## 1

## Preliminaries

Queueing theory is an intricate and yet highly practical field of mathematical study that has vast applications in performance evaluation. It is a subject usually taught at the advanced stage of an undergraduate programme or the entry level of a postgraduate course in Computer Science or Engineering. To fully understand and grasp the essence of the subject, students need to have certain background knowledge of other related disciplines, such as probability theory and transform theory, as a prerequisite.

It is not the intention of this chapter to give a fine exposition of each of the related subjects but rather meant to serve as a refresher and highlight some basic concepts and important results in those related topics. These basic concepts and results are instrumental to the understanding of queueing theory that is outlined in the following chapters of the book. For more detailed treatment of each subject, students are directed to some excellent texts listed in the references.

### 1.1 PROBABILITY THEORY

In the study of a queueing system, we are presented with a very dynamic picture of events happening within the system in an apparently random fashion. Neither do we have any knowledge about when these events will occur nor are we able to predict their future developments with certainty. Mathematical models have to be built and probability distributions used to quantify certain parameters in order to render the analysis mathematically tractable. The

[^0]importance of probability theory in queueing analysis cannot be overemphasized. It plays a central role as that of the limiting concept to calculus. The development of probability theory is closely related to describing randomly occurring events and has its roots in predicting the random outcome of playing games. We shall begin by defining the notion of an event and the sample space of a mathematical experiment which is supposed to mirror a real-life phenomenon.

### 1.1.1 Sample Spaces and Axioms of Probability

A sample space $(\Omega)$ of a random experiment is a collection of all the mutually exclusive and exhaustive simple outcomes of that experiment. A particular simple outcome $(\omega)$ of an experiment is often referred to as a sample point. An event $(E)$ is simply a subset of $\Omega$ and it contains a set of sample points that satisfy certain common criteria. For example, an event could be the even numbers in the toss of a dice and it contains those sample points \{[2], [4], [6]\}. We indicate that the outcome $\omega$ is a sample point of an event $E$ by writing $\{\omega \in E\}$. If an event $E$ contains no sample points, then it is a null event and we write $E=\varnothing$. Two events $E$ and $F$ are said to be mutually exclusive if they have no sample points in common, or in other words the intersection of events $E$ and $F$ is a null event, i.e. $E \cap F=\varnothing$.

There are several notions of probability. One of the classic definitions is based on the relative frequency approach in which the probability of an event $E$ is the limiting value of the proportion of times that $E$ was observed. That is

$$
\begin{equation*}
P(E)=\lim _{N \rightarrow \infty} \frac{N_{E}}{N} \tag{1.1}
\end{equation*}
$$

where $N_{E}$ is the number of times event $E$ was observed and $N$ is the total number of observations. Another one is the so-called axiomatic approach where the probability of an event $E$ is taken to be a real-value function defined on the family of events of a sample space and satisfies the following conditions:

Axioms of probability
(i) $0 \leq P(E) \leq 1$ for any event in that experiment
(ii) $P(\Omega)=1$
(iii) If $E$ and $F$ are mutually exclusive events, i.e. $E \in F=\varnothing$, then $P(E \cup F)$ $=P(E)+P(F)$


Figure 1.1 A closed loop of $M$ queues

There are some fundamental results that can be deduced from this axiomatic definition of probability and we summarize them without proofs in the following propositions.

## Proposition 1.1

(i) $P(\varnothing)=0$
(ii) $P(\bar{E})+P(E)=1$ for any event $E$ in $\Omega$, where $\bar{E}$ is the compliment of $E$.
(iii) $P(E \cup F)=P(E)+P(F)-P(E \cap F)$, for any events $E$ and $F$.
(iv) $P(E) \leq P(F)$, if $E \subseteq F$.
(v) $P\left(\bigcup_{i} E_{i}\right)=\sum_{i} P\left(E_{i}\right)$, for $E_{i} \cap E_{j}=\varnothing$, when $i \neq j$.

## Example 1.1

By considering the situation where we have a closed loop of $M$ identical queues, as shown in Figure 1.1, then calculate the probability that Queue 1 is non-empty (it has at least one customer) if there are $N$ customers circulating among these queues.

## Solution

To calculate the required probability, we need to find the total number of ways of distributing those $N$ customers among $M$ queues. Let $X_{i}(>0)$ be the number of customers in Queue $i$, then we have

$$
X_{1}+X_{2}+\ldots+X_{M}=N
$$

The problem can now be formulated by having these $N$ customers lined up together with $M$ imaginary zeros, and then dividing them into $M$ groups. These $M$ zeros are introduced so that we may have empty queues. They also ensure that one of the queues will contain all the customers, even in the case where


Figure 1.2 $N$ customers and $M$ zeros, $(N+M-1)$ spaces
all zeros are consecutive because there are only $(M-1)$ spaces among them, as shown in Figure 1.2.

