2905 Queueing Theory and Simulation

PART III: HIGHER DIMENSIONAL AND NON-MARKOVIAN QUEUES

16 Queueing Systems with Two Types of Customers

In this section, we discuss queueing systems with two types of customers. The queueing system has no waiting space. There are two possible cases: **infinite-server case** and **finite-server case**.

16.1 Infinite-Server Queue

Consider the infinite-server queue with two types of customers. The arrival process of customers of type i $(i = 1, 2)$ is Poisson with rate λ_i and their service times are independent, identically distributed, exponential random variables with mean μ_i^{-1} $(i = 1, 2)$.

• We define the 2-dimensional states $\{E_{j_1,j_2}\}$, where j_i is the number of customers of type i in the system, with corresponding equilibrium distribution $\{p(j_1, j_2)\}$, then clearly the Markov property still holds.

• Here $p(j_1, j_2)$ is the steady state probability that there are j_1 **type 1** customers and j_2 **type 2** customers in the system.

• By equating expected **rate out** to expected **rate in** for each state, the equilibrium state equations are

$$\boxed{\begin{aligned}(\lambda_1+\lambda_2+j_1\mu_1+j_2\mu_2)p(j_1,j_2) \;=\; & \lambda_1 p(j_1-1,j_2)+(j_1+1)\mu_1 p(j_1+1,j_2)\\ & +\lambda_2 p(j_1,j_2-1)+(j_2+1)\mu_2 p(j_1,j_2+1)\end{aligned}} \tag{1}$$

and $[p(-1,j_2)=p(j_1,-1)=0\ ;\ j_1=0,1,\ldots\ ;\ j_2=0,1,\ldots.]$

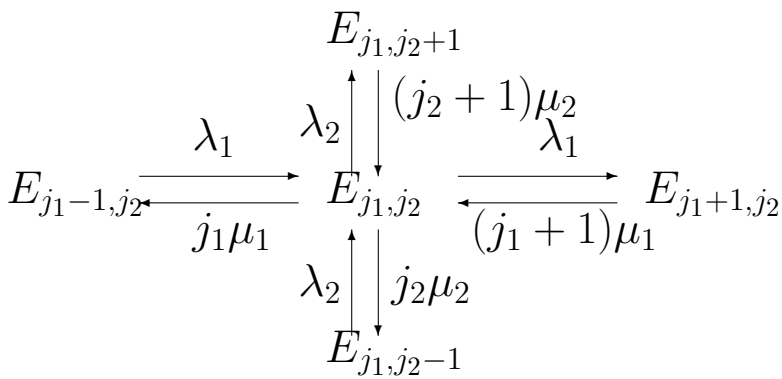


Figure 16.1 The Markov Chain of the System at State E_{j_1,j_2}.

• In addition, the probabilities must satisfy the normalization equation:

$$\boxed{\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} p(j_1,j_2)=1.}$$

• In this case, we already know the answer. Since the number of servers is infinite, the two types of customers do not affect one another.

• Thus the **marginal distribution** of the number of customers of each type is that which would be obtained by solving the corresponding one-dimensional problem, namely the Poisson distribution:

$$\boxed{\begin{cases} p_1(j) &= \displaystyle\sum_{k=0}^{\infty} p(j,k) = \frac{(\lambda_1/\mu_1)^j}{j!} e^{-(\lambda_1/\mu_1)}, \\ p_2(j) &= \displaystyle\sum_{k=0}^{\infty} p(k,j) = \frac{(\lambda_2/\mu_2)^j}{j!} e^{-(\lambda_2/\mu_2)} \ . \end{cases}} \tag{2}$$

• Since the number of customers present of each type is independent of the number present of the other type, therefore

$$\boxed{\begin{aligned} p(j_1,j_2) &= p_1(j_1)p_2(j_2) \\ &= \frac{(\lambda_1/\mu_1)^{j_1}}{j_1!} \frac{(\lambda_2/\mu_2)^{j_2}}{j_2!} e^{-[(\lambda_1/\mu_1)+(\lambda_2/\mu_2)]}. \end{aligned}} \tag{3}$$

16.1.1 Solution of Product Form

The fact that the solution $p(j_1, j_2)$ can be decomposed into **a product of two factors** has enabled us to solve the problem with ease.

• In fact one may try

$$\boxed{p(j_1, j_2) = \frac{(\lambda_1/\mu_1)^{j_1}}{j_1!} \frac{(\lambda_2/\mu_2)^{j_2}}{j_2!} C} \tag{4}$$

where the constant C is determined from the normalization equation. In this case

$$\sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} p(j_1, j_2) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \frac{(\lambda_1/\mu_1)^{j_1}}{j_1!} \frac{(\lambda_2/\mu_2)^{j_2}}{j_2!} C = C e^{\lambda_2/\mu_2} \sum_{j_1=0}^{\infty} \frac{(\lambda_1/\mu_1)^{j_1}}{j_1!} = C e^{[(\lambda_1/\mu_1)+(\lambda_2/\mu_2)]}.$$

• Hence we have the constant

$$C = e^{-[(\lambda_1/\mu_1)+(\lambda_2/\mu_2)]}.$$

• In practice, a good strategy for finding solutions to equations of the form (1) is to assume a product solution of the form (4); and see if such a solution satisfies Eq. (1).

• If it works, then the solution has been obtained. If it doesn't, then try a different approach. In this case it works!

16.2 Multiple-server Queue with Blocked Customers Cleared

The situation is similar to that of the previous example except now the system has finitely many servers.

• The system is again described by Eq. (1), which is valid now only for $j_1 + j_2 < s$.

• When $j_1 + j_2 = s$, then the states E_{j_1+1,j_2} and E_{j_1,j_2+1} cannot occur and the equation becomes

$$\boxed{(j_1\mu_1 + j_2\mu_2)p(j_1, j_2) = \lambda_1 p(j_1 - 1, j_2) + \lambda_2 p(j_2, j_2 - 1).} \tag{5}$$

• Observe that Eq. (5) can be obtained from Eq. (1) by deleting the first two terms on the left and the last two terms on the right. The product-form solution (4) satisfies Eq. (1) and also the equation with only the deleted terms

$$\boxed{(\lambda_1 + \lambda_2)p(j_1, j_2) = (j_1 + 1)\mu_1 p(j_1 + 1, j_2) + (j_2 + 1)\mu_2 p(j_1, j_2 + 1),}$$

therefore it also satisfies Eq. (5). Thus the product solution in Eq. (4) is a solution of this problem.

