The theory of queues is concerned with the development of mathematical models to predict the behaviour of systems that provide services for randomly arising demands. Since the demands for service are assumed to be governed by some probability law, the theory of queues has been developed within the framework of the theory of stochastic processes. In order to investigate the stochastic properties of a particular queueing system, it is necessary to formulate a model which is based on (1) the input process, (2) the queue discipline, and (3) the service mechanism which characterize the queueing system. We have observed that the stochastic processes occurring in the theory of queues are in general non-Markovian. In section 1.1 we present the different representation of queueing discipline. We develop the different techniques adopted in solving the equation we come across in the context of queueing systems. The transform technique is studied in section 1.2. The recent technique of Neuts and Ramaswami [45] namely matrix geometric technique is analysed in section 1.3. When one needs numerical approximations to the solution of the integral equation, different quadrature formulae have been suggested. We present one such formula for solving the integral equation in queueing system in section 1.4. We illustrate the method by a numerical example.
SECTION 1.1. REPRESENTATION OF QUEUES

1.1.1. Markov chain representation

A method of reducing non-Markovian queueing processes to Markov chains which is of great importance in queueing theory has been developed by Kendall [29,30]. This method, called the method of the imbedded Markov chain, can be described abstractly as follows: Let the state of the queueing system at time $t$ be denoted by the random variable $Y(t)$, so that, in any realization of the process $\{Y(t), t \geq 0\}$, the history of the system can be represented as a function $Y(\bullet)$ of time with domain $(-\infty, \infty)$. Let the set $\Omega_t$ denote the collection of functions with domain $(-\infty, t]$ and having the same range as $Y(\bullet)$. For each $t \in (-\infty, \infty)$ let $\theta_t$ be a specified subset of $\Omega_t$ and corresponding to any actual realization of the process let $T$ be the set of those values of $t \in (-\infty, \infty)$ for which $\theta_t$ contains as an element the contraction of $Y(\bullet)$ to the reduced domain $(-\infty, t]$.

We next define the random variable

$$X(t) = \mathfrak{f}_t \{ Y(\tau) : \tau \leq t, t \in T \}$$

where $\mathfrak{f}_t$ is some specified functional with domain $\theta_t$. If we can select a domain $\theta_t$ and a functional $\mathfrak{f}_t$, for $t \in (-\infty, \infty)$, such that

2. The set $T$ almost certainly has no finite point of accumulation. (Hence the elements $t$ can be ordered as follows: $\ldots, t_{n-1}, t, t_{n+1}, \ldots$)

2. If $X_n = X(t_n)$ for each $t_n \in T$, then

$$\mathcal{P} \{ X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots \} = \mathcal{P} \{ X_{n+1} = x_{n+1} | X_n = x_n \}$$

for all $n$;

the process $\{X_n, n = 0, 1, \ldots\}$ will then be said to be an imbedded Markov chain.
1.1.2. Markov Process Representation

We observe that it is only for queueing systems of type $M/M/s$ that the stochastic process associated with the fluctuations in queue size are Markovian. Hence, in these cases if we let the random variable $X(t)$ denote the number of customers in the queue at time $t$, i.e., the queue size at time $t$, the stochastic process \{X(t), t \geq 0\} is a Markov process with a denumerable number of states, and its stochastic properties can be obtained from the solutions of the Kolmogorov differential equations representing the process.

Let $P(t) = (p_{ij}(t))$ denote the matrix of transition probabilities associated with the process \{X(t)\}. We note that $P(t)$ satisfies the system of Kolmogorov equations

$$\frac{Dp(t)}{dt} = P(t) A(t), \quad \frac{Dp(t)}{dt} = A(t) P(t)$$

\text{... (1)}

with

$$P(0) = I \text{ the identity matrix}$$

In (1) $A(t) = (a_{ij}(t))$ is the matrix of infinitesimal transition probabilities. Therefore, in terms of the matrix elements the above equations become

$$\frac{D p_{ij}(t)}{dt} = \sum_{k=0}^{\infty} P_{ik}(t) a_{kj}(t)$$

$$\frac{D p_{ij}(t)}{dt} = \sum_{k=0}^{\infty} a_{ik}(t) P_{kj}(t) \quad i, j = 0, 1, \ldots$$

with

$$p_{ij}(0) = \delta_{ij} = 0 \text{ for } i \neq j$$

$$= 1 \text{ for } i = j$$
In order to solve the Kolmogorov equations, it is necessary to specify the functions $a_j(t)$, which in the study of queueing systems will be functions of time, and the parameters characterizing the inter-arrival time and service-time distribution functions. Once the $a_j(t)$ are specified, Kolmogorov equations can be solved.

1.1.3. Integrodifferential Equation Representation

We now consider another approach to the study of non-Markovian queueing systems of the type GI/M/1. In this approach, due to Takács [63], the state of the queueing system is characterized by a single random variable which denotes the waiting time of a customer joining the queue at time $t$. The state of the queueing system was characterized by the number of customers in the queue at a time $t$; hence, it was necessary to consider an infinite system of differential-difference equations in order to represent the process. In the Takács theory it is only necessary to consider a single integrodifferential equation for the distribution function of the waiting time.

Let the random variable $W(t)$ denote the waiting time of a customer joining the queue at time $t$. If at time $t = 0$ the server is not busy,
W(0) = 0. Should the server be busy at t = 0, we denote by W(0) = wo the random variable that denotes the time when the server ceases to be busy. If at some time t', W(t') = 0, then it remains zero until a customer joins the queue.

Let \{t_n\} and \{\xi_n\} denote the sequence of arrival times and service times, respectively. The random variable W(t) changes as follows: It jumps upward discontinuously every time a customer with nonzero service time joins the queue. Otherwise W(t) approaches zero with slope-1 until it jumps again or reaches 0. At 0 it remains equal to 0 until another jump occurs. The magnitudes of the jumps W(t) experiences are the service times of the customers, \xi_1, \xi_2, \ldots, arriving at times t_1, t_2 \ldots Hence, W(t) changes both discontinuously and continuously.

The changes in W(t) are illustrated in Fig. 1.1.1.

In view of the above, W(t) can be defined as follows: Let W(0) = wo; then for \(t_n < t < t_{n+1}, n = 0, 1, \ldots,\) with \(t_0 = 0,\)

\[
W(t) = \begin{cases} 
W(t_n) - (t - t_n) & \text{for } W(t_n) > t - t_n \\
0 & \text{for } W(t_n) \leq t - t_n 
\end{cases} \quad \ldots (2)
\]

and if \(t = t_n,\)

\[
W(t_n) = W(t_{n-0}) + \xi_n \quad \ldots (3)
\]

Hence, if the probability laws governing the sequences of arrival times and service times are known, W(t) can be determined for all \(t \geq 0\) from the initial condition \(W(0) = w_o\) and Eqs.(2) and (3).

We know that in the case when the arrival times are Poisson and the service times are mutually independent random variables, the process \{W(t), t \geq 0\} is
Markovian. However, if the arrival times are not Poisson but the differences 
\( \tau_n = t_n - t_{n-1} \) form a recurrent process, the process is non-Markovian. In this case the
arrival times \( t_n \) are the regeneration points of the process. From the nature of the
changes which \( W(t) \) experiences we can conclude that the process \( \{W(t), t \geq 0\} \) is a
Markov process of the mixed type; i.e., it is a Markov process characterized by both
discontinuous and continuous changes of state.

We now consider the integral equation representation of queueing system due
to Lindley [37]. It is assumed that:

1. The process \( \{\tau_n\} \) is a recurrent process, and the distribution function \( A(t) \)
is arbitrary.

2. The service times \( \xi_1, \xi_2, \ldots \) are equidistributed mutually independent random variables with common distribution function \( B(\xi) \).

3. There is only one server at the counter.

In view of the above assumptions, the queueing system we consider is of the
type GI/G/1.