We can select $M-1$ of the $(N+M-1)$ spaces between customers as our separating points and hence the number of combinations is given by

$$
\binom{N+M-1}{M-1}=\binom{N+M-1}{N}
$$

When Queue 1 is empty, the total number of ways of distributing $N$ customers among $(M-1)$ queues is given by

$$
\binom{N+M-2}{N}
$$

Therefore, the probability that Queue 1 is non-empty:

$$
\begin{aligned}
& =1-\binom{N+M-2}{N} /\binom{N+M-1}{N} \\
& =1-\frac{M-1}{N+M-1}=\frac{N}{N+M-1}
\end{aligned}
$$

## Example 1.2

Let us suppose a tourist guide likes to gamble with his passengers as he guides them around the city on a bus. On every trip, there are about 50 random passengers. Each time he challenges his passengers by betting that if there is at least two people on the bus that have the same birthday, then all of them would have to pay him $\$ 1$ each. However, if there were none for that group on that day, he would repay each of them $\$ 1$. What is the likelihood (or probability) of the event that he wins his bet?

## Solution

Let us assume that each passenger is equally likely to have their birthday on any day of the year (we will neglect leap years). In order to solve this problem
we need to find the probability that nobody on that bus has the same birthday. Imagine that we line up these 50 passengers, and the first passenger has 365 days to choose as his/her birthday. The next passenger has the remainder of 364 days to choose from in order for him/her not to have the same birthday as the first person (i.e. he has a probability of $364 / 365$ ). This number of choices reduces until the last passenger. Therefore:

$$
\begin{aligned}
& P(\text { None of the } 50 \text { passengers has the same birthday })= \\
& \underbrace{\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right) \ldots\left(\frac{365-49}{365}\right)}_{49 t e r m s}
\end{aligned}
$$

Therefore, the probability that the tourist guide wins his bet can be obtained by Proposition 1.1 (ii):

$$
\begin{aligned}
& P(\text { At least } 2 \text { passengers out of } 50 \text { has the same birthday })= \\
& \quad 1-\left(\frac{\prod_{j=1}^{49}(365-j)}{365^{49}}\right)=0.9704 .
\end{aligned}
$$

The odds are very much to the favour of tourist guide, although we should remember this probability has a limiting value of (1.1) only.

### 1.1.2 Conditional Probability and Independence

In many practical situations, we often do not have information about the outcome of an event but rather information about related events. Is it possible to infer the probability of an event using the knowledge that we have about these other events? This leads us to the idea of conditional probability that allows us to do just that!

Conditional probability that an event $E$ occurs, given that another event $F$ has already occurred, denoted by $P(E \mid F)$, is defined as

$$
\begin{equation*}
P(E \mid F)=\frac{P(E \cap F)}{P(F)} \quad \text { where } \quad P(F) \neq 0 \tag{1.2}
\end{equation*}
$$

Conditional probability satisfies the axioms of probability and is a probability measure in the sense of those axioms. Therefore, we can apply any results obtained for a normal probability to a conditional probability. A very useful expression, frequently used in conjunction with the conditional probability, is the so-called Law of Total Probability. It says that if $\left\{A_{i} \in \Omega, i=1,2, \ldots, n\right\}$ are events such that
(i) $A_{i} \cap A_{j}=\varnothing$ if $i \neq j$
(ii) $P\left(A_{i}\right)>0$
(iii) $\bigcup_{i=1}^{n} A_{i}=\Omega$
then for any event $E$ in the same sample space:

$$
\begin{equation*}
P(E)=\sum_{i=1}^{n} P\left(E \cap A_{i}\right)=\sum_{i=1}^{n} P\left(E \mid A_{i}\right) P\left(A_{i}\right) \tag{1.3}
\end{equation*}
$$

This particular law is very useful for determining the probability of a complex event $E$ by first conditioning it on a set of simpler events $\{A i\}$ and then by summing up all the conditional probabilities. By substituting the expression (1.3) in the previous expression of conditional probability (1.2), we have the well-known Bayes' formula:

$$
\begin{equation*}
P(E \mid F)=\frac{P(E \cap F)}{\sum_{i} P\left(F \mid A_{i}\right) P\left(A_{i}\right)}=\frac{P(F \mid E) P(E)}{\sum_{i} P\left(F \mid A_{i}\right) P\left(A_{i}\right)} \tag{1.4}
\end{equation*}
$$

Two events are said to be statistically independent if and only if

$$
P(E \cap F)=P(E) P(F)
$$

From the definition of conditional probability, this also implies that

$$
\begin{equation*}
P(E \mid F)=\frac{P(E \cap F)}{P(F)}=\frac{P(E) P(F)}{P(F)}=P(E) \tag{1.5}
\end{equation*}
$$

Students should note that the statistical independence of two events $E$ and $F$ does not imply that they are mutually exclusive. If two events are mutually exclusive then their intersection is a null event and we have

$$
\begin{equation*}
P(E \mid F)=\frac{P(E \cap F)}{P(F)}=0 \quad \text { where } \quad P(F) \neq 0 \tag{1.6}
\end{equation*}
$$

## Example 1.3

Consider a switching node with three outgoing links A, B and C. Messages arriving at the node can be transmitted over one of them with equal probability. The three outgoing links are operating at different speeds and hence message transmission times are 1,2 and 3 ms , respectively for $\mathrm{A}, \mathrm{B}$ and C. Owing to
the difference in trucking routes, the probability of transmission errors are 0.2 , 0.3 and 0.1 , respectively for A, B and C. Calculate the probability of a message being transmitted correctly in 2 ms .