• In particular, if we don't distinguish the two types of customers, then the probability $p(j)$ that there are j customers (including type 1 and type 2) in service is given by

$$p(j) = \sum_{j_1+j_2=j} p(j_1, j_2) = \sum_{j_1=0}^{j} p(j_1, j - j_1).$$

- With the help of binomial theorem, we have

$$p(j) \;=\; C\sum_{j_1=0}^{j}\frac{(\frac{\lambda_1}{\mu_1})^{j_1}}{j_1!}\frac{(\frac{\lambda_2}{\mu_2})^{j-j_1}}{(j-j_1)!} = \frac{C}{j!}\sum_{j_1=0}^{j}\frac{j!}{j_1!(j-j_1)!}(\frac{\lambda_1}{\mu_1})^{j_1}(\frac{\lambda_2}{\mu_2})^{j-j_1} = C\frac{1}{j!}\Big(\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{\mu_2}\Big)^{j}.$$

- The normalization equation

$$\sum_{k=0}^{s} p(k) = 1$$

implies that

$$C = \left\{\sum_{k=0}^{s}\frac{1}{k!}\left(\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{\mu_2}\right)^{k}\right\}^{-1}.$$

We conclude that

$$p(j) = \frac{a^j/j!}{\sum\limits_{k=0}^{s} a^k/k!} \quad (j=0,1,\ldots,s) \quad \text{where} \quad a = \frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{\mu_2}.$$

- We note that for the case of infinite-server queue (we let $s \to \infty$) we have

$$p(j) = \frac{a^j e^{-a}}{j!} \quad (j=0,1,\ldots,).$$

17 Queues in Tandem

Consider **two** sets of servers arranged **in tandem**, so that the output from the first set of servers is the input of the second set of servers.

• Assume that the arrival process at the first stage of this tandem queueing system is Poisson with rate λ, the service times in the first stage are exponentially distributed with mean μ_1^{-1}, and the queueing discipline is blocked customers delayed.

• The customers completing service in the first stage will enter the second stage (and wait if all servers in second stage are busy), where the service times are assumed to be exponentially distributed with mean μ_2^{-1}. The number of servers in stage i is $s_i (i = 1, 2)$.

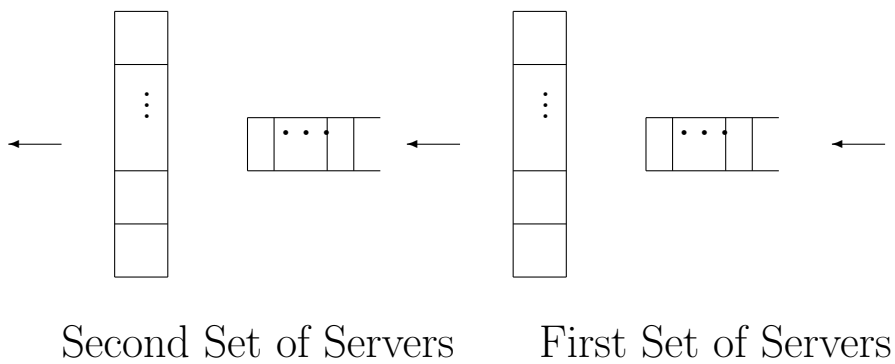


Figure 17.1 Two Queues in Tandem.

• Let $p(j_1, j_2)$ be the probability that there are j_1 customers in stage 1 and j_2 customers in stage 2. Let

$$\mu_i(j) = \begin{cases} j\mu_i & (j = 0, 1, \ldots, s_i) \\ s_i\mu_i & (j = s_i + 1, s_i + 2, \ldots) \end{cases}$$

be the departure rates. Then equating **rate in** to **rate out** for each state, we obtain

$$\begin{aligned} (\lambda + \mu_1(j_1) + \mu_2(j_2))\, p(j_1, j_2) \;&=\; \lambda p(j_1 - 1, j_2) + \mu_1(j_1 + 1)p(j_1 + 1, j_2 - 1) \\ &\quad + \mu_2(j_2 + 1)p(j_1, j_2 + 1) \end{aligned} \tag{6}$$

$$[p(-1, j_2) = p(j_1, -1) = 0 \;;\; j_1 = 0, 1, \ldots \;;\; j_2 = 0, 1, \ldots]$$

• Since the first stage of this system is precisely an **Erlang delay system**, so that the marginal distribution of the number of customers in stage one is given by the **Erlang delay probabilities**.

• Let us try (hopefully it would work) a product solution of the form

$$\boxed{p(j_1, j_2) = p_1(j_1)p_2(j_2)} \tag{7}$$

with the factor $p_1(j_1)$ given by the Erlang delay probabilities:

$$p_1(j_1) = \begin{cases} C_1 \dfrac{(\lambda/\mu_1)^{j_1}}{j_1!} \; (j_1 = 0, 1, \ldots, s_1 - 1) \;, \\ C_1 \dfrac{(\lambda/\mu_1)^{j_1}}{s_1! s_1^{j_1 - s_1}} \; (j_1 = s_1, s_1 + 1, \ldots) \;. \end{cases} \tag{8}$$

• We shall substitute the assumed product solutions (7) and (8) into the equilibrium state equations (6), with the hope that the system of equations will be reduced to a one-dimensional set of equations that can be solved for the remaining factor $p_2(j_2)$. Indeed, (6) is reduced to

$$[\lambda + \mu_2(j_2)]p_2(j_2) = \lambda p_2(j_2 - 1) + \mu_2(j_2 + 1)p_2(j_2 + 1) \tag{9}$$

$$[p_2(-1) = 0\ ; \quad j_2 = 0, 1, \ldots].$$

Eq. (9) is the equilibrium state equations that define the Erlang delay probabilities.