Let the random variable \( W(t) \) denote the waiting time of the customer arriving
at the service at time \( t \). In general, the process \( \{W(t), t \geq 0\} \) is non-Markovian. Now
let \( W(t_n-0) = W_n \) denote the waiting time of the \( n^{th} \) customer to arrive at the counter.
Hence, the process \( \{W_n, n = 0, 1, \ldots\} \) is the imbedded Markov chain associated with
the queueing system. If at time \( t = 0 \) the server is free, then \( W_0 = 0 \); however, if the
server is not free, \( W_0 \) denotes the time that elapses before the server is free. If \( W_0 \) is
known, the random variables \( W_n \) may be determined successively from the following
equation:
\[ W_{n+1} = W_n + \xi_n - \tau_{n+1} \quad \text{if} \quad \tau_{n+1} - \xi_n < W_n \]
\[ = 0 \quad \text{if} \quad \tau_{n+1} - \xi_n \geq W_n \]

The interpretation of this equation should be clear.

Let \( F_n(t) = \mathcal{P}\{W_n \leq t\} \) denote the distribution function of \( W_n \); that is, \( F_n(t) \) is the distribution function of the waiting time of the \( n^{th} \) customer to arrive at the service. It is to be noted that, since the random variables \( W_n \) are non-negative, \( F_n(t) = 0 \) for \( t < 0 \), and that \( F_n(0) \) is the probability that the \( n^{th} \) customer will not have to wait.

If we start with \( F_0(t) \), the sequence of distribution functions \( \{F_n(t)\} \) may be determined by the recurrence formula

\[ F_{n+1}(t) = \int_0^\infty K(t, u) \, Df_n(u) \quad n = 0, 1, 2, \ldots \]

where the kernel function \( K(t, u) \) expresses the transition probabilities of the imbedded Markov chain \( \{W_n\} \); that is,

\[ K(t, u) = P\{W_{n+1} \leq t \mid W_n = u\} \]
\[ = \int_0^\infty [1 - A(w+u-t) \, D]\]

The computations indicated may be carried out by utilizing the recurrence formulae

\[ K_n(t) = \int_0^\infty B(t-u) \, Df_n(u) \]

and

\[ F_{n+1}(t) = \int_0^\infty [1 - A(u-t)] \, Dk_n(u) \]
Theorem 1.1.1 (Lindley [37])

A necessary and sufficient condition that \( \lim_{n \to \infty} F_n(t) = F(t) \) exists is that either

\[
\mathbb{E}\{\xi_n\} \leq \mathbb{E}\{\tau_n\}, \text{ or } \tau_n - \xi_n = 0.
\]

If \( \mathbb{E}\{\xi_n\} \geq \mathbb{E}\{\tau_n\} \) and \( \tau_n - \xi_n \neq 0 \), then \( \lim_{n \to \infty} F_n(t) = 0 \) for every \( t \geq 0 \).

The limiting distribution \( F(t) \), if it exists, is independent of the initial distribution \( F_0(t) \) and is the unique solution of the integral equation

\[
F(t) = \int_0^\infty K(t, u) Df(u) \quad .... (4)
\]

Equation (4), which is an integral equation of the Wiener-Hopf type, has been solved by Lindley for the case of a queueing system of type D/Ek/1. Its solution has been considered by Smith [59] for cases in which more general arrival-time distributions are assumed. In particular, Smith has shown that, if the distribution function of the service times is exponential, so is that of the waiting time, regardless of the assumptions made concerning the distribution function of the arrival times.

SECTION 1.2. TRANSFORM TECHNIQUE

In this section, within the framework of the Takács theory we shall determine the following, using transform technique (1) the distribution function of the waiting time, (2) the distribution function of the length of the service period, and (3) the number of customers served in a service period.
Let $F(t, w) = \mathbb{P}(W(t) \leq w)$ denote the distribution function of $W(t)$. Hence, $F(t, w)$ gives the probability that the waiting time of a customer joining the queue at time $t$ does not exceed $w$. We now assume that:

2. The process $\{t_n\}$ is Poisson with parameter $\lambda(t)$; i.e., the probability of a customer arriving in the interval $(t, t + \Delta t)$ is $\lambda(t) \Delta t + o(\Delta t)$, where $\lambda(t)$ is a real-valued, non-negative, continuous and bounded function of $t$.

2. The service times $\xi_n$ are equidistributed mutually independent random variables with common distribution function $B(w)$.

3. The initial condition $W(0) = \omega_0$ is in general arbitrary, with distribution function

$$F_0(w) = \mathbb{P}(\omega_0 \leq w); \text{ or, in particular } \omega_0 = 0 \text{ with}$$

$$F_0(w) = \begin{cases} 0 & \text{for } w < 0 \\ 1 & \text{for } w \geq 0 \end{cases}$$

**Theorem 1.2.1**

The distribution function $F(t, w)$ of the waiting time $W(t)$ satisfies the integrodifferential equation

$$\frac{\partial F(t,w)}{\partial t} = -\lambda(t) F(t, w) + \lambda(t) \int_0^w B(w-z) d_z F(t,z) \quad \ldots (5)$$

where all the derivatives occurring in (5) are either right-hand derivatives ($w \geq 0$) or left-hand derivatives ($w < 0$). $F(t, w)$ has a jump of magnitude $F(t, 0)$ at $w = 0$, and
for $w > 0$ it is continuous for all $t$. By imposition of the initial condition $F(0, w) = 1$ ($w > 0$), Eq.(5) has a unique solution.

**Proof**

In order to derive Eq.(5), we consider the changes in the interval $(t, t + \Delta t)$ that can bring about the occurrence of the event \{W(t + \Delta t) \leq w\}. We have the following:

1. In the interval $(t, t + \Delta t)$, the probability that no customer joins the queue is $1 - \lambda(t) \Delta t + o(\Delta t)$. In this case $P\{W(t) < w + \Delta t\} = F(t, w + \Delta t)$.  

2. In the interval $(t, t + \Delta t)$, the probability that one customer joins the queue is $\lambda(t) \Delta t + o(\Delta t)$. In this case if $W(t) = z$ ($0 < z < w$), we must have $\xi \leq w - z$, the probability that it obtains being $B(w-z)$. The distribution function of $z$ is $F(t, z)$.  

3. In the interval $(t, t + \Delta t)$, the probability that more than one customer joins the queue is $o(\Delta t)$.

Since these events are mutually exclusive, we obtain the equation

$$F(t + \Delta t, w) = [1 - \lambda(t) \Delta t] F(t, w + \Delta t) + \lambda(t) \Delta t \int_0^w B(w-z)dz F(t, z) + o(\Delta t) \quad \cdots \cdots (6)$$

We now have to determine $F(t, w + \Delta t)$. It can be shown that for $w > 0$, $F(t, w)$ is a continuous function of $w$. For $w = 0$, $F(t, w)$ has in general a jump of magnitude $F(t, 0)$. Hence, for $\Delta t > 0$ we have
where the partial derivative with respect to \( w \) is a right-hand derivative which exists for \( w \geq 0 \).

If we now insert (7) in (6) and let \( \Delta t \to 0 \), we obtain Eq.(5). Now, for \( \Re(s) \geq 0 \), let
\[
\phi(t, s) = \int_0^\infty e^{-sw} Df(t, w) \text{ and } \psi(s) = \int_0^\infty e^{-sw} Db(w)
\]
do not the Laplace-Stieltjes transforms of \( F(t, w) \) and \( B(w) \), respectively. From (5) we obtain the partial differential equation
\[
\frac{\partial \phi(t, s)}{\partial t} = [s - \lambda(t) + \lambda(t) \psi(s)] \phi(t, s) - Sf(t, 0)
\]
which is to be solved with the initial condition \( \phi(0, s) = 1 \). The unique solution of Eq.(8) that satisfies the above initial condition is
\[
\phi(t, s) = \exp \left\{ st - [1 - \psi(s)] \Lambda(t) \right\}
\]
\[
, [1-s \int_0^t \exp \{-st + [1 - \psi(s)] \Lambda(\tau)\} F(\tau, 0) d\tau]
\]
where \( F(\tau, 0) \) is the probability that at time \( \tau \) the server is not busy and
\[
\Lambda(t) = \int_0^t \lambda(\tau) d\tau
\]
Hence, the distribution function of the waiting time can be determined uniquely from (9) by the inversion theorem for Laplace transforms.
We now determine the limiting distribution

$$F^*(w) = \lim_{t \to \infty} F(t, w)$$

Let

$$2 = \int_0^\infty w \, Db(w)$$

denote the expected value of the random variable $\xi_n$ and assume that the Poisson density is asymptotically constant, i.e.,

$$\lim_{t \to \infty} \lambda(t) = \lambda > 0$$

We now state the following two results due to (Bharucha – Reid [4]):

**Result 1.2.1**

If $\lambda \mu < 1$, the limiting distribution $F^*(w)$ exists, is independent of the initial distribution $F_0(w)$, and is uniquely determined by the equations

$$F^*(0) = 1 - \lambda \mu$$

and

$$\frac{Df^*(w)}{dw} = \lambda \left[ F^*(w) - \int_0^w B(w-z) \, Df^*(z) \right]$$

where the derivative with respect to $w$ is the right-hand derivative.