## Solution

Denote the event that a message is transmitted correctly by $F$, then we are given

$$
\begin{aligned}
& P(F \mid A \text { Link })=1-0.2=0.8 \\
& P(F \mid B \text { Link })=1-0.3=0.7 \\
& P(F \mid C \text { Link })=1-0.1=0.9
\end{aligned}
$$

The probability that a message being transmitted correctly in 2 ms is simply the event $(F \cap B)$, hence we have

$$
\begin{aligned}
P(F \cap B) & =P(F \mid B) \times P(B) \\
& =0.7 \times \frac{1}{3}=\frac{7}{30}
\end{aligned}
$$

### 1.1.3 Random Variables and Distributions

In many situations, we are interested in some numerical value that is associated with the outcomes of an experiment rather than the actual outcomes themselves. For example, in an experiment of throwing two die, we may be interested in the sum of the numbers $(X)$ shown on the dice, say $X=5$. Thus we are interested in a function which maps the outcomes onto some points or an interval on the real line. In this example, the outcomes are $\{2,3\},\{3,2\},\{1,4\}$ and $\{4,1\}$, and the point on the real line is 5 .

This mapping (or function) that assigns a real value to each outcome in the sample space is called a random variable. If $X$ is a random variable and $x$ is a real number, we usually write $\{X \leq x\}$ to denote the event $\{\omega \in \Omega$ and $\mathrm{X}(\omega) \leq \mathrm{x}\}$. There are basically two types of random variables; namely the discrete random variables and continuous random variables. If the mapping function assigns a real number, which is a point in a countable set of points on the real line, to an outcome then we have a discrete random variable. On the other hand, a continuous random variable takes on a real number which falls in an interval on the real line. In other words, a discrete random variable can assume at most a finite or a countable infinite number of possible values and a continuous random variable can assume any value in an interval or intervals of real numbers.

A concept closely related to a random variable is its cumulative probability distribution function, or just distribution function (PDF). It is defined as

$$
\begin{align*}
F_{X}(x) & \equiv P[X \leq x] \\
& =P[\omega: X(\omega) \leq x] \tag{1.7}
\end{align*}
$$

For simplicity, we usually drop the subscript $X$ when the random variable of the function referred to is clear in the context. Students should note that a distribution function completely describes a random variable, as all parameters of interest can be derived from it. It can be shown from the basic axioms of probability that a distribution function possesses the following properties:

Proposition 1.2
(i) $F$ is a non-negative and non-decreasing function, i.e. if $x_{1} \leq x_{2}$ then $F\left(x_{1}\right) \leq F\left(x_{2}\right)$
(ii) $F(+\infty)=1 \& F(-\infty)=0$
(iii) $F(b)-F(a)=P[a<X \leq b]$

For a discrete random variable, its probability distribution function is a disjoint step function, as shown in Figure 1.3. The probability that the random variable takes on a particular value, say $x$ and $x=0,1,2,3 \ldots$, is given by

$$
\begin{align*}
p(x) \equiv P[X=x] & =P[X<x+1]-P[X<x] \\
& =\{1-P[X \geq x+1]\}-\{1-P[X \geq x]\} \\
& =P[X \geq x]-P[X \geq x+1] \tag{1.8}
\end{align*}
$$

The above function $p(x)$ is known as the probability mass function ( $p m f$ ) of a discrete random variable $X$ and it follows the axiom of probability that


Figure 1.3 Distribution function of a discrete random variable $X$


Figure 1.4 Distribution function of a continuous RV

$$
\begin{equation*}
\sum_{x} p(x)=1 \tag{1.9}
\end{equation*}
$$

This probability mass function is a more convenient form to manipulate than the $P D F$ for a discrete random variable.

In the case of a continuous random variable, the probability distribution function is a continuous function, as shown in Figure 1.4, and pmf loses its meaning as $P[X=x]=0$ for all real $x$.

A new useful function derived from the $P D F$ that completely characterizes a continuous random variable $X$ is the so-called probability density function ( $p d f$ ) defined as

$$
\begin{equation*}
f_{X}(x) \equiv \frac{d}{d x} F_{X}(x) \tag{1.10}
\end{equation*}
$$

It follows from the axioms of probability and the definition of $p d f$ that

$$
\begin{gather*}
F_{X}(x)=\int_{-\infty}^{x} f_{X}(\tau) d \tau  \tag{1.11}\\
P[a \leq X \leq b]=\int_{a}^{b} f_{X}(x) d x \tag{1.12}
\end{gather*}
$$

and

$$
\begin{equation*}
\int_{-\infty}^{\infty} f_{X}(x)=P[-\infty<X<\infty]=1 \tag{1.13}
\end{equation*}
$$

The expressions (1.9) and (1.13) are known as the normalization conditions for discrete random variables and continuous random variables, respectively.

We list in this section some important discrete and continuous random variables which we will encounter frequently in our subsequent studies of queueing models.