• We conclude that $p_2(j_2)$ is given by

$$\boxed{p_2(j_2) = \begin{cases} C_2 \dfrac{(\lambda/\mu_2)^{j_2}}{j_2!} & (j_2 = 0, 1, 2, \ldots, s_2 - 1)\ , \\ C_2 \dfrac{(\lambda/\mu_2)^{j_2}}{s_2! s_2^{j_2 - s_2}} & (j_2 = s_2, s_2 + 1, \ldots,). \end{cases}} \tag{10}$$

• Using $\sum_{j=0}^{\infty} p_i(j) = 1$ $(i = 1, 2)$, we get

$$C_i = \left(\sum_{k=0}^{s_i - 1} \frac{\rho_i^k}{k!} + \frac{\rho_i^{s_i}}{s_i!(1 - \rho_i/s_i)} \right)^{-1} \quad \text{where} \quad \rho_i = \frac{\lambda}{\mu_i} \quad (i = 1, 2).$$

This equation implies that a proper joint distribution exists only when $\lambda/\mu_1 < s_1$ and $\lambda/\mu_2 < s_2$.

• It is a remarkable result that the number of customers in the second stage also follows the Erlang delay distribution, that is, the distribution of customers in the second stage is the same as if the first stage were not present and the customers arrived according to a Poisson process directly at the second stage.

• It also suggests that the output from the first stage is a Poisson process. This is in fact true in general and we state without proof the Burke's theorem as follows.

Proposition 1 (Burke's Theorem) The statistical equilibrium output of an $M/M/s$ queue with arrival rate λ and mean service times μ^{-1} is a Poisson process with the same rate λ.

Remark: In 1956 Burke showed that the departure process of an M/M/s queue was Poisson. [P. Burke, (1956) The Output of a Queueing System, *Oper. Res.* (4) 699-704.]

• In the following, we are going to prove the Burke's Theorem.

We consider an M/M/s queue in steady state. Let λ be the arrival rate of customers and μ be the service rate and

$$q_i \ (i = 0, 1, 2, \ldots)$$

be the steady state probability that there are j customers in the system. We recall that q_i can be obtained by using the Erlang's delay formula.

Let τ be the inter-departure interval and let $N(t)$ be the state of the system at time t after a departure (so that there have been no further departure by time t) and we define

$$S_n(t) = P(N(t) = n \text{ and } \tau > t).$$

- We note that

$$\boxed{\begin{array}{rcl} S_0(t+h) & = & P(\text{no arrival of customer in } (t, t+h)) \times S_0(t) = (1 - \lambda h + o(h))S_0(t); \\ S_n(t+h) & = & P(\text{an arrival of customer in } (t, t+h)) \times S_{n-1}(t) \\ & & +P(\text{no arrival of customer in } (t, t+h)) \times S_n(t) \\ & = & (\lambda h + o(h))(1 - n\mu h + o(h))S_{n-1}(t) + (1 - \lambda h + o(h))(1 - n\mu h + o(h))S_n(t), \quad n \leq s; \\ S_n(t+h) & = & P(\text{an arrival of customer in } (t, t+h)) \times S_{n-1}(t) \\ & & +P(\text{no arrival of customer in } (t, t+h)) \times S_n(t) \\ & = & (\lambda h + o(h))(1 - s\mu h + o(h))S_{n-1}(t) + (1 - \lambda h + o(h))(1 - s\mu h + o(h))S_n(t), \quad n > s. \end{array}}$$

- By re-arranging the terms, we have the following equations:

$$\begin{cases} \dfrac{S_0(t+h) - S_0(t)}{h} = -\lambda S_0(t) + \dfrac{o(h)}{h}; \\ \dfrac{S_n(t+h) - S_n(t)}{h} = \lambda S_{n-1}(t) - (\lambda + n\mu)S_n(t) + \dfrac{o(h)}{h}, \quad n \leq s; \\ \dfrac{S_n(t+h) - S_n(t)}{h} = \lambda S_{n-1}(t) - (\lambda + s\mu)S_n(t) + \dfrac{o(h)}{h}, \quad n > s. \end{cases}$$

- By letting $h \to 0$ one can obtain the following differential equations.

$$\begin{cases} \dfrac{dS_0(t)}{dt} = -\lambda S_0(t); \\ \dfrac{dS_n(t)}{dt} = \lambda S_{n-1}(t) - (\lambda + n\mu) S_n(t) \quad n \leq s; \\ \dfrac{dS_n(t)}{dt} = \lambda S_{n-1}(t) - (\lambda + s\mu) S_n(t) \quad n > s. \end{cases} \tag{11}$$

- We note that

$$S_n(0) = q_n \quad \text{for} \quad n = 0, 1, \ldots$$

This is because we assume that the system is in steady state.

- Moreover, we have

$$\boxed{q_n = \frac{\lambda}{\min\{n, s\}\mu} q_{n-1}.}$$

- By solving the differential equations inductively one may obtain

$$\boxed{S_n(t) = q_n e^{-\lambda t}.}$$

• Alternative one may compute

$$S_n'(t) = -q_n \lambda e^{-\lambda t}.$$

We then have for $n \leq s$

$$\begin{aligned}
\lambda S_{n-1}(t) - (\lambda + n\mu) S_n(t) &= \lambda q_{n-1} e^{-\lambda t} - (\lambda + n\mu) q_n e^{-\lambda t} \\
&= (\lambda q_{n-1} - (\lambda + n\mu) q_n) e^{-\lambda t} \\
&= (\lambda q_{n-1} - n\mu q_n) e^{-\lambda t} - \lambda q_n e^{-\lambda t} \\
&= 0 + S_n'(t).
\end{aligned}$$

For $n > s$ we have

$$\begin{aligned}
\lambda S_{n-1}(t) - (\lambda + s\mu) S_n(t) &= \lambda q_{n-1} e^{-\lambda t} - (\lambda + s\mu) q_n e^{-\lambda t} \\
&= (\lambda q_{n-1} - (\lambda + s\mu) q_n) e^{-\lambda t} \\
&= (\lambda q_{n-1} - s\mu q_n) e^{-\lambda t} - \lambda q_n e^{-\lambda t} \\
&= 0 + S_n'(t).
\end{aligned}$$