If $\lambda \mu \geq 1$, the limiting distribution $F^*(w)$ does not exist; however,

$$\lim_{t \to \infty} F(t, w) = 0 \text{ for all } w.$$

**Result 1.2.2**

If $\lambda \mu < 1$, the limiting distribution $F^*(w)$ exists, is independent of the initial distribution $F_0(w)$, and is uniquely determined by the Laplace-Stieltjes transform
\[ \phi^*(s) = \int_0^\infty e^{-sw} Df^*(w) \Re(s) \geq 0 \]
satisfying
\[ \phi^*(s) = \frac{2 \lambda \mu}{2 \lambda ([1-\psi(s)]/s)} \tag{10} \]

The expression for the Laplace-Stieltjes transform of \( F^*(w) \) is equivalent to the solution obtained by Khintchine [32]. If, following Khintchine, we assume that for \( \lambda \mu < 1 \) the limiting distribution function \( F^*(w) \) exists, then Eq.(10) can be obtained from the fundamental integrodifferential equation (5). Hence, the Laplace-Stieltjes transform is given by
\[ \phi^*(s) = \frac{F^*(0)}{2 \lambda ([1-\psi(s)]/s)} \]

Since \( F^*(w) \) is a distribution function, we must have \( \lim_{s \to 0} \phi^*(s) = 1 \), and since \( \lim_{s \to 0} [1-\psi(s)]/s = \mu \), we have \( F^*(0) = 1 - \lambda \mu \).

**Illustration 1.2.1**

We consider the case when \( \xi_n \) has an exponential distribution,
i.e., \( B(w) = 1 - e^{-w/\mu} \) for \( w \geq 0 \)
\[ = 0 \quad \text{for} \quad w < 0 \]
Hence \( \psi(s) = \frac{2}{1 + \mu s} \) and \( \phi^*(s) = \frac{(1-\lambda \mu)(1+\mu s)}{(1-\lambda \mu) + \mu s} \)
Inversion yields the limiting distribution

\[ F^*(w) = 1 - \lambda \mu \exp \left\{ \begin{array}{c} - \mu \\ 1 - \lambda \mu \end{array} \right\} w \]

Let us now consider the determination of the distribution function of the length of the service period. We restrict our attention to the homogeneous case; i.e., we assume that the Poisson parameter \( \lambda(t) = \lambda \). To further simplify matters, we assume that \( W(0) = w_0 = 0 \). Let the random variable \( \theta_n \) denote the duration of the \( n^{th} \) service period and let \( G(x) = \mathcal{F}(\theta_n \leq x) \) denote the distribution function of the service period, which is assumed to be the same for all \( n \).

To determine \( G(x) \), we proceed as follows: We assume that a customer joins the queue and is served without waiting. If the duration of his service time is \( \tau \), then the probability that in the interval \( (0, \tau) \), \( n \) customers arrive at the counter is

\[ \frac{(\lambda \tau)^n}{n!} e^{-\lambda \tau} \]

\[ \text{.... (11)} \]

If \( n = 0 \), only one customer is served, and the associated distribution function is \( H(x) \). If \( n \geq 1 \), the server, after serving the first customer, starts to attend one of the customers waiting in the queue.

Let \( G_n(x) \) denote the \( n \)-fold convolution of \( G(x) \), i.e., \( G_n(x) \) is the distribution of \( n \) mutually independent random variables each of which has the distribution function \( G(x) \). Then

\[ G_n(x) = \int_0^x G_{n-1} (x - \tau) Dg(\tau) \quad n = 2, 3, \ldots \]
where $G_1(x) = G(x)$. Now, the length of the service period does not exceed $x$ if the service of the first customer lasts a time $\tau$ ($0 < \tau \leq x$), the distribution function of which is $B(\tau)$. During this service time, the probability that $n$ customers $n = 0, 1, \ldots$ arrive is given by (11). Also, the probability that the service time of these $n$ customers, and those arriving in the meantime, does not exceed $x-\tau$ is $G_n(x-\tau)$. Hence, we have

$$G(x) = \sum_{n=0}^{\infty} \frac{(-\tau)^n}{n!} \int_0^x \lambda^n e^{\lambda x} G_n(x-\tau) B(\tau) \quad \ldots (12)$$

Let $\Gamma(s)$ denote the Laplace-Stieltjes transform of $G(x)$. From Eq.(12) we obtain the functional equation

$$\Gamma(s) = \sum_{n=0}^{\infty} \frac{[\Gamma(s)]^n}{n!} \int_0^x (\lambda x)^n e^{\lambda x} B(x)$$

$$= \sum_{n=0}^{2} \frac{(-1)^n \lambda^n [\Gamma(s)]^n \psi^{(n)}(s+\lambda)}{n!}$$

$$= \psi [s + \lambda - \lambda \Gamma(s)]$$

This functional equation for $\Gamma(s)$ was first obtained by Kendall [30]. The distribution function $G(x)$ can be determined uniquely from $\Gamma(s)$ by inversion.

**Theorem 1.2.2 (Takács [63])**

The Laplace transform $\Gamma(s)$ of the distribution function $G(x)$ is the uniquely defined analytic solution of the functional equation

$$\Gamma(s) = \psi [s + \lambda - \lambda \Gamma(s)] \quad \ldots (13)$$

valid for $\Re(s) \geq 0$, where $\Gamma(s)$ is subject to the condition $\Gamma(\infty) = 0$. 
Let $p$ denote the smallest positive number for which

$$\psi[\lambda (1 - p)] = p$$

Then

$$\lim_{n \to \infty} G(x) = p$$

If $\lambda \mu \leq 1$, $p = 1$ and $G(x)$ is an honest distribution function, while if $\lambda \mu > 1$, $p < 1$ and $G(x)$ is a dishonest distribution function, i.e., in such cases the service period can be infinite with probability $(1-p)$.

**Illustration 1.2.2**

We again consider the case when $B(x)$ is the exponential distribution function.

Hence $\psi(s) = (1 + \mu s)^{-1}$ and

$$\Gamma(s) = \frac{1 + \mu(s + \lambda) - \sqrt{(1 + \mu(s + \lambda))^2 - 4\lambda \mu}}{2 \lambda \mu}$$

Since $\Gamma(\infty) = 0$, we take the root of (14) with positive sign. By inverting (14), we obtain

$$\frac{Dg(x)}{dx} = \exp \left\{ - \frac{(1 + \lambda \mu)x}{\mu} \right\} I_1 \left\{ \frac{2\sqrt{\lambda \mu} x}{\mu} \right\} \frac{1}{\sqrt{\lambda \mu} x}$$

where $I_1(x) = J_1(ix)/i$ and $J_1(x)$ denotes the Bessel function of order 1 with imaginary argument. Here

$$G(\infty) = \begin{cases} 1 & \text{for } \lambda \mu \leq 1 \\ \frac{1}{\lambda \mu} & \text{for } \lambda \mu > 1 \end{cases}$$
The moments of $G(x)$ can be determined from the functional equation (13). However, the expected duration of the service period can be obtained directly from (12).