## (i) Bernoulli random variable

A Bernoulli trial is a random experiment with only two outcomes, 'success' and 'failure', with respective probabilities, $p$ and $q$. A Bernoulli random variable $X$ describes a Bernoulli trial and assumes only two values: 1 (for success) with probability $p$ and 0 (for failure) with probability $q$ :

$$
\begin{equation*}
P[X=1]=p \quad \& \quad P[X=0]=q=1-p \tag{1.14}
\end{equation*}
$$

(ii) Binomial random variable

If a Bernoulli trial is repeated $k$ times then the random variable $X$ that counts the number of successes in the $k$ trials is called a binomial random variable with parameters $k$ and $p$. The probability mass function of a binomial random variable is given by

$$
\begin{equation*}
B(k ; n, p)=\binom{n}{k} p^{k} q^{n-k} \quad k=0,1,2, \ldots, n \quad \& \quad q=1-p \tag{1.15}
\end{equation*}
$$

## (iii) Geometric random variable

In a sequence of independent Bernoulli trials, the random variable $X$ that counts the number of trials up to and including the first success is called a geometric random variable with the following pmf:

$$
\begin{equation*}
P[X=k]=(1-p)^{k-1} p \quad k=1,2, \ldots \infty \tag{1.16}
\end{equation*}
$$

## (iv) Poisson random variable

A random variable $X$ is said to be Poisson random variable with parameter $\lambda$ if it has the following mass function:

$$
\begin{equation*}
P[X=k]=\frac{\lambda^{k}}{k!} e^{-\lambda} \quad k=0,1,2, \ldots \tag{1.17}
\end{equation*}
$$

Students should note that in subsequent chapters, the Poisson mass function is written as

$$
\begin{equation*}
P[X=k]=\frac{\left(\lambda^{\prime} t\right)^{k}}{k!} e^{-\lambda^{\prime} t} \quad k=0,1,2, \ldots \tag{1.18}
\end{equation*}
$$

Here, the $\lambda$ in expression (1.17) is equal to the $\lambda^{\prime} t$ in expression (1.18).

## (v) Continuous uniform random variable

A continuous random variable $X$ with its probabilities distributed uniformly over an interval ( $\mathrm{a}, \mathrm{b}$ ) is said to be a uniform random variable and its density function is given by

$$
f(x)=\left\{\begin{array}{cl}
\frac{1}{b-a} & a<x<b  \tag{1.19}\\
0 & \text { otherwise }
\end{array}\right.
$$

The corresponding distribution function can be easily calculated by using expression (1.11) as

$$
F(x)=\left\{\begin{array}{cc}
0 & x<a  \tag{1.20}\\
\frac{x-a}{b-a} & a \leq x<b \\
1 & x \geq b
\end{array}\right.
$$

(vi) Exponential random variable

A continuous random variable $X$ is an exponential random variable with parameter $\lambda>0$, if its density function is defined by

$$
f(x)=\left\{\begin{array}{cc}
\lambda e^{-\lambda x} & x>0  \tag{1.21}\\
0 & x \leq 0
\end{array}\right.
$$

The distribution function is then given by

$$
F(x)=\left\{\begin{array}{cc}
1-\lambda e^{-\lambda x} & x>0  \tag{1.22}\\
0 & x \leq 0
\end{array}\right.
$$

(vii) Gamma random variable

A continuous random variable $X$ is said to have a gamma distribution with parameters $\alpha>0$ and $\lambda>0$, if its density function is given by

$$
f(x)=\left\{\begin{array}{cc}
\frac{\lambda^{\alpha}(x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x>0  \tag{1.23}\\
0 & x \leq 0
\end{array}\right.
$$

where $\Gamma(\alpha)$ is the gamma function defined by

$$
\begin{equation*}
\Gamma(\alpha)=\int_{0}^{\infty} x^{\alpha-1} e^{-x} d x \quad \alpha>0 \tag{1.24}
\end{equation*}
$$

There are certain nice properties about gamma functions, such as

$$
\begin{align*}
& \Gamma(k)=(k-1) \Gamma(k-1)=(k-1)!\alpha=n \text { a positive integer } \\
& \Gamma(\alpha)=\alpha \Gamma(\alpha-1) \alpha>0 \text { a real number } \tag{1.25}
\end{align*}
$$

## (viii) Erlang-k or $\boldsymbol{k}$-stage Erlang Random Variable

This is a special case of the gamma random variable when $\alpha(=k)$ is a positive integer. Its density function is given by

$$
f(x)=\left\{\begin{array}{cc}
\frac{\lambda^{k}(x)^{k-1}}{(k-1)!} e^{-\lambda x} & x>0  \tag{1.26}\\
0 & x \leq 0
\end{array}\right.
$$

## (ix) Normal (Gaussian) Random Variable

A frequently encountered continuous random variable is the Gaussian or Normal with the parameters of $\mu$ (mean) and $\sigma_{X}$ (standard deviation). It has a density function given by

$$
\begin{equation*}
f_{X}(x)=\frac{1}{\sqrt{2 \pi \sigma_{X}^{2}}} e^{-(x-\mu)^{2} / 2 \sigma_{X}^{2}} \tag{1.27}
\end{equation*}
$$