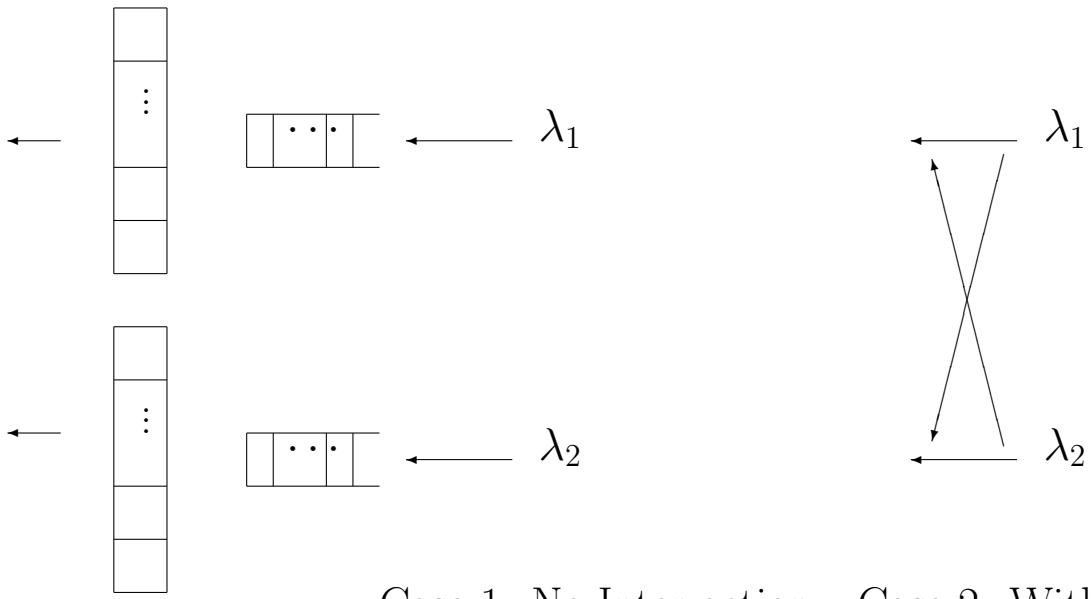
• We then have

$$\boxed{P(\tau > t) = \sum_{n=0}^{\infty} S_n(t) = e^{-\lambda t} \sum_{n=0}^{\infty} q_n = e^{-\lambda t}.}$$

This implies that the p.d.f. of τ is given by $\lambda e^{-\lambda t}$, i.e. the inter-departure time is exponentially distributed with rate λ.

• Finally this implies that the departure process of an M/M/s queue in steady state is a Poisson process with the same rate of the input process.

18 Queues in Parallel



Case 1: No Inter-action Case 2: With Overflow

Figure 18.1 Two Queues in Parallel.

Case (i) We assume there is no inter-action between the two queues. There are $s_i (i = 1, 2)$ servers in queue $i (i = 1, 2)$ and there is infinite many waiting spaces in each queue and we assume blocked customers are delayed. The arrival rate of queue i is λ_i and the service completion rate of a server in queue i is μ_i.

Hence we have the steady state probability that queue i has j customers given by

$$p_i(j) = \begin{cases} \frac{\rho_i^j}{j!} C_i & (j = 0, 1, \cdots, s_i) \\ \frac{\rho_i^j}{s_i! s_i^{j-s_i}} C_i & (j = s_i + 1, \ldots) \end{cases}$$

where $\rho_i = \lambda_i / \mu_i$. Moreover if $\rho_i < s_i$ then

$$C_i = \left(\sum_{k=0}^{s_i - 1} \frac{\rho_i^k}{k!} + \sum_{k=s_i}^{\infty} \frac{\rho_i^k}{s_i! s_i^{k-s_i}} \right)^{-1} = \left(\sum_{k=0}^{s_i - 1} \frac{\rho_i^k}{k!} + \frac{\rho_i^{s_i}}{(s_i - 1)!(s_i - \rho_i)} \right)^{-1}.$$

• If $\rho_i \geq s_i$, the infinite geometric sum diverges then $C_i = 0$ and hence $p_i(j) = 0$ for all finite j. If $\rho_i < s_i$ then the steady state probability that there are i customers in Queue 1 and j customers in Queue 2 is given by

$$\boxed{p(i, j) = p_1(i) p_2(j) \quad i, j = 0, 1, 2, \ldots.}$$

Case (ii) There are $s_i (i = 1, 2)$ servers in queue $i (i = 1, 2)$ and there are finite many waiting spaces in each queue and we assume blocked customers are cleared. The arrival rate of queue i is λ_i and the service completion rate of a server in queue i is μ_i.

• We assume there is **inter-action** between the two queues as follows. Whenever Queue 2(1) is full, an arrived customer of type 2(1) will overflow to Queue 1(2) provided that Queue 1(2) is not yet full. Let us consider a simple example as follows. We assume that Queue 1 and 2 are M/M/1/1 queue. The following figure gives the Markov chain of the queueing system.

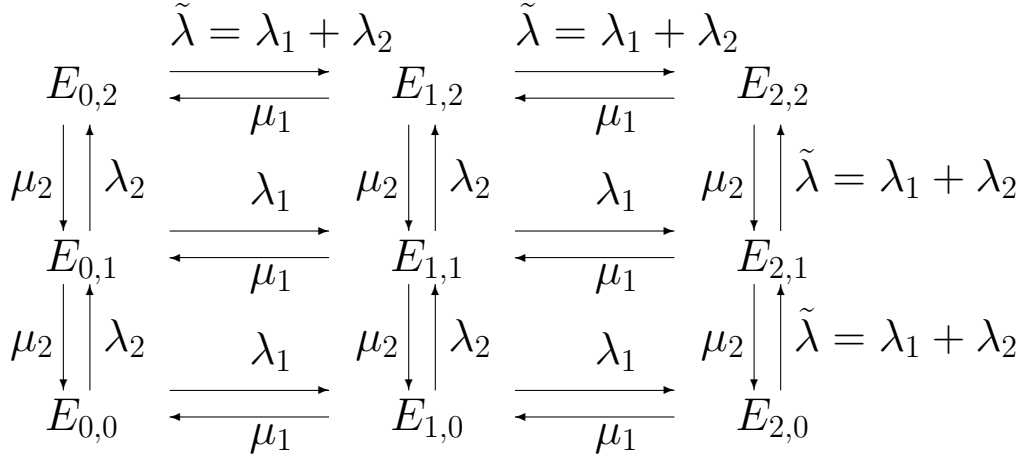


Figure 17.2 The Markov Chain of the Two Queue Overflow System.