By definition,

$$\mathcal{E}\{\theta\} = \int_0^\infty xDg(x)$$

hence

$$\mathcal{E}\{\theta\} = \mu + \sum_{n=1}^\infty n\mu \int_0^\infty \frac{(\lambda x)^n}{n!} e^{\lambda x} Dh(\tau)$$

As a final problem we consider the determination of the number of customers that will be served in a service period. Let $P_i$ denote the probability that $i$ ($i=1, 2, \ldots$) customers are served in a service period. Now the probability that $n$ customers arrive at the counter while one customer is being served is

$$q_n = \int \frac{(\lambda x)^n}{2^n n!} e^{\lambda x} Dh(x) \quad n = 0, 1, 2, \ldots$$

Now it is clear that

$$P_i = \sum q_{n_1} q_{n_2} \ldots q_{n_i}$$

$$n_1 + \ldots + n_i = i - 1$$

$$n_1 + \ldots + n_k \geq k \quad (k = 1, 2, \ldots, i-1)$$

We now introduce the generating function

$$F(s) = \sum_{i=1}^\infty P_i s^i \quad |s| \leq 1$$

and the recurrence relations

$$P_1 = q_0$$

$$P_i = \sum_{n=1}^{i-1} q_n \sum_{n_1 + \ldots + n_i = i-1} P_{i_1} P_{i_2} \ldots P_{i_n}$$
From the above recurrence relations, forming the generating function, we obtain the functional equation

\[ F(s) = s \psi \left[ \lambda - \lambda F(s) \right] \]

Where we have used the relation

\[ \sum_{i=0}^{\infty} q_i s^i = \psi \left[ \lambda (1 - s) \right] . \]

SECTION 1.3. MATRIX GEOMETRIC TECHNIQUE

1.3.1. General Study

We start with the following definitions.

Definition 1.3.1

Consider a discrete time Markov chain \( \{X_t, t \in \mathbb{N}\} \) on the two-dimensional state space \( \{(n, i) : n \geq 0, 1 \leq i \leq m\} \), which we partition as \( U_{n \geq 0} \), where \( \zeta(n) = \{(n, 1), (n, 2), \ldots, (n, m)\} \) for \( n \geq 0 \). The first co-ordinate \( n \) is called the level, and the second coordinate \( j \) is called the phase of the state \( (n, j) \). We shall also use the word level to denote the whole subject \( \zeta(n) \).

Definition 1.3.2

The Markov chain is called a QBD if one-step transitions from a state are restricted to states in the same level or in the two adjacent levels.
Remark 1.3.1

The transition probabilities are assumed to be level-independent. More precisely, for \( n \) or \( n' \) greater than or equal to 1, the probability \( P[X_1 = (n', j) | X_0 = (n, i)] \) may depend on \( i, j \) and \( n' - n \), but not on the specific values of \( n \) and \( n' \). Thus, the transition matrix is block-tridiagonal and has the following form:

\[
P = \begin{pmatrix}
B & A_0 & 0 & 0 & \ldots \\
A_2 & A_1 & A_0 & 0 & \ldots \\
0 & A_2 & A_1 & A_0 & \ldots \\
0 & 0 & A_2 & A_1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

where \( A_0, A_1, A_2 \) and \( B \) are square matrices of order \( m \).

We temporarily assume that the QBD is aperiodic and positive recurrent and denote by \( \pi \) its stationary probability vector. It is the unique solution of the system \( \pi = \pi P, \pi 1 = 1 \).

We partition the vector \( \pi \) by levels into subvectors \( \pi_n, n \geq 0 \), where \( \pi_n \) has \( m \) components. The defining system may be decomposed as

\[
\pi_0 (B - I) + \pi_1 A_2 = 0, \\
\pi_{n-1} A_0 + \pi_n (A_1 - I) + \pi_{n+1} A_2 = 0 \quad \text{for } n \geq 1, \\
\sum_{n \geq 0} \pi_n 1 = 1.
\]
Notation

Let $D$ be a proper subset of $S$, let $P_D = \{P_{ij} : i, j \in D\}$ be the submatrix of transition probabilities between states of $D$, and let us denote by $N_D$ the matrix of expected sojourn times in the states of $D$ before the first visit to any state outside of $D$. One has that

$$[N_D]_{ij} = \mathbb{E} \left[ \sum_{n \geq 0} I \{ X_n = j \text{ and } S_D \geq n \} | X_0 = i \right],$$

where $I\{\cdot\}$ is the indicator function and $S_D$ is the total sojourn time in $D$.

Result 1.3.1

Assume that the Markov chain is irreducible. Its limiting probability vector satisfies

$$\pi_D = \pi_T \tau_{TD}$$

for all disjoint subsets of states $T$ and $D$, where the matrix $\tau_{TD}$ records the expected numbers of visits to the states of $D$ between two visits to $T$.

Result 1.3.2

The following relations hold for any disjoint subsets $T$ and $D$:

$$\tau_{TD} = \tau_{TD} \tau_{DD}$$

and

$$\tau_{DD} = \sum_{v \geq 0} (\tau_{DD})^v = (I - \tau_{DD})^{-1}.$$

Theorem 1.3.1

If the QBD is positive recurrent, then there exists a nonnegative matrix $N$ of order $m$ such that
\[ \pi_{n+1} = \pi_n A_0 N \text{ for } n \geq 0. \]

The matrix \( N \) is such that \( N_{ij} \) (\( 1 \leq i, j \leq m \)) is equal to the expected number of visits to the state \((n, j)\), starting from the state \((n, i)\), before the first visit to any of the states in \( \mathcal{C}(n-1) \) and is independent of \( n \geq 1 \).

We may also write
\[ \pi_n = \pi_0 R^n \quad \text{for } n \geq 0, \]
\[ (15) \]
where \( R = A_0 N \) is such that, for any \( n \geq 0 \), \( R_{ij} \) (\( 1 \leq i, j \leq m \)) is the expected number of visits to \((n + 1, j)\) before a return to \( \mathcal{C}(0) \cup \ldots \cup \mathcal{C}(n) \), given that the process starts in \((n, i)\).

**Proof**

Fix \( n \geq 0 \) and partition the state space as \( T \cup T^c \), where \( T = \mathcal{C}(0) \cup \ldots \cup \mathcal{C}(n) \), \( T^c = \mathcal{C}(n+1) \cup \mathcal{C}(n+2) \cup \ldots \). Also we have

\[
P_{T^c} = \begin{pmatrix}
A_1 & A_0 & 0 & 0 & \ldots \\
A_2 & A_1 & A_0 & 0 & \ldots \\
0 & A_2 & A_1 & A_0 & \ldots \\
0 & 0 & A_2 & A_1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & \ldots \\
\end{pmatrix}
\]

and

\[
P_{TT^c} = \begin{pmatrix}
0 & 0 & 0 & \ldots \\
\ldots & \ldots & \ldots & \ddots \\
\ldots & \ldots & \ldots & \ddots \\
0 & 0 & 0 & \ldots \\
A_0 & 0 & 0 & \ldots \\
\end{pmatrix}
\]
By Result 1.3.1 and 1.3.2, we have that $\pi_T = \pi_T P_T \pi_T (I - P_T)^{-1}$, or in expanded form,

$$(\pi_{n+1}, \pi_{n+2}, \ldots) = (\pi_0, \ldots, \pi_n) P_T \pi_T (I - P_T)^{-1},$$

where $(I - P_T)^{-1} = \sum_{i=0}^\infty P_T^i$ converges since we have assumed that the QBD is irreducible.

We decompose the matrix $N_T = (I - P_T)^{-1}$ into blocks $\{N_{kk'}, k, k' \geq 1\}$, where $N_{kk'}$ is the matrix of expected number of visits to the states in $\mathcal{E}(n+k')$, starting from a state in $\mathcal{E}(n+k)$, before the first visit to any of the states in $T = \mathcal{E}(0) \cup \ldots \cup \mathcal{E}(n)$.

