1. Modeling Ex. 7.1

Let \( X(t) \) be the number of customers waiting at the taxi stand at time \( t \). This number goes up by 1 whenever a customer arrives, and goes down by one whenever there is at least one customer and a taxi arrives. Thus \( \{X(t), t \geq 0\} \) is a birth and death process with birth rate \( \lambda \) in all states \( i \geq 0 \), and death rates \( \mu \) in all states \( i \geq 1 \). Hence it is an \( M|M|1 \) queue with arrival rate \( \lambda \) and service rate \( \mu \).

2. Modeling Ex. 7.6. Suppose a customer starts service at time 0 and finishes service at time \( S \). This means the server is up at time 0 and \( S \).

Let \( W \sim \text{exp}(\mu) \) be amount of time it takes to service the customer if there are no failures, and \( U \sim \text{exp}(\theta) \) be the first uptime of the server. If \( U > W \) then \( S = \min(W, U) \). If \( U < W \) then the server fails at time \( U \) before the service is completed, then the server stays down for \( R \sim \text{exp}(\alpha) \) amount of time. From then on it again takes \( S \) amount of time to finish the service, due to memoryless property of the exponential distribution. Hence \( S \sim U + R + S \). Using this analysis, we get

\[
\tilde{G}(s) = E(e^{-sS}) = \frac{\mu}{\theta + \mu} \cdot \frac{\theta + \mu}{s + \theta + \mu} + \frac{\theta}{s + \theta + \mu} \cdot \frac{\theta + \mu}{s + \alpha} \cdot \tilde{G}(s).
\]

This yields:

\[
\tilde{G}(s) = \frac{\mu(s + \alpha)}{(s + \theta + \mu)(s + \alpha) - \theta \alpha}.
\]

Since the server is up whenever a new customer enters service for the first time, it is clear that the service times are iid with LST \( \phi(s) \). The arrival process is PP(\( \lambda \)). Hence this is an \( M|G|1 \) queue.

3. Modeling Ex. 7.12. Under the random routing scheme, the Bernoulli splitting of Poisson processes implies that \( \{X_i(t), t \geq 0\} \) is the queue length process in an \( M/M/1 \) queue with inter-arrival times iid exp(\( \lambda/2 \)) and iid exp(\( \mu \)) service times. The two queues are independent. Under the alternate routing scheme, \( \{X_i(t), t \geq 0\} \) is the queue length process in a \( G/M/1 \) queues with inter-arrival times iid with common distribution erl(2, \( \lambda \)) and iid exp(\( \mu \)) service times. The two queues are dependent.

4. Computational Ex. 7.3. Follow the analysis of \( W_n \) in Subsection 7.3.1. We have \( X_n^* = 0 \Leftrightarrow W_n^q = 0 \). Hence, letting \( n \to \infty \),

\[
P(W_n^q = 0) = 1 - \rho.
\]
For $j \geq 1$, we have
\[ P(W^q_n \leq x | X^*_n = j) = 1 - \sum_{r=0}^{j-2} e^{-\mu x} \frac{(\mu x)^r}{r!}. \]
Substituting, we get
\[ F^q(x) = \sum_{j=0}^{\infty} (1 - \rho)^j \rho^j P(W^q_n \leq x | X^*_n = j) \left( 1 - \sum_{r=0}^{j-1} e^{-\mu x} \frac{(\mu x)^r}{r!} \right), \]
which, after some algebra, reduces to
\[ F^q(x) = 1 - \rho e^{-(\mu - \lambda)x}, \quad x \geq 0. \]
The expected value can be calculated as
\[ W^q = \frac{1}{\mu} \frac{\rho}{1 - \rho}. \]
This satisfies $L^q = \lambda W^q$.
5. (a) We must have $\lambda \alpha_k < \mu_k$. Hence the feasible region is
\[ 0 \leq \alpha_k < \frac{\mu_k}{\lambda}, \quad 1 \leq k \leq K, \]
\[ \alpha_1 + \ldots + \alpha_K = 1. \]
(b) The long run cost per unit time for the entire system is given by
\[ c(\alpha) = \sum_{k=1}^{K} h_k \frac{\lambda \alpha_k}{\mu_k - \lambda \alpha_k}. \]
(c) We need to minimize $c(\alpha)$ subject to $\alpha_k < \frac{\mu_k}{\lambda}$ for $1 \leq k \leq K$, and $\alpha_1 + \ldots + \alpha_K = 1$. We ignore the inequality constraints, and use Lagrangian multipliers to solve the constrained optimization. We get the KK conditions as
\[ \frac{\partial c(\alpha)}{\alpha_k} = \frac{\lambda h_k \mu_k}{(\mu_k - \lambda \alpha_k)^2} = a(\text{constant}), \quad 1 \leq k \leq K. \]
This yields
\[ \alpha_k = \sqrt{\frac{\mu_k}{\lambda}} \left( \sqrt{\frac{\mu_k}{\lambda}} - \sqrt{\frac{h_k}{a}} \right). \]
Thus the constant $a$ is chosen to satisfy
\[ \sum_{k=1}^{K} \sqrt{\frac{\mu_k}{\lambda}} \left( \sqrt{\frac{\mu_k}{\lambda}} - \sqrt{\frac{h_k}{a}} \right) = 1. \]
The solution is given by

\[ a = \frac{\lambda}{\mu} \left( \sum_{k=1}^{K} \sqrt{h_k \mu_k} \right)^2. \]

It can be seen that with this \( a \) the resulting \( \alpha_k \) will be automatically feasible.