- The generator matrix for this queueing problem is given by

$$A_8 = \begin{array}{c} (0,0) \\ (1,0) \\ (2,0) \\ (0,1) \\ (1,1) \\ (2,1) \\ (0,2) \\ (1,2) \\ (2,2) \end{array} \left(\begin{array}{ccc|ccc|ccc} * & \mu_1 & 0 & \mu_2 & 0 & 0 & 0 & 0 & 0 \\ \lambda_1 & * & \mu_1 & 0 & \mu_2 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & * & 0 & 0 & \mu_2 & 0 & 0 & 0 \\ \hline \lambda_2 & 0 & 0 & * & \mu_1 & 0 & \mu_2 & 0 & 0 \\ 0 & \lambda_2 & 0 & \lambda_1 & * & \mu_1 & 0 & \mu_2 & 0 \\ 0 & 0 & \tilde{\lambda} & 0 & \lambda_1 & * & 0 & 0 & \mu_2 \\ \hline 0 & 0 & 0 & \lambda_2 & 0 & 0 & * & \mu_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & \tilde{\lambda} & * & \mu_1 \\ 0 & 0 & 0 & 0 & 0 & \tilde{\lambda} & 0 & \tilde{\lambda} & * \end{array} \right). \tag{12}$$

Here $*$ is such that the column sum is zero. Unfortunately there is no analytic solution for the steady state probabilities of this system. Direct method or numerical method are common methods for solving the steady state probabilities.

19 The M/G/1 Queue: Imbedded Markov Chain Technique

In the preceding sections we will concentrate mainly on queues with Poisson input and exponential service times.

• These assumptions imply that the future evolution of the system from some time t depends only on the state of the system at time t, and is independent of the history of the system prior to time t (i.e. a Markov process).

• In fact, the systems that we have considered are birth-and-death processes, which are special cases of Markov processes.

• In this section, we shall consider models which **do not** constitute Markov processes. One of the most powerful methods for the analysis of certain queueing models is the **Imbedded Markov Chain** technique (IMC).

• We shall concentrate on queueing models, to which the IMC. analysis is applicable. We shall assume, unless specified otherwise, that all queues are BCD and customers are served in order of arrival.

19.1 Imbedded Markov Chain

Consider a general queueing system. Let $N(t)$ be the number of customers in the system at time t. Then, by the law of total probability,

$$P\{N(t+x)=j\} = \sum_{i=0}^{\infty} P\{N(t+x)=j|N(t)=i\}P\{N(t)=i\} \quad (t \geq 0, x \geq 0; j = 0, 1, \ldots). \tag{13}$$

- To obtain the state probabilities, all that is required is to evaluate the conditional probabilities

$$P\{N(t+x)=j|N(t)=i\}$$

called the **transition probabilities**, and then solve Eq. (13). For a birth-and-death process, the transition probabilities can be easily obtained by choosing h small and letting $h \to 0$;

$$P\{N(t+h)=j|N(t)=i\} = \begin{cases} \lambda_i h + o(h) & j = i+1 \\ \mu_i h + o(h) & j = i-1 \\ 1-(\lambda_i+\mu_i)h + o(h) & j = i \\ o(h) & \text{otherwise.} \end{cases} \tag{14}$$

With the transition probabilities Eq. (14), the state probabilities can then be derived:

$$p_j(t+h) = \lambda_{j-1} h p_{j-1}(t) + \mu_{j+1} h p_{j+1}(t) + [1-(\lambda_j+\mu_j)h]p_j(t) + o(h).$$

• But for a M/G/1 queue (and many other queueing models) the transition probabilities depend not only on the state of the system at time t but also on the past history of the system (such as how long a server has been serving a particular customer); that is, the system does not have the Markov property at arbitrary time t.

• However, there are time epochs at which the Markov property holds. Such a point is called a **renewal point** (In a Markov process, every point t is a renewal point). If we focus on a discrete set of renewal points then the states of the system can be considered as a Markov chain, this is called a Imbedded Markov Chain (IMC).

• With regard to the M/G/1 queue, the departure epochs are renewal points. For whenever a customer leaves the system, either the system becomes empty or a previously waiting customer starts service, in either case the transition probabilities (and therefore the future evolution of the system) depend only on the number of customers in the system at the departure epoch.

• We proceed to study the M/G/1 queue by using the IMC defined on the set of departure epochs.

• Let N_k^* be **the number of customers in the system** at the instant **the kth customer completes service and departs**. Then

$$P\{N_{k+1}^* = j\} = \sum_{i=0}^{\infty} P\{N_{k+1}^* = j | N_k^* = i\} P\{N_k^* = i\} \quad (j = 0, 1, \ldots; k = 1, 2, \ldots). \tag{15}$$

• Observe that

$$\begin{aligned} P\{N^*_{k+1} = j | N^*_k = 0\} &= P\{j \text{ customers arrive during the service time} \\ &\qquad \text{of the } (k+1)\text{th customer}\} \\ &= \int_0^\infty \underbrace{\frac{(\lambda t)^j}{j!} e^{-\lambda t}}_{\text{Poisson arrival}} \times \underbrace{f(t)}_{\text{p.d.f. of service time}} dt \quad (j \geq 0) \end{aligned} \tag{16}$$

where $f(t)$ is the service time density function. Similarly for $i > 0$ and $j \geq i - 1$

$$\begin{aligned} P\{N^*_{k+1} = j | N^*_k = i\} &= P\{(j - i + 1) \text{ customers arrive} \\ &\qquad \text{during the service time of the } (k+1)\text{th customer}\} \\ &= \int_0^\infty \frac{(\lambda t)^{j-i+1}}{(j-i+1)!} e^{-\lambda t} f(t) dt \end{aligned} \tag{17}$$

and of course,

$$P\{N^*_{k+1} = j | N^*_k = i\} = 0 \quad \text{if} \quad i > 0 \ , \ j < i - 1 \ .$$

• Let us denote the right-hand-side of Eq. (16) by p_j. Then the transition probabilities are

$$P\{N^*_{k+1} = j | N^*_k = i\} = \begin{cases} p_j & \text{if } i = 0 \\ p_{j-i+1} & \text{if } i > 0, \text{ and } j \geq i - 1 \\ 0 & \text{if } i > 0, \text{ and } j < i - 1. \end{cases} \tag{18}$$

Note that the transition probabilities are **independent** of the value of the index k.

