Box 31.1 M/M/1 Queue

1. Parameters:
   λ = arrival rate in jobs per unit time
   μ = service rate in jobs per unit time
2. Traffic intensity: $\rho = \lambda/\mu$ = utilization
3. Stability condition: Traffic intensity ρ must be less than 1.
4. Probability of zero jobs in the system: $p_0 = 1 - \rho$
5. Probability of n jobs in the system: $p_n = (1-\rho)\rho^n,\ n = 0, 1, \ldots, \infty$
6. Mean number of jobs in the system: $E[n] = \rho/(1-\rho)$
7. Variance of number of jobs in the system: $\mathrm{Var}[n] = \rho/(1-\rho)^2$
8. Probability of k jobs in the queue:

$$P(n_q = k) = \begin{cases} 1-\rho^2, & k = 0 \\ (1-\rho)\rho^{k+1}, & k > 0 \end{cases}$$

9. Mean number of jobs in the queue: $E[n_q] = \rho^2/(1-\rho)$
10. Variance of number of jobs in the queue:
    $\mathrm{Var}[n_q] = \rho^2(1+\rho-\rho^2)/(1-\rho)^2$
11. Cumulative distribution function of the response time:
    $F(r) = 1 - e^{-r\mu(1-\rho)}$
12. Mean response time: $E[r] = (1/\mu)/(1-\rho)$
13. Variance of the response time: $\mathrm{Var}[r] = \dfrac{1/\mu^2}{(1-\rho)^2}$
14. q-Percentile of the response time: $E[r]\ln[100/(100-q)]$
15. 90-Percentile of the response time: $2.3E[r]$
16. Cumulative distribution function of waiting time:
    $F(w) = 1 - \rho e^{-\mu w(1-\rho)}$
17. Mean waiting time: $E[w] = \rho\dfrac{1/\mu}{1-\rho}$
18. Variance of the waiting time: $\mathrm{Var}[w] = (2-\rho)\rho/[\mu^2(1-\rho)^2]$
19. q-Percentile of the waiting time: $\max\left(0, \dfrac{E[w]}{\rho}\ln[100\rho/(100-q)]\right)$
20. 90-Percentile of the waiting time: $\max\left(0, \dfrac{E[w]}{\rho}\ln[10\rho]\right)$
21. Probability of finding *n or more* jobs in the system: ρ^n
22. Probability of serving n jobs in one busy period:

$$\frac{1}{n}\binom{2n-2}{n-1}\frac{\rho^{n-1}}{(1+\rho)^{2n-1}}$$