)^n}{n!}, & n < m \\ p_0 \dfrac{\rho^n m^m}{m!}, & n \geq m \end{cases}$$

6. Probability of queueing:

$$\varrho = P(\geq m \text{ jobs}) = \frac{(m\rho)^m}{m!(1-\rho)} p_0$$

   In the remaining formulas below we will use ϱ as defined here.
7. Mean number of jobs in the system: $E[n] = m\rho + \rho\varrho/(1-\rho)$
8. Variance of number of jobs in the system:

$$\text{Var}[n] = m\rho + \rho\varrho\left[\frac{1+\rho-\rho\varrho}{(1-\rho)^2} + m\right]$$

9. Mean number of jobs in the queue: $E[n_q] = \rho\varrho/(1-\rho)$
10. Variance of number of jobs in the queue:
    $\text{Var}[n_q] = \varrho\rho(1+\rho-\varrho\rho)/(1-\rho)^2$
11. Average utilization of each server: $U = \lambda/(m\mu) = \rho$
12. Cumulative distribution function of response time:

$$F(r) = \begin{cases} 1 - e^{-\mu r} - \dfrac{\varrho}{1-m+m\rho} e^{-m\mu(1-\rho)r} - e^{-\mu r}, & \rho \neq (m-1)/m \\ 1 - e^{-\mu r} - \varrho\mu r e^{-\mu r}, & \rho = (m-1)/m \end{cases} \quad r > 0$$

Box 31.2 Continued

13. Mean response time:

$$E[r] = \frac{1}{\mu}\left(1 + \frac{\varrho}{m(1-\rho)}\right)$$

14. Variance of the response time:

$$\mathrm{Var}[r] = \frac{1}{\mu^2}\left[1 + \frac{\varrho(2-\varrho)}{m^2(1-\rho)^2}\right]$$

15. Cumulative distribution function of waiting time: $F(w) = 1 - \varrho e^{-m\mu(1-\rho)w}$
16. Mean waiting time: $E[w] = E[n_q]/\lambda = \varrho/[m\mu(1-\rho)]$
17. Variance of the waiting time: $\mathrm{Var}[w] = \varrho(2-\varrho)/[m^2\mu^2(1-\rho)^2]$
18. q-Percentile of the waiting time: $\max\left(0, \frac{E[w]}{\varrho}\ln\frac{100\varrho}{100-q}\right)$.
19. 90-Percentile of the waiting time: $\frac{E[w]}{\varrho}\ln(10\varrho)$

Once again, ϱ in these formulas is the probability of m or more jobs in the system: $\varrho = [(m\rho)^m/\{m!(1-\rho)\}]p_0$. For $m = 1$, ϱ is equal to ρ and all of the formulas become identical to those for M/M/1 queues.