If we introduce the notations

$$\begin{cases} a_j^{(k)} = P\{N_k^* = j\}\ , \quad a^{(k)} = (a_0^{(k)}, a_1^{(k)}, a_2^{(k)}, \ldots)^T \\ p_{ij} = P\{N_{k+1}^* = i | N_k^* = j\}\ , \quad \text{and} \\ P = (p_{ij})^T \quad (P \text{ is called the transition matrix}) \end{cases}$$

where A^T denotes the transpose of the matrix A. Then

$$\boxed{a^{(k+1)} = Pa^{(k)}} \tag{19}$$

where P is a transition probability matrix given by

$$P = \begin{pmatrix} p_0 & p_0 & 0 & 0 & 0 & \cdots \\ p_1 & p_1 & p_0 & 0 & 0 & \cdots \\ p_2 & p_2 & p_1 & p_0 & 0 & \cdots \\ p_3 & p_3 & p_2 & p_1 & p_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

- It can be shown by using the theory of Markov chains that a unique probability distribution

$$\boxed{\pi_j^* = \lim_{k\to\infty} a_j^{(k)} \quad (j = 0, 1, 2, \ldots)} \tag{20}$$

independent of the initial conditions, exists if and only if $\rho < 1$. (If $\rho \geq 1$ then $\pi_j^* = 0$ for all finite j).

• Furthermore, $\{\pi_j^*\}$ is a stationary distribution, i.e. if

$$\pi^* = (\pi_0^*, \pi_1^*, \ldots,)^T$$

then

$$\pi^* = P\pi^*. \tag{21}$$

• The proofs of (20) and (21) are the only place where the formal theory of Markov chains is used. We shall **accept without proof** that the existence of the limiting distribution (20).

• Eq. (21) now becomes

$$\boxed{\pi_j^* = p_j\pi_0^* + \sum_{i=1}^{j+1} p_{j-i+1}\pi_i^* \quad (j = 0, 1, \ldots).} \tag{22}$$

The normalization equation is

$$\sum_{j=0}^{\infty} \pi_j^* = 1. \tag{23}$$

• **We want to solve π_i^* given that p_i is known**. Now let us define the probability generating function

$$\boxed{g(z) = \sum_{j=0}^{\infty} \pi_j^* z^j.} \tag{24}$$

• Substituting Eq. (22) into Eq. (24) we have

$$g(z) = \sum_{j=0}^{\infty} \underbrace{\left(p_j \pi_0^* + \sum_{i=1}^{j+1} p_{j-i+1} \pi_i^* \right)}_{\pi_j^*} z^j = \pi_0^* \sum_{j=0}^{\infty} p_j z^j + \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} p_{j-i+1} \pi_i^* z^j. \tag{25}$$

If we also define the probability generating function of p_i as

$$\boxed{h(z) = \sum_{j=0}^{\infty} p_j z^j.} \tag{26}$$

• We obtain after changing the order of summation,

$$\begin{aligned} g(z) &= h(z)\pi_0^* + \sum_{i=1}^{\infty} \sum_{j=i-1}^{\infty} p_{j-(i-1)} z^j \pi_i^* \\ &= h(z)\pi_0^* + \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} p_k z^{k+(i-1)} \pi_i^* \\ &= h(z)\pi_0^* + \sum_{i=1}^{\infty} h(z) \pi_i^* z^{i-1} \\ &= h(z)\pi_0^* + \frac{1}{z} h(z) \left(g(z) - \pi_0^* \right) \ . \end{aligned}$$

- Re-arranging the terms we get

$$\boxed{g(z) = \frac{h(z)(z-1)}{z-h(z)}\pi_0^*.} \tag{27}$$

- To determine π_0^* we have to make use of the result that

$$g(1) = \sum_{j=0}^{\infty} \pi_j^* = 1.$$

- From Eq. (27), we have

$$\begin{aligned} 1 &= g(1) \\ &= \lim_{z\to 1} \frac{h(z)(z-1)}{z-h(z)}\pi_0^* \quad (0/0 \text{ form}) \\ &= \lim_{z\to 1} \frac{h'(z)(z-1)+h(z)}{1-h'(z)}\pi_0^* \quad (\text{L'Hospital Rule}) \\ &= \frac{1}{1-h'(1)}\pi_0^* \quad (h(1)=1). \end{aligned}$$

- What is $h'(1)$? We shall investigate this.

- Hence

$$
\begin{aligned}
\pi_0^* &= 1 - h'(1) \\
&= 1 - \sum_{j=1}^{\infty} j p_j \\
&= 1 - \sum_{j=1}^{\infty} j \left(\int_0^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} f(t) dt \right) \\
&= 1 - \sum_{j=1}^{\infty} \left(\int_0^{\infty} \frac{(\lambda t)^j}{(j-1)!} e^{-\lambda t} f(t) dt \right) \\
&= 1 - \int_0^{\infty} \left(\sum_{j=1}^{\infty} \frac{(\lambda t)^j}{(j-1)!} \right) e^{-\lambda t} f(t) dt \\
&= 1 - \int_0^{\infty} \left(\lambda t e^{\lambda t} \right) e^{-\lambda t} f(t) dt \\
&= 1 - \int_0^{\infty} \lambda t f(t) dt \\
&= 1 - \frac{\lambda}{\mu} \quad (\mu^{-1} \text{is the expected service time}) \\
&= 1 - \rho.
\end{aligned}
$$

We remark that

$$\pi_0^* = 1 - \rho$$

holds for any service probability distribution and $0 < \rho < 1$.

20 Laplace-Stieltjes Transform

Let us now introduce the method of Laplace-Stieltjes transform and show that the generating function $h(z)$ (and hence $g(z)$) can be expressed in terms of the Laplace-Stieltjes transform of the service time distribution function.