Because of the extremely sparse structure of the matrix $P_T \pi_T$, we have that

$$P_T \pi_T (I - P_T)^{-1} = \begin{pmatrix}
0 & 0 & 0 & \ldots \\
\vdots & \ddots & \ddots & \\
\vdots & \ddots & 0 & 0 \\
0 & 0 & 0 & \ldots \\
A_0 N_{11} & A_0 N_{12} & A_0 N_{13} & \ldots
\end{pmatrix}$$

and that

$$(\pi_{n+1}, \pi_{n+2}, \ldots) = (\pi_n A_0 N_{11}, \pi_n A_0 N_{12}, \ldots).$$

In particular, we have that $\pi_{n+1} = \pi_n A_0 N_{11} = \pi_n A_0 N$ if we define $N = N_{11}$.

We now make the first of many uses of the homogeneity of the matrix $P$. Irrespective of the value chosen for $n$ in defining the partition $T U T^c$, the matrix $P_T$ is the same. Therefore, the matrix $N_{11} = N$ is independent of $n$ and it records the
expected number of visits to $\ell(n+1)$, starting from $\ell(n+1)$, before the first visit to 
$\ell(0) \cup \ldots \cup \ell(n)$ for all values of $n \geq 0$, as we claim in the statement of the theorem.

Furthermore, the structure of $P_{TT^c}$ is independent of $n$; thus the argument 
above applies to every level, and we have that $\pi_{n+1} = \pi_n \Lambda_0 N$ for all $n \geq 0$, which may 
obviously be written as $\pi_n = \pi_0 R^n$ for all $n \geq 0$ if we set $R = \Lambda_0 N$. Clearly, $\Lambda_0 N$ 
records the expected number of visits to $\ell(n+1)$ starting from $\ell(n)$, avoiding 
$U_{0 \leq k \leq n} \ell(k)$, which concludes the proof.

**Remark 1.3.2**

The matrix $R = \Lambda_0 N_{11}$ is also interpreted as recording the expected rate of 
visit to the states of $\ell(n+1)$ per unit of time spent in the states of $\ell(n)$ measured in the 
local time of the whole subset $T = \ell(0) \cup \ldots \cup \ell(n)$. Because of the structure of the 
QBD, the Markov chain is forced to pass through the level $n$ when it leaves the 
complementary set $T^c$. Therefore, we may restrict ourselves to the local time of the 
level $n$ only and interpret $R_{ij}$ as the expected rate of visit to $(n + 1, j)$ per unit of the 
local time of $\ell(n)$ spent in $(n, i)$.

**Corollary 1.3.1**

If the QBD is positive recurrent and if $m$ is finite, then the spectral radius 
in $\lambda_\rho(R)$ of the matrix $R$ is strictly less than 1.

**Proof**

If the QBD is positive recurrent, then the series $\sum_{n \geq 0} \pi_n$ must converge. Since $\pi_n = \pi_0 R^n$ for $n \geq 0$, the series $\sum_{n \geq 0} R^n$ must converge, and this implies that $\lambda_\rho(R) < 1$. 
Remark 1.3.3

When m is infinite, it is proved in Tweedie [70] that there exists $\theta > 1$ such that $\sum_{n \geq 0} \theta^n (R^n)_{ij}$ diverges for some i and j and $\sum_{n \geq 0} r^n R^n$ converges whenever $r < \theta$. The quantity $\theta^{-1} < 1$ is the natural analogue of the spectral radius for finite matrices.

The matrix-geometric property (15) leads to very simple expressions for various marginal distributions and moments. For instance, the vector

$$\phi = \sum_{n \geq 0} \pi_n = \pi_o (I-R)^{-1}$$

is the marginal probability distribution of the phase. The marginal distribution $\{p_n, n \geq 0\}$ of the level is given by

$$p_n = \pi_n 1 = \pi_o R^n 1 \quad \text{for } n \geq 0$$

and has a discrete phase-type representation as we show below.

Theorem 1.3.2

The marginal distribution $\{p_n, n \geq 0\}$ of the level has the PH representation $$(\xi, V)$$
with m phases, $\xi = \pi_o R(I-R)^{-1}$, and $V = \Delta^{-1} R^t \Delta$, where $\Delta = \text{diag} (\xi)$ and $R^t$ is the transpose of $R$.

Proof

The vector $\xi$ is clearly strictly positive, and we have that $\xi 1 = 1 - \pi_o 1 < 1$. Thus, $\xi$ satisfies the conditions for being the initial probability vector in a phase-type representation. Furthermore, $V$ is nonnegative and

$$1 - V 1 = \Delta^{-1} (I-R^t) \Delta 1$$
$$= \Delta^{-1} (I-R^t) \xi^t$$
$$= \Delta^{-1} R^t \pi_o^t$$
$$> 0,$$
so that the matrix $V$ is substochastic. Thus, $(\xi, V)$ is the representation of a phase-type distribution. Since

$$p_n = \pi_0 R^n 1 = e(R)^n \pi_0^t,$$

where $e = (1, 1, \ldots, 1)$, we have that

$$p_n = e \Delta V^n \Delta^{-1} \pi_0^t = \xi V^{n-1} (V \Delta^{-1} \pi_0^t) = \xi V^{n-1} (\Delta^{-1} R^t \pi_0^t) = \xi V^{n-1} (1-V1),$$

which proves that $\{p_n\}$ is the density of the PH $(\xi, V)$ distribution.

We may directly evaluate the factorial moments $M_k$ of the distribution $\{p_n : n \geq 0\}$ as

$$M_k = k! \xi (I-V)^k V^{k-1} 1 \text{ for } k \geq 1.$$

**Theorem 1.3.3**

The matrices $\{N_{ik}, k \geq 1\}$ satisfy the following recurrence equation:

$$N_{i,k+1} = N_{i1} A_0 N_{ik} \text{ for } k \geq 1.$$

Hence,

$$N_{1,k} = N_{11} (A_0 N_{11})^{k-1} = N_{11} R^{k-1} \text{ for } k \geq 1.$$

**Proof**

We assume that $n \geq 0$ is arbitrary but fixed and that the initial state is in $\mathcal{E}(n+1)$. We define $Z_j$ as the number of visits to the state $(n + k + 1, j)$ before the first visit to any state in $\mathcal{E}(n)$ (or, equivalently, in $\mathcal{E}(0) \cup \cdots \cup \mathcal{E}(n)$). As seen above,

$$\langle N_{1k+1} \rangle_{ij} = \mathbb{E}[Z_j | X_0 = (n + 1, i)].$$
We denote by $\tau$ the epoch of the first visit to $\mathcal{A}(n)$ and by $0 = \theta_0 < \theta_1 < \theta_2 < \ldots < \tau$ the successive epochs of visit to $\mathcal{A}(n+1)$, before the first visit to $\mathcal{A}(n)$. We shall simply compute $(N_{1k+1})_{ij}$ by counting the number of visits to $(n + k + 1, j)$ during each interval $(\theta_v, \theta_{v+1})$ and the number of such intervals. In order to do this, we define the random variables $K_s$ as the number of visits to $(n + 1, s)$ before the first visit to $\mathcal{A}(n)$; we have that $E[K_s \mid X_0 = (n + 1, i)] = (N_{11})_s$.

We finally define the random variables $W_j$ as the number of visits to $(n + k + 1, j)$ before the first visit to either $\mathcal{A}(n+1)$ or $\mathcal{A}(n)$. Since $E[W_j \mid X_0 = (n + 2, i)] = (N_{1k})_{ij}$, we obtain by a simple argument that $E[W_j \mid X_0 = (n + 1, i)] = (A_0 N_{1k})_{ij}$.

By the strong Markov property, we have that

$$E[Z_j \mid X_0 = (n + 1, i)] =$$

$$2 \sum_{1 \leq s \leq m} E[K_s \mid X_0 = (n+1, i)] E[W_j \mid X_0 = (n+1, s)],$$

from which the theorem follows.