The normal distribution is often denoted in a short form as $N\left(\mu, \sigma_{X}^{2}\right)$.
Most of the examples above can be roughly separated into either continuous or discrete random variables. A discrete random variable can take on only a finite number of values in any finite observations (e.g. the number of heads obtained in throwing 2 independent coins). On the other hand, a continuous random variable can take on any value in the observation interval (e.g. the time duration of telephone calls). However, samples may exist, as we shall see later, where the random variable of interest is a mixed random variable, i.e. they have both continuous and discrete portions. For example, the waiting time distribution function of a queue in Section 4.3 can be shown as

$$
\begin{aligned}
F_{W}(t) & =\left(1-\rho e^{-\mu(1-\rho) t}\right) & & t \geq 0 \\
& =0 & & t<0 .
\end{aligned}
$$

This has a discrete portion that has a jump at $t=0$ but with a continuous portion elsewhere.

### 1.1.4 Expected Values and Variances

As discussed in Section 1.1.3, the distribution function or $p m f$ ( $p d f$, in the case of continuous random variables) provides a complete description of a random
variable. However, we are also often interested in certain measures which summarize the properties of a random variable succinctly. In fact, often these are the only parameters that we can observe about a random variable in real-life problems.

The most important and useful measures of a random variable $X$ are its expected value ${ }^{1} E[\mathrm{X}]$ and variance $\operatorname{Var}[\mathrm{X}]$. The expected value is also known as the mean value or average value. It gives the average value taken by a random variable and is defined as

$$
\begin{equation*}
E[X]=\sum_{k=0}^{\infty} k P[X=k] \quad \text { for discrete variables } \tag{1.28}
\end{equation*}
$$

and

$$
\begin{equation*}
E[X]=\int_{0}^{\infty} x f(x) d x \quad \text { for continuous variables } \tag{1.29}
\end{equation*}
$$

The variance is given by the following expressions. It measures the dispersion of a random variable $X$ about its mean $E[X]$ :

$$
\begin{align*}
\sigma^{2} & =\operatorname{Var}[X] \\
& =E\left[(X-E[X])^{2}\right] \quad \text { for discrete variables }  \tag{1.30}\\
& =E\left[X^{2}\right]-(E[X])^{2} \\
\sigma^{2}= & \operatorname{Var}[X] \\
= & \int_{0}^{\infty}(x-E[X])^{2} f(x) d x \quad \text { for continuous variables } \\
= & \int_{0}^{\infty} x^{2} f(x) d x-2 E[X] \int_{0}^{\infty} x f(x) d x+\int_{0}^{\infty} f(x) d x \\
= & E\left[X^{2}\right]-(E[X])^{2} \quad \tag{1.31}
\end{align*}
$$

$\sigma$ refers to the square root of the variance and is given the special name of standard deviation.

Example 1.4
For a discrete random variable $X$, show that its expected value is also given by

$$
E[X]=\sum_{k=0}^{\infty} P[X>k]
$$

[^1]
## Solution

By definition, the expected value of $X$ is given by

$$
\begin{aligned}
E[X]= & \sum_{k=0}^{\infty} k P[X=k]=\sum_{k=0}^{\infty} k\{P[X \geq k]-P[X \geq k+1]\} \\
= & P[X \geq 1]-[X \geq 2]+2 P[X \geq 2]-2 P[X \geq 3] \\
& +3 P[X \geq 3]-3 P[X \geq 4]+4 P[X \geq 4]-4 P[X \geq 5]+\ldots \\
= & \sum_{k=1}^{\infty} P[X \geq k]=\sum_{k=0}^{\infty} P[X>k]
\end{aligned}
$$

Example 1.5
Calculate the expected values for the Binomial and Poisson random variables.

Solution

1. Binomial random variable

$$
\begin{aligned}
E[X] & =\sum_{k=1}^{n} k\binom{n}{k} p^{k}(1-p)^{n-k} \\
& =n p \sum_{k=1}^{n}\binom{n-1}{k-1} p^{k-1}(1-P)^{n-k} \\
& =n p \sum_{j=0}^{n}\binom{n-1}{j} p^{j}(1-p)^{(n-1)-j} \\
& =n p
\end{aligned}
$$

2. Poisson random variable

$$
\begin{aligned}
E[X] & =\sum_{k=0}^{\infty} k \frac{\lambda^{k}}{k!} e^{-\lambda} \\
& =e^{-\lambda}(\lambda) \sum_{k=1}^{\infty} \frac{(\lambda)^{k-1}}{(k-1)!} \\
& =e^{-\lambda}(\lambda) e^{\lambda} \\
& =\lambda
\end{aligned}
$$

Table 1.1 Means and variances of some common random variables

| Random variable | $\boldsymbol{E}[\boldsymbol{X}]$ | $\boldsymbol{V a r}[\boldsymbol{X}]$ |
| :--- | :--- | :--- |
| Bernoulli | $p$ | $p q$ |
| Binomial | $n p$ | $n p q$ |
| Geometric | $1 / p$ | $q / p^{2}$ |
| Poisson | $\lambda$ | $\lambda$ |
| Continuous uniform | $(a+b) / 2$ | $(b-a)^{2} / 12$ |
| Exponential | $1 / \lambda$ | $1 / \lambda^{2}$ |
| Gamma | $\alpha / \lambda$ | $\alpha / \lambda^{2}$ |
| Erlang- $k$ | $1 / \lambda$ | $1 / k \lambda^{2}$ |
| Gaussian | $\mu$ | $\sigma_{X}^{2}$ |

Table 1.1 summarizes the expected values and variances for those random variables discussed earlier.