Definition: A **Stieltjes integral** of a function $g(x)$ with respect to a function $h(x)$ from a to b is

$$\boxed{\int_a^b g(x)dh(x) = \lim_{\|\triangle\|\to 0} \sum_{k=1}^{n} g(\xi_k)[h(x_k) - h(x_{k-1})]}$$

where

$$x_{k-1} \leq \xi_k \leq x_k \ (k = 1, 2, \ldots, n),$$

and

$$\|\triangle\| = \max(x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}).$$

Note:

(1) If $h(x)$ is continuously differentiable on $[a, b]$, then

$$\int_a^b g(x)dh(x) = \int_a^b g(x)h'(x)dx,$$

where the RHS is the **ordinary Riemann integral**.

(2) If

$$h(x) = \begin{cases} 1 & \text{for } x \geq c \\ 0 & \text{for } x < c \end{cases}$$

then

$$\int_a^b g(x)dh(x) = g(c). \tag{28}$$

20.1 Definition and basic properties of Laplace-Stieltjes transform

Definition: The ordinary Laplace transform of $L(f(t))$ of a distribution function, $f(t)$, for a continuous nonnegative random variable is given by

$$\boxed{L[f(t)] = \int_0^\infty e^{-st} f(t)dt \quad (s \geq 0).}$$

The Laplace-Stieltjes transform $L^*[F(t)]$ of a distribution function, $F(t)$, for a continuous nonnegative random variable is given by

$$\boxed{L^*[F(t)] = \phi(s) = \int_0^\infty e^{-st} dF(t) \quad (s \geq 0).}$$

For a discrete probability distribution $\{f_n\}$, the Laplace-Stieltjes transform

$$L^*[\{f_n\}] = \phi(s) = \sum_{n=0}^{\infty} e^{-ns} f_n.$$

Some Basic properties:

(1) If $g(z)$ is the **generating function** of $\{f_n\}$, then

$$L^*[\{f_n\}] = \sum_{n=0}^{\infty} f_n (e^{-s})^n = g(e^{-s}).$$

(2) If $F(t)$ is differentiable, $F'(t) = f(t)$, then $\phi(s)$ is the **ordinary Laplace transform** $L[f(t)]$ of $f(t)$:

$$L^*[F(t)] = \int_0^{\infty} e^{-st} dF(t) = \int_0^{\infty} e^{-st} f(t) dt = L[f(t)].$$

(3)

$$\begin{aligned}
(-1)^n \frac{d^n}{ds^n} \phi(s)\Big|_{s=0} &= (-1)^n \int_0^{\infty} \frac{d^n}{ds^n} e^{-st}\Big|_{s=0} dF(t) \\
&= (-1)^n \int_0^{\infty} (-1)^n t^n dF(t) \\
&= \int_0^{\infty} t^n dF(t).
\end{aligned}$$

(4) For the Laplace transformation, it is well known that if

$$\phi_1(s) = L(f_1(t)) \quad \text{and} \quad \phi_2(s) = L(f_2(t))$$

then

$$\begin{aligned}
\phi_1(s)\phi_2(s) &= \int_0^\infty e^{-sy} f_1(y)dy \times \int_0^\infty e^{-sx} f_2(x)dx = \int_0^\infty \int_0^\infty e^{-s(x+y)} f_1(y) f_2(x) dx dy \\
(\text{Let } t = x+y) &= \int_0^\infty e^{-st} \int_0^t f_2(x) f_1(t-x) dx dt = L\left(\int_0^t f_2(x) f_1(t-x) dx\right).
\end{aligned}$$

If $F_1(t)$ and $F_2(t)$ have Laplace-Stieltjes transforms $\phi_1(s)$ and $\phi_2(s)$ respectively then the Laplace-Stieltjes transforms of $\phi(s)$ of the **convolution**:

$$\boxed{F(t) = \int_0^t F_1(t-x) dF_2(x) = \int_0^t F_1(t-x) f_2(x) dx}$$

is the product

$$\boxed{\phi(s) = \int_0^\infty e^{-st} \left(\int_0^t f_2(x) f_1(t-x) dx\right) dt = \int_0^\infty e^{-st} dF(t) = \phi_1(s)\phi_2(s).}$$

• Given X_1 and X_2 are independent nonnegative random variables with distribution functions $F_1(t)$ and $F_2(t)$ and Laplace-Stieltjes transform $\phi_1(s)$ and $\phi_2(s)$.

• Then the distribution function of **the sum $(X_1 + X_2)$** has Laplace-Stieltjes transform equal to the product of the component transforms.

20.2 Applications

We consider

$$p_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dF(t) \quad (j = 0, 1, \ldots)$$

and

$$h(z) = \sum_{j=0}^{\infty} p_j z^j.$$

Then

$$\begin{aligned} h(z) &= \sum_{j=0}^{\infty} \left(\int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dF(t) \right) z^j \\ &= \int_0^\infty \left(\sum_{j=0}^{\infty} \frac{(\lambda t z)^j}{j!} \right) e^{-\lambda t} dF(t) \\ &= \int_0^\infty e^{\lambda z t} \times e^{-\lambda t} dF(t) \\ &= \int_0^\infty e^{-\lambda(1-z)t} dF(t). \end{aligned}$$

The last integral is the Laplace-Stieltjes transform with argument $\lambda(1-z)$ of the distribution function $F(t)$:

$$\boxed{h(z) = \phi(\lambda - \lambda z).} \tag{29}$$

- Let us now come back to the generating function $g(z)$ in Eq. (27).

- By Eq. (29)

$$g(z) = \frac{(z-1)\phi(\lambda - \lambda z)}{z - \phi(\lambda - \lambda z)}\pi_0^*.$$

Since $\pi_0^* = (1-\rho)$, we have the final result

$$\boxed{g(z) = \frac{(z-1)\phi(\lambda - \lambda z)}{z - \phi(\lambda - \lambda z)}(1-\rho) \quad \text{for} \quad (\rho < 1\ ,\ |z| \leq 1).} \tag{30}$$

Remark: From Eq. (30), we can, in principle, find the mean number of customers in the system

$$E(N^*) = g'(1).$$

and

$$\boxed{\pi_1^* = g'(0), \ldots, \pi_k^* = \frac{g^{(k)}(0)}{k!}.}$$

21 M/G/1 Queue

We now turn to the waiting time (in the queue) distribution

$$W(t) = P\{W \leq t\}$$

for the M/G/1 queue with service in order of arrival. Let $W(t)$ have Laplace-Stieltjes transform $\omega(s)$,

$$\omega(s) = \int_0^\infty e^{-st} dW(t), \tag{31}$$

and denote by $\eta(s)$ the Laplace-Stieltjes transform of the distribution function of the **sojourn time** $S(t)$ (the sum of the waiting time and the service time).