**Remark 1.3.4**

We have given in Theorem 1.3.1 one interpretation of the matrix $R$. We may now give one to its successive powers: we repeat the argument at the end of the proof above and find that for any $n \geq 0$, $k \geq 1$, $1 \leq i, j \leq m$, $(R^k)_{ij}$ is the expected number of visits to $(n+k, j)$ before the first return to $\mathcal{A}(0) \cup \mathcal{A}(1) \cup \ldots \cup \mathcal{A}(n)$ given that the process starts
in \((n, i)\). The quantity \((R^k)_{ij}\) may also be interpreted as the expected rate of visit to \((n+k, j)\) per unit of the local time of \(\alpha(n)\) spent in \((n, i)\).

The characterization, which we have obtained so far for the stationary distribution \(\pi\) is not very satisfactory. Indeed, it is expressed in terms of \(N = N_{11}\), i.e., one block in the infinite matrix \(N_T^c\), which is the minimal nonnegative solution of the infinite linear system \(N_T^c (I-P_T^c) = I\). We look therefore for other characterizations, better suited for the numerical evaluation of \(R\).

Let us assume that \(X_0\) is in \(\alpha(n)\). Define \(\tau\) as the first epoch of visit to the level \(\alpha(n-1)\) and \(\theta\) as the first epoch of return to the level \(\alpha(n)\). Let us define the matrices \(U\) and \(G\) as follows,

\[
U_{ij} = P[\theta < \tau \text{ and } X_\theta = (n, j) | X_0 = (n, i)];
\]

the matrix \(U\) records the probability, starting from \(\alpha(n)\), of returning to \(\alpha(n)\) before visiting \(\alpha(n-1)\);

\[
G_{ij} = P[\tau < \infty \text{ and } X_\tau = (n-1, j) | X_0 = (n, i)];
\]

the matrix \(G\) records the probability, starting from \(\alpha(n)\), of visiting \(\alpha(n-1)\) in a finite time. In view of the homogeneity of the matrix \(P\), the values of \(U\) and of \(G\) do not depend on \(n \geq 1\).

The three matrices \(R\), \(U\) and \(G\) are related by a collection of equations. For instance, by conditioning on the number of returns to \(\alpha(n)\) before the first visit to \(\alpha(n-1)\), we find that
\[ N = \sum_{i \geq 0} U^i = (I - U)^{-1} \]
and therefore that
\[ R = A_0 (I - U)^{-1} ; \]
similarly,
\[ G = \sum_{i \geq 0} U^i A_2 = (I - U)^{-1} A_2 \]

If we condition on the state after the first transition, we find that
\[ U = A_1 + A_0 G ; \]
indeed, in order to return to \( \xi(n) \), avoiding \( \xi(n-1) \), the QBD may either remain in \( \xi(n) \) at time 1, with probabilities recorded in \( A_1 \), or move up to \( \xi(n+1) \), with probabilities recorded in \( A_0 \); from \( \xi(n+1) \), the QBD returns to \( \xi(n) \) (necessarily before reaching \( \xi(n-1) \)) with probabilities recorded in \( G \).

This last equation may also be written as
\[
U = A_1 + A_0 (I - U)^{-1} A_2 \quad \text{.... (16)}
\]
\[ = A_1 + RA_2 \]

Thus we have the following result: If any one of the matrices \( U \), \( G \), or \( R \) is known, then we may determine the other two by applying one of the following equations:
\[
R = A_0 (I - U)^{-1},
\]
\[ G = (I - U)^{-1} A_2, \]
\[ U = A_1 + A_0 G, \]
\[ U = A_1 + RA_2. \]

We also observe that (16) only involves the unknown matrix \( U \) and might serve as a basis from which to actually compute \( U \) and then \( R \). That equation is not linear, but it
is more parsimonious than our earlier characterization of N from a linear but infinite system. We may also obtain similar equations for R and G as given below.

The three matrices U, G, and R, respectively, satisfy the following equations:

\[
U = A_1 + A_0 (I - U)^{-1} A_2 \quad \ldots \quad (17)
\]
\[
G = A_2 + A_1 G + A_0 G^2, \quad \ldots \quad (18)
\]
\[
R = A_0 + RA_1 + R^2 A_2. \quad \ldots \quad (19)
\]

**Remark 1.3.5**

These equations are not solely the result of formal manipulations, they also express interesting probabilistic properties.

The case of (18) is particularly obvious. The left-hand side records the distribution of the first state visited in \( \mathcal{E}(n-1) \), conditioned on the initial state being in \( \mathcal{E}(n) \); we see in the right-hand side a decomposition according to the first transition: the first term corresponds to the case where the QBD directly moves from \( \mathcal{E}(n) \) to \( \mathcal{E}(n-1) \) in one transition with probabilities recorded in \( A_2 \); as for the second term, with probabilities recorded in \( A_1 \), the QBD remains in \( \mathcal{E}(n) \), from where it still has to move eventually to \( \mathcal{E}(n-1) \), with probabilities recorded in \( G \); finally, for the last term, note that with probabilities recorded in \( A_0 \) the QBD moves up to \( \mathcal{E}(n+1) \), from where it must eventually return to \( \mathcal{E}(n) \), with probabilities recorded in \( G \) and then to \( \mathcal{E}(n-1) \), again with probabilities recorded in \( G \).
The interpretation of (19) is also simple. It is to be recalled that $R$ records the expected number of visits to $\mathcal{A}(n)$, starting from $\mathcal{A}(n-1)$, before the first return to $\mathcal{A}(n-1)$. In the right-hand side of (19), these visits to $\mathcal{A}(n)$ are decomposed into three groups according to the level from which a visit to $\mathcal{A}(n)$ occurs. There has to be a first visit immediately upon starting from $\mathcal{A}(n-1)$ with probability recorded in $A_0$, which explains the first term. Each visit to $\mathcal{A}(n)$ - the expected number is recorded in $R$-may immediately be followed by another visit to $\mathcal{A}(n)$ with probability recorded in $A_1$; this explains the second term. Finally, each visit to $\mathcal{A}(n+1)$ – with expected number recorded in $R^2$ may be followed by a visit to $\mathcal{A}(n)$ with probability recorded in $A_2$, which explains the third term.

The interpretation of (17) is immediate if we write it as $U = A_1 + \sum_{i \geq 0} A_0 U^i A_2$. Indeed, the term $A_1$ records the probability of immediately returning to $\mathcal{A}(n)$; the term $A_0 U^i A_2$ records the probability of returning to $\mathcal{A}(n)$ after exactly $i+1$ visits to $\mathcal{A}(n+1)$.

1.3.2. The PH/PH/1 queue

The PH/PH/1 queue is a single server system with renewal arrivals, where both the interarrival times and the service times have a PH distribution. In the example of Figure 1.3.1 the interarrival times have the Erlang distribution $F_{m,v}(\cdot)$ and the service time distribution has two phases: the first phase has parameter $\mu$; the second phase has parameter $\mu'$ and is performed with probability $q$. 
Fig. 1.3.1. Diagrammatic representation of a PH/PH/1 queue. The interarrival times have the Erlang distributions $F_{nv}$. The service time distribution has two phases: the first phase has parameter $\mu$; the second phase has parameter $\mu'$ and is performed with probability $q$.

This may still be represented as a Markov process $\{(N(t), \phi(t)), t \geq 0\}$ on the state space $U_{\geq 0} \times \mathbb{N}$, but the levels themselves become two dimensional here since we need to record $\phi(t) = (\phi_s(t), \phi_a(t))$, where $\phi_s(t)$ is the index of the node occupied by the customer in service, when the server is busy, while $\phi_a(t)$ is the position of the token in the arrival process. Thus, we have that $\xi(0) = \{(0, j), 1 \leq j \leq m\}$ and $\xi(n) = \{(n, i, j) : i = 1 \text{ or } 2, 1 \leq j \leq m\}$ for $n \geq 1$. The possible transitions are enumerated in Table 1.3.1.