## Example 1.6

Find the expected value of a Cauchy random variable $X$, where the density function is defined as

$$
f(x)=\frac{1}{\pi\left(1+x^{2}\right)} u(x)
$$

where $u(x)$ is the unit step function.

## Solution

Unfortunately, the expected value of $E[X]$ in this case is

$$
E[X]=\int_{-\infty}^{\infty} \frac{x}{\pi\left(1+x^{2}\right)} u(x) d x=\int_{0}^{\infty} \frac{x}{\pi\left(1+x^{2}\right)} d x=\infty
$$

Sometimes we get unusual results with expected values, even though the distribution of the random variable is well behaved.

Another useful measure regarding a random variable is the coefficient of variation which is the ratio of standard deviation to the mean of that random variable:

$$
C_{x} \equiv \frac{\sigma_{X}}{E[X]}
$$

### 1.1.5 Joint Random Variables and Their Distributions

In many applications, we need to investigate the joint effect and relationships between two or more random variables. In this case we have the natural extension of the distribution function to two random variables $X$ and $Y$, namely the joint distribution function. Given two random variables $X$ and $Y$, their joint distribution function is defined as

$$
\begin{equation*}
F_{X Y}(x, y) \equiv P[X \leq x, Y \leq y] \tag{1.32}
\end{equation*}
$$

where $x$ and $y$ are two real numbers. The individual distribution function $F_{X}$ and $F_{Y}$, often referred to as the marginal distribution of $X$ and $Y$, can be expressed in terms of the joint distribution function as

$$
\begin{align*}
& F_{X}(x)=F_{X Y}(x, \infty) \\
& F_{Y}(y)=P[X \leq x, Y \leq \infty]  \tag{1.33}\\
& X Y \\
&(\infty, y)=P[X \leq \infty, Y \leq y]
\end{align*}
$$

Similar to the one-dimensional case, the joint distribution also enjoys the following properties:
(i) $\quad F_{X Y}(-\infty, y)=F_{X Y}(x,-\infty)=0$
(ii) $F_{X Y}(-\infty,-\infty)=0$ and $F_{X Y}(\infty, \infty)=1$
(iii) $F_{X Y}\left(x_{1}, y\right) \leq F_{X Y}\left(x_{2}, y\right)$ for $x_{1} \leq x_{2}$
(iv) $F_{X Y}\left(x, y_{1}\right) \leq F_{X Y}\left(x, y_{2}\right)$ for $y_{1} \leq y_{2}$
(v) $P\left[x_{1}<X \leq x_{2}, y_{1}<Y \leq y_{2}\right]=F_{X Y}\left(x_{2}, y_{2}\right)-F_{X Y}\left(x_{1}, y_{2}\right)$

$$
-F_{X Y}\left(x_{2}, y_{1}\right)+F_{X Y}\left(x_{1}, y_{1}\right)
$$

If both $X$ and $Y$ are jointly continuous, we have the associated joint density function defined as

$$
\begin{equation*}
f_{X Y}(x, y) \equiv \frac{d^{2}}{d x d y} F_{X Y}(x, y) \tag{1.34}
\end{equation*}
$$

and the marginal density functions and joint probability distribution can be computed by integrating over all possible values of the appropriate variables:

$$
\begin{align*}
& f_{X}(x)=\int_{-\infty}^{\infty} f_{X Y}(x, y) d y \\
& f_{Y}(y)=\int_{-\infty}^{\infty} f_{X Y}(x, y) d x \tag{1.35}
\end{align*}
$$

$$
F_{X Y}(x, y)=\int_{-\infty}^{x} \int_{-\infty}^{y} f_{X Y}(u, v) d u d v
$$

If both are jointly discrete then we have the joint probability mass function defined as

$$
\begin{equation*}
p(x, y) \equiv P[X=x, Y=y] \tag{1.36}
\end{equation*}
$$

and the corresponding marginal mass functions can be computed as

$$
\begin{align*}
& P[X=x]=\sum_{y} p(x, y) \\
& P[Y=y]=\sum_{x} p(x, y) \tag{1.37}
\end{align*}
$$

With the definitions of joint distribution and density function in place, we are now in a position to extend the notion of statistical independence to two random variables. Basically, two random variables $X$ and $Y$ are said to be statistically independent if the events $\{x \in E\}$ and $\{y \in F\}$ are independent, i.e.:

$$
P[x \in E, \mathrm{y} \in F]=P[x \in E] \cdot P[y \in F]
$$

From the above expression, it can be deduced that $X$ and $Y$ are statistically independent if any of the following relationships hold:

- $F_{X Y}(x, y)=F_{X}(x) \cdot F_{Y}(y)$
- $f_{X Y}(x, y)=f_{X}(x) \cdot f_{Y}(y) \quad$ if both are jointly continuous
- $P[x=x, Y=y]=P[X=x] \cdot P[Y=y] \quad$ if both are jointly discrete

We summarize below some of the properties pertaining to the relationships between two random variables. In the following, $X$ and $Y$ are two independent random variables defined on the same sample space, c is a constant and $g$ and $h$ are two arbitrary real functions.
(i) Convolution Property

If $Z=X+Y$, then

- if $X$ and $Y$ are jointly discrete

$$
\begin{equation*}
P[Z=k]=\sum_{i+j=k} P[X=i] P[Y=j]=\sum_{i=0}^{k} P[X=i] P[Y=k-i] \tag{1.38}
\end{equation*}
$$

- if X and Y are jointly continuous

$$
\begin{align*}
f_{Z}(z) & =\int_{0}^{\infty} f_{X}(x) f_{Y}(z-x) d x=\int_{0}^{\infty} f_{X}(z-y) f_{Y}(y) d y \\
& =f_{X}(x) \otimes f_{Y}(y) \tag{1.39}
\end{align*}
$$

where $\otimes$ denotes the convolution operator.
(ii) $E[c X]=c E[X]$
(iii) $E[X+Y]=E[X]+E[Y]$
(iv) $E[g(X) h(Y)]=E[g(X)] \cdot E[h(Y)] \quad$ if $X$ and $Y$ are independent
(v) $\operatorname{Var}[c X]=c^{2} \operatorname{Var}[X]$
(vi) $\operatorname{Var}[X+Y]=\operatorname{Var}[X]+\operatorname{Var}[Y] \quad$ if $X$ and $Y$ are independent
(vi) $\operatorname{Var}[X]=E\left[X^{2}\right]-(E[X])^{2}$

## Example 1.7: Random sum of random variables

Consider the voice packetization process during a teleconferencing session, where voice signals are packetized at a teleconferencing station before being transmitted to the other party over a communication network in packet form. If the number $(N)$ of voice signals generated during a session is a random variable with mean $E(N)$, and a voice signal can be digitized into $X$ packets, find the mean and variance of the number of packets generated during a teleconferencing session, assuming that these voice signals are identically distributed.

## Solution

Denote the number of packets for each voice signal as $X_{i}$ and the total number of packets generated during a session as $T$, then we have

$$
T=X_{1}+X_{2}+\ldots+X_{N}
$$

To calculate the expected value, we first condition it on the fact that $N=k$ and then use the total probability theorem to sum up the probability. That is:

$$
\begin{aligned}
E[T] & =\sum_{i=1}^{N} E[T \mid N=k] P[N=k] \\
& =\sum_{i=1}^{N} k E[X] P[N=k] \\
& =E[X] E[N]
\end{aligned}
$$

To compute the variance of $T$, we first compute $E\left[T^{2}\right]$ :

$$
\begin{aligned}
E\left[T^{2} \mid N=k\right] & =\operatorname{Var}[T \mid N=k]+(E[T \mid N=k])^{2} \\
& =k \operatorname{Var}[X]+k^{2}(E[X])^{2}
\end{aligned}
$$

and hence we can obtain

$$
\begin{aligned}
E\left[T^{2}\right] & =\sum_{k=1}^{N}\left(k \operatorname{Var}[X]+k^{2}(E[X])^{2}\right) P[N=k] \\
& =\operatorname{Var}[X] E[N]+E\left[N^{2}\right](E[X])^{2}
\end{aligned}
$$

Finally:

$$
\begin{aligned}
\operatorname{Var}[T] & =E\left[T^{2}\right]-(E[T])^{2} \\
& =\operatorname{Var}[X] E[N]+E\left[N^{2}\right](E[X])^{2}-(E[X])^{2}(E[N])^{2} \\
& =\operatorname{Var}[X] E[N]+(E[X])^{2} \operatorname{Var}[N]
\end{aligned}
$$

## Example 1.8

Consider two packet arrival streams to a switching node, one from a voice source and the other from a data source. Let $X$ be the number of time slots until a voice packet arrives and $Y$ the number of time slots till a data packet arrives. If $X$ and $Y$ are geometrically distributed with parameters $p$ and $q$ respectively, find the distribution of the time (in terms of time slots) until a packet arrives at the node.