• Since the waiting time and service time of a customer are independent random variables, we have

$$\boxed{\eta(s) = \omega(s)\phi(s).} \tag{32}$$

Since service is in arrival order, the customers left behind by an arbitrary customer must all have arrived during his sojourn time. Therefore the probability generating function $g(z)$ of the number of customers left behind by an arbitrary customer is given by

$$\boxed{g(z) = \sum_{j=0}^{\infty} \left(\int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} dS(t) \right) z^j = \eta(\lambda - \lambda z).} \tag{33}$$

Using Eqs. (30), (32) and (33) we have

$$g(z) = \eta(\lambda - \lambda z) = \omega(\lambda - \lambda z)\phi(\lambda - \lambda z) \quad \text{and} \quad g(z) = \frac{(z-1)\phi(\lambda - \lambda z)}{z - \phi(\lambda - \lambda z)}(1-\rho).$$

Putting $s = \lambda - \lambda z$ we have

$$\omega(s) \;=\; \frac{-\frac{s}{\lambda}(1-\rho)}{(\frac{\lambda - s}{\lambda}) - \phi(s)} = \frac{s(1-\rho)}{s - \lambda[1-\phi(s)]}. \tag{34}$$

Using the fact that $E(W) = -\omega'(0)$ [Refer to basic property (3) of Section 20.1], it is a straightforward (an exercise) but tedious task (requiring two applications of L'Hôpital's rule) to establish that

$$\boxed{E(W) = \frac{\rho\tau}{2(1-\rho)}\left(1 + \frac{\sigma^2}{\tau^2}\right)}. \tag{35}$$

Here τ and σ^2 are respectively the mean and variance of the service time. By Little's formula

$$E(Q) = \frac{\rho^2}{2(1-\rho)}\left(1 + \frac{\sigma^2}{\tau^2}\right) \tag{36}$$

where Q is the mean number of customers waiting in the queue.

Remark: When the service times are exponential, then $\tau^2 = \sigma^2$ and (35) becomes

$$E(W) = \frac{\rho\tau}{1-\rho},$$

in agreement with formula for Erlang delay systems.

21.1 Special Case: M/D/1 Queue

Assuming all service times are constant and equal to $1/\mu$. Then the service time distribution function is given by

$$H(t) = \begin{cases} 0 & t < \frac{1}{\mu} \\ 1 & t \geq \frac{1}{\mu}. \end{cases}$$

We have [using (26)]

$$h(z) = \phi(\lambda - \lambda z) = \int_0^\infty e^{-\lambda(1-z)t} dH(t) = e^{-(1-z)\rho} \ , \quad (\rho = \lambda/\mu) \ .$$

Hence (30) becomes

$$g(z) = \frac{(1-z)(1-\rho)}{1 - ze^{(1-z)\rho}}. \tag{37}$$

Thus we have

$$\begin{cases} \pi_0^* = g(0) = (1-\rho) \\ \pi_1^* = g'(0) = (1-\rho)(e^\rho - 1) \end{cases} \tag{38}$$

To find π_n^* for $n = 2, 3, \ldots$ it is more convenient to employ the following method. Considering (37), we expand

$$\frac{1}{(1 - ze^{\rho(1-z)})}$$

as a geometric series giving

$$
\begin{aligned}
g(z) &= (1-z)(1-\rho)\sum_{k=0}^{\infty}(ze^{\rho(1-z)})^k \quad \text{for} \quad |ze^{\rho(1-z)}| < 1 \\
&= (1-z)(1-\rho)\sum_{k=0}^{\infty} e^{k\rho} z^k e^{-k\rho z} \\
&= (1-z)(1-\rho)\sum_{k=0}^{\infty} e^{k\rho} \sum_{m=0}^{\infty} \frac{(-k\rho z)^m}{m!} z^k \\
&= (1-z)(1-\rho)\sum_{k=0}^{\infty}\sum_{n=k}^{\infty} e^{k\rho}(-1)^{n-k}\frac{(k\rho)^{n-k}}{(n-k)!} z^n \qquad (\text{Let } n = m+k) \\
&= (1-z)(1-\rho)\left\{\sum_{n=0}^{\infty}\sum_{k=0}^{n} e^{k\rho}(-1)^{n-k}\frac{(k\rho)^{n-k}}{(n-k)!} z^n\right\} \\
&= (1-\rho)\left\{\sum_{n=0}^{\infty}\sum_{k=0}^{n} e^{k\rho}(-1)^{n-k}\frac{(k\rho)^{n-k}}{(n-k)!} z^n - \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} e^{k\rho}(-1)^{n-k-1}\frac{(k\rho)^{n-k-1}}{(n-k-1)!} z^n\right\}.
\end{aligned}
\tag{39}
$$

Thus for $n \geq 2$

$$
\pi_n^* = (1-\rho)\left(\sum_{k=1}^{n} e^{k\rho}(-1)^{n-k}\frac{(k\rho)^{n-k}}{(n-k)!} + \sum_{k=1}^{n-1} e^{k\rho}(-1)^{n-k}\frac{(k\rho)^{n-k-1}}{(n-k-1)!}\right). \tag{40}
$$

A Summary on Higher Dimensional and Non-Markovian Queueing Systems

- Product form solutions of higher dimensional queueing systems.

- Queues in tandem and queue in parallel.

- The statement of Burke's theorem on the M/M/s queue.

- The Imbedded Markov chain technique.

- The M/G/1 Queue:

$$g(z) = \frac{h(z)(z-1)(1-\rho)}{z - h(z)}$$

where $g(z)$ is the generating function of the steady state probabilities of the number of customers in the system and $h(z)$ is the generating function of the transition probabilities.

- The definition and properties of the Laplace-Stieltjes transform.

- The expected queueing time in the M/G/1 queue:

$$E(W) = \frac{\rho\tau}{2(1-\rho)}\left(1 + \frac{\sigma^2}{\tau^2}\right).$$