**Table 1.3.1**

Transitions for the PH/PH/1 Queue of Figure 1.3.1

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Rate</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, j)</td>
<td>(0, j-1)</td>
<td>$\nu$</td>
<td>for $2 \leq j \leq m$</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1,1,m)</td>
<td>$\nu$</td>
<td></td>
</tr>
<tr>
<td>(1,1,j)</td>
<td>(0, j)</td>
<td>$\mu p$</td>
<td>for $1 \leq j \leq m$</td>
</tr>
<tr>
<td>(n, 1,j)</td>
<td>(n-1, 1, j)</td>
<td>$\mu p$</td>
<td>for $n \geq 2, 1 \leq j \leq m$</td>
</tr>
<tr>
<td>(n,1,j)</td>
<td>(n,2,j)</td>
<td>$\mu q$</td>
<td>for $n \geq 1, 1 \leq j \leq m$</td>
</tr>
<tr>
<td>(1,2,j)</td>
<td>(0, j)</td>
<td>$\mu'$</td>
<td>for $1 \leq j \leq m$</td>
</tr>
<tr>
<td>(n,2,j)</td>
<td>(n-1,1,j)</td>
<td>$\mu'$</td>
<td>for $n \geq 2, 1 \leq j \leq m$</td>
</tr>
<tr>
<td>(n,i,j)</td>
<td>(n,i,j-1)</td>
<td>$\nu$</td>
<td>for $n \geq 1, i = 1, 2, 2 \leq j \leq m$</td>
</tr>
<tr>
<td>(n,i,1)</td>
<td>(n+1,i,m)</td>
<td>$\nu$</td>
<td>for $n \geq 1, i = 1, 2$</td>
</tr>
</tbody>
</table>
The states are grouped by levels, so that \((n, i, j)\) precedes \((n', i', j')\) if \(n < n'\). Within each level, the states are grouped according to the service phase, so that \((n, 1, j)\) precedes \((n, 2, j')\) for all \(j, j'\). Finally, for a fixed level and a fixed service phase, the states are ordered by arrival phase, so that \((n, i, j)\) precedes \((n, i, j')\) if \(j < j'\).

We may verify that the infinitesimal generator \(Q\) has the form

\[
Q = \begin{pmatrix}
B_1 & B_0 & 0 & 0 & \ldots \\
B_2 & A_1 & A_0 & 0 & \ldots \\
0 & A_2 & A_1 & A_0 & \ldots \\
0 & 0 & A_2 & A_1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{pmatrix}
\]

Here, the matrix \(B_1\) is of order \(m\) and is equal to the matrix

\[
S = \begin{pmatrix}
-\nu & . & . & \ldots & . & . \\
\nu & -\nu & . & \ldots & . & . \\
. & \nu & -\nu & \ldots & . & . \\
. & . & \nu & -\nu & \ldots & . \\
. & . & . & \nu & -\nu & \ldots \\
. & . & . & . & \nu & -\nu \\
\end{pmatrix}
\]

The matrix \(B_0\) has \(m\) rows and \(2m\) columns and is equal to \([s.\sigma \quad 0]\), where

\[
s = \begin{pmatrix}
\nu \\
0 \\
. \\
. \\
\end{pmatrix}
\quad \text{and} \quad \sigma = [0, 0 \ldots 1],
\]
The matrix $B_2$ has $2m$ rows and $m$ columns and is given by

$$B_2 = \begin{bmatrix} \mu \phi \\ \mu' I \end{bmatrix}$$

The matrices $A_0$, $A_1$ and $A_2$ have order $2m$ and are given by

$$A_0 = \begin{bmatrix} s. \sigma & 0 \\ 2 & s. \sigma \end{bmatrix}$$

$$A_1 = \begin{bmatrix} -\mu I + S & \mu Q_i \\ 0 & -\mu' I + S \end{bmatrix}$$

$$A_2 = \begin{bmatrix} \mu \phi & 0 \\ \mu' I & 0 \end{bmatrix}$$

In a more compact form, using Kronecker products we may write that

$$B_1 = S,$$

$$B_0 = \tau \otimes s. \sigma,$$

$$B_2 = t \otimes I,$$

$$A_0 = I \otimes s. \sigma,$$

$$A_1 = T \otimes I + I \otimes S,$$

$$A_2 = t. \tau \otimes I.$$

**SECTION 1.4. NUMERICAL TECHNIQUES USED IN QUEUEING THEORY**

Consider the linear non-homogenous integral equation

$$\phi(x) = f(x) + \int_a^b K(x, t) \phi(t) dt$$

.... (20)
Approximating the definite integral in Equation (20) by a quadrature formula, equation (20) can then be written as

\[ \phi(x) = f(x) + (b-a) (c_1 K(x, t_1) \phi(t_1) + c_2 K(x, c_2) \phi(t_2) + \ldots + c_n K(x, t_n) \phi(t_n)), \quad \ldots \ (21) \]

where \( t_j, j = 1(1)n, \) are the subinterval points of the interval \([a, b]\), and \( c_j, j = 1(1)n, \) are the weighting coefficients known from the quadrature formula used. Since Equation (21) should be valid for all values of \( x \) in \([a, b]\), it should be true for \( x = t_j, j = 1(1)n. \) Thus, from Equation (21), we obtain

\[ \phi(t_j) = f(t_j) + (b-a) (c_1 K(t_j, t_1) \phi(t_1) + c_2 K(t_j, t_2) \phi(t_2) + \ldots + c_n K(t_j, t_n) \phi(t_n)), j = 1(1)n. \quad \ldots \ (22) \]

Set

\[ \phi(t_j) = \phi_j, \quad f(t_j) = f_j, \quad j = 1(1)n. \quad \ldots \ (23) \]

We then obtain

\[ \phi_1 = f_1 + (b-a) (c_1 K(t_1, t_1) \phi_1 + c_2 K(t_1, t_2) \phi_2 + \ldots + c_n K(t_1, t_n) \phi_n) \]

\[ \phi_2 = f_2 + (b-a) (c_1 K(t_2, t_1) \phi_1 + c_2 K(t_2, t_2) \phi_2 + \ldots + c_n K(t_2, t_n) \phi_n) \]

\[ \vdots \]

\[ \phi_n = f_n + (b-a) (c_1 K(t_n, t_1) \phi_1 + c_2 K(t_n, t_2) \phi_2 + \ldots + c_n K(t_n, t_n) \phi_n) \]

In Equation (24), all the quantities are known except \( \phi_j, j = 1(1)n. \) Put

\[ (b-a) c_j K(t_i, t_j) = d_{ij}, \quad i = 1(1)n, j = 1(1)n, \quad i \neq j, \]

\[ (b-a) c_j K(t_j, t_j) - 1 = d_{jj}, \quad j = 1(1)n, \quad \ldots \ (25) \]

\[ - f_j = c_j, \quad j = 1(1)n. \]

Equation (24) then becomes

\[ D \phi = e, \quad \ldots \ (26) \]
where

\[
D = \begin{pmatrix}
d_{11} & d_{12} & \cdots & d_{1n} \\
d_{21} & d_{22} & \cdots & d_{2n} \\
& \ddots & \ddots & \ddots \\
d_{n1} & d_{n2} & \cdots & d_{nn}
\end{pmatrix}, \quad \phi = \begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_n
\end{pmatrix}, \quad e = \begin{pmatrix}
e_1 \\
e_2 \\
\vdots \\
e_n
\end{pmatrix}
\]

\[\ldots (27)\]

It is to be noted that Equation (26) is a system of linear nonhomogenous equations. By solving Equation (26), we obtain \(\phi_j, j = 1(1)n\), or, equivalently, \(\phi(x)\) at \(x = t_j, j = 1(1)n\). If we are interested in obtaining a functional form of \(\phi(x)\), then a suitable curve has to be fitted.

Consider the homogenous integral equation

\[
\phi(x) = \lambda \int_a^b K(x, t) \phi(t) \, dt.
\]

\[\ldots (28)\]

Here, \(f(x)\) is absent or \(f(x) = 0\) for all \(x\) in \([a, b]\). Obviously, we can obtain, from Equation (24), an eigenvalue problem of the form

\[
D'\phi = \lambda \phi,
\]

\[\ldots (29)\]

where \(\lambda\) is an unknown scalar, called an eigenvalue of \(D'\) given as

\[
D' = \begin{pmatrix}
d'_{11} & d'_{12} & \cdots & d'_{1n} \\
d'_{21} & d'_{22} & \cdots & d'_{2n} \\
& \ddots & \ddots & \ddots \\
d'_{n1} & d'_{n2} & \cdots & d'_{nn}
\end{pmatrix}
\]

\[\ldots (30)\]

\[
d'_{ij} = (b - a) c_{ij} K(t_i, t_j), \quad i = 1(1)n, \quad j = 1(1)n.
\]

\[\ldots (31)\]

Since \(D'\) is completely known, we can compute the eigenvalues of \(D'\). Once an eigenvalue \(\lambda\) is known, we solve the homogenous system
and obtain a nontrivial solution for $\phi$.

The same technique can be used to find the stationary queueing time distribution of a single server queue. We have the integral equation of a single server queue as follows:

The stationary distribution of single-server queue-size can be obtained easily, when our assumptions involve the presence of randomness.

In many single-server queueing situations our convenient assumptions fail to hold. It may be that service-times have a quite general distribution $B(x)$, and that the intervals between arrivals of customers, while independent, have a general distribution $A(x)$. A particular example covered by this more general set-up arises when the intervals between arrivals are of constant length, i.e., when items arrive regularly at the service-point.

We call the general arrangement involving the distributions $A(x)$ and $B(x)$ the single-server queue with general independent arrivals and general service-time.

Let us write $C_n$ for the $n^{th}$ customer, $X_n$ for his queueing-time, $Y_n$ for his service-time, and $Z_n$ for the interval between the arrival of $C_n$ and the arrival of $C_{n+1}$. Then $X_n$, $Y_n$, $Z_n$ are independent quantities, as a moment’s reflection will show. Indeed, $Y_n$ and $Z_n$ are independent of the entire history of the queueing-process up to the arrival of $C_n$. The random variable $Y_n$ has the distribution $B(x)$, and $Z_n$ has the
distribution $A(x)$. We shall put $b_i = \varepsilon_n$ and $a_i = E Z_n$; thus $\rho$, the traffic intensity, is equal to $b_i/a_i$. We shall write $F_n(x)$ for the distribution of $X_n$. If $\rho < 1$, it is very reasonable to expect there to be a stationary distribution of queueing-time. The existence of the stationary distribution can be proved, but here we shall assume its existence and concentrate on the problem of its calculation. That is, we assume that as $n \to \infty$, $F_n(x)$ tends to a limit, $F(x)$, and consider the problem of calculating $F(x)$.

We see that $X_n$ and $X_{n+1}$ are related by the equation

$$X_{n+1} = X_n + Y_n - Z_n,$$

provided that the quantity on the right is positive. If $X_n + Y_n - Z_n$ is negative, the customer $C_{n+1}$ arrives to find the server free, and so then we have, simply $X_{n+1} = 0$. In general

$$X_{n+1} = \max(X_n + Y_n - Z_n, 0).$$

The numerical value of $X_n + Y_n - Z_n$, when it is algebraically negative, measures the time during which the server was free between the serving of $C_n$ and the arrival of $C_{n+1}$. In either event, whether $X_n + Y_n - Z_n$ is positive or negative, it should be clear that provided that $x \geq 0$ we have

$$\text{Prob} \{ X_{n+1} \leq x \} = \text{Prob} \{ X_n + Y_n - Z_n \leq x \}. \quad \ldots (33)$$

To proceed further, let us put

$$K(x) = \text{Prob} \{ Y_n - Z_n < x \}$$

for the distribution function of the quantity $Y_n - Z_n$. Evidently

$$K(x) = \int_0^\infty B(x + z) ' (z),$$

so that $K(x)$ is in principle calculable from known distributions. Further, since $X_n$ and $Y_n - Z_n$ are independent, the distribution function of $X_n + Y_n - Z_n$ is
Thus (33) is telling us that for all $x \geq 0$ we must have

$$F_{n+1}(x) = \int_0^\infty K(x-y) Df_n(y).$$

If we let $n \to \infty$ in this equation we have that $F(x)$, the stationary queueing-time distribution, satisfies the equation

$$F(x) = \int_0^\infty K(x-y) Df(y), \text{ for all } x > 0 \quad \text{.... (34)}$$

This equation may be called the integral equation of the queue. It is of the Wiener-Hopf type. With the proper choice of $k(x)$ and $f(x) = 0$, we can solve for the stationary queueing time distribution from the above equation (34).

**Illustration 14.1**

Solve the Fredholm integral equation

$$\phi(x) = 1 + \lambda \int_0^1 (1 - 3xt) \phi(t), \text{ dt, } \lambda < 2,$$

for $\lambda = 1$.

Replacing

$$\int_0^1 (1 - 3xt) \phi(t) \text{ dt}$$

by the Simpson 1/3 rule and setting $x = t_j$, $j = 1 (1)3$, we obtain

$$\phi(t_j) = 1 + \frac{(1 - 0)}{6} \left( (1-3t_j \times 0) \phi(0) + 4(1-3t_j \times ½) \phi(1/2) \right)$$
\begin{align*}
+ (1 - 3t_j x 1) \phi(1)), \quad j = 1(1)3,
\end{align*}
\begin{align*}
t_1 &= 0, \quad t_2 = \frac{1}{2} \quad t_3 = 1.
\end{align*}

Let
\begin{align*}
\frac{1}{6} (1-3t_1 x 0) - 1 &= -\frac{5}{6} = d_{11}, \\
\frac{1}{6} x 4(1-3t_1 x \frac{1}{2}) &= 2/3 = d_{12}, \\
\frac{1}{6} (1-3t_1 x 1) &= 1/6 = d_{13}, \\
\frac{1}{6} (1-3t_2 x 0) &= 1/6 = d_{21}, \\
\frac{1}{6} x 4(1-3t_2 x \frac{1}{2}) - 1 &= 2/3 (1-3/4) - 1 = -5/6 = d_{22}, \\
\frac{1}{6} x (1-3t_2 x 1) &= 1/6 x -1/2 = -1/12 = d_{23}, \\
\frac{1}{6} (1-3t_3 x 0) &= 1/6 = d_{31}, \\
\frac{1}{6} x 4(1-3t_3 x \frac{1}{2}) &= -1/3 = d_{32}, \\
\frac{1}{6} (1-3t_3 x 1) - 1 &= -4/3 = d_{33}.
\end{align*}

Also, let
\begin{align*}
e_1 &= -1, \quad e_2 = -1, \quad e_3 = -1, \quad \phi = [\phi_1 \quad \phi_2 \quad \phi_3]^T,
\end{align*}
where
\begin{align*}
\phi_1 &= \phi(0), \quad \phi_2 = \phi(\frac{1}{2}), \quad \phi_3 = \phi(1)
\end{align*}

Then, \(D \phi = e\) or
\begin{align*}
\begin{pmatrix}
-5/6 & 2/3 & 1/6 \\
1/6 & -5/6 & -1/12 \\
1/6 & -1/3 & -4/3
\end{pmatrix}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3
\end{pmatrix}
= \begin{pmatrix}
-1 \\
1 \\
-1
\end{pmatrix}
\end{align*}
or, multiplying both sides by 12, we obtain

\[
\begin{pmatrix}
-10 & 8 & 2 \\
2 & -10 & -1 \\
2 & -4 & -16
\end{pmatrix}
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3
\end{pmatrix}
= 
\begin{pmatrix}
-12 \\
-12 \\
-12
\end{pmatrix}
\]

Hence, \( \phi_1 = \phi(0) = 8/3, \ \phi_2 = \phi(1/2) = 5/3, \ \phi_3 = \phi(1) = 2/3. \) This result tallies exactly with that obtained from the actual solution

\[
\phi(x) = \frac{(4 + 2\lambda(2-3x))/(4-\lambda^2)}
\]

for \( \lambda = 1. \)