## Solution

Let $Z$ be the time until a packet arrives at the node, then $Z=\min (X, Y)$ and we have

$$
\begin{gathered}
P[Z>k]=P[X>k] P[Y>k] \\
1-F_{Z}(k)=\left\{1-F_{X}(k)\right\}\left\{1-F_{Y}(k)\right\}
\end{gathered}
$$

but

$$
\begin{aligned}
F_{X}(k) & =\sum_{j=1}^{\infty} p(1-p)^{j-1}=p \frac{1-(1-p)^{k}}{1-(1-p)} \\
& =1-(1-p)^{k}
\end{aligned}
$$

Similarly

$$
F_{Y}(k)=1-(1-q)^{k}
$$

Therefore, we obtain

$$
\begin{aligned}
F_{Z}(k) & =1-(1-p)^{k}(1-q)^{k} \\
& =1-[(1-p)(1-q)]^{k}
\end{aligned}
$$

## Theorem 1.1

Suppose a random variable $Y$ is a function of a finite number of independent random variables $\left\{X_{i}\right\}$, with arbitrary known probability density functions ( $p d f$ ). If

$$
\mathrm{Y}=\sum_{i=1}^{n} \mathrm{X}_{\mathrm{i}}
$$

then the $p d f$ of $Y$ is given by the density function:

$$
\begin{equation*}
g_{Y}(y)=f_{X 1}\left(x_{1}\right) \otimes f_{X 2}\left(x_{2}\right) \otimes f_{X 3}\left(x_{3}\right) \otimes \ldots f_{X n}\left(x_{n}\right) \tag{1.40}
\end{equation*}
$$

The keen observer might note that this result is a general extension of expression (1.39). Fortunately the convolution of density functions can be easily handled by transforms ( $z$ or Laplace).

## Example 1.9

Suppose the propagation delay along a link follows the exponential distribution:

$$
f_{X}\left(x_{i}\right)=\exp \left(-x_{i}\right) \quad \text { for } x_{i} \geq 0 \text { for } i=1 \ldots 10
$$

Find the density function $g(y)$ where $y=x_{1}+x_{2}+\ldots x_{10}$.

Solution
Consider the effect of the new random variable by using Theorem 1.1 above, where each exponential random variable are independent and identically distributed with $g(y)=\frac{y^{i-1} e^{-y}}{(i-1)!}$ for $y \geq 0$ as shown in Figure 1.5.


Figure 1.5 The density function $g(y)$ for $I=1 \ldots 10$

### 1.1.6 Independence of Random Variables

Independence is probably the most fertile concept in probability theorems, for example, it is applied to queueing theory under the guise of the well-known Kleinrock independence assumption.

## Theorem 1.2

## [Strong law of large numbers]

For $n$ independent and identically distributed random variables $\left\{X_{n}, n \geq 1\right\}$ :

$$
\begin{equation*}
Y_{n}=\left\{X_{1}+X_{2} \ldots X_{n}\right\} / n \rightarrow E\left[X_{1}\right] \quad \text { as } \quad n \rightarrow \infty \tag{1.41}
\end{equation*}
$$

That is, for large $n$, the arithmetic mean of $Y_{n}$ of $n$ independent and identically distributed random variables with the same distribution is close to the expected value of these random variables.

Theorem 1.3
[Central Limit theorem]
Given $Y_{n}$ as defined above:

$$
\begin{equation*}
\left\{Y_{n}-E\left[X_{1}\right]\right\} \sqrt{n} \approx N\left(0, \sigma^{2}\right) \text { for } n \gg 1 \tag{1.42}
\end{equation*}
$$

where $N\left(0, \sigma^{2}\right)$ denotes the random variable with mean zero and variance $\sigma^{2}$ of each $X_{n}$.

The theorem says that the difference between the arithmetic mean of $Y_{n}$ and the expected value $E\left[X_{1}\right]$ is a Gaussian distributed random variable divided by $\sqrt{n}$ for large $n$.

## 1.2 z-TRANSFORMS - GENERATING FUNCTIONS

If we have a sequence of numbers $\left\{f_{0}, f_{1}, f_{2}, \ldots f_{k} \ldots\right\}$, possibly infinitely long, it is often desirable to compress it into a single function - a closed-form expression that would facilitate arithmetic manipulations and mathematical proofing operations. This process of converting a sequence of numbers into a single function is called the $z$-transformation, and the resultant function is called the $z$-transform of the original sequence of numbers. The $z$-transform is commonly known as the generating function in probability theory.

The $z$-transform of a sequence is defined as

$$
\begin{equation*}
F(z) \equiv \sum_{k=0}^{\infty} f_{k} z^{k} \tag{1.43}
\end{equation*}
$$

where $z^{k}$ can be considered as a 'tag' on each term in the sequence and hence its position in that sequence is uniquely identified should the sequence need to be recovered. The $z$-transform $F(\mathrm{z})$ of a sequence exists so long as the sequence grows slower than $a^{k}$, i.e.:

$$
\lim _{k \rightarrow \infty} \frac{\left|k_{k}\right|}{a^{k}}=0
$$

for some $a>0$ and it is unique for that sequence of numbers.
$z$-transform is very useful in solving difference equations (or so-called recursive equations) that contain constant coefficients. A difference equation is an equation in which a term (say $k$ th) of a function $f(\bullet)$ is expressed in terms of other terms of that function. For example:

$$
f_{k-1}+f_{k+1}=2 f_{k}
$$

This kind of difference equation occurs frequently in the treatment of queueing systems. In this book, we use $\Leftrightarrow$ to indicate a transform pair, for example, $f_{k} \Leftrightarrow F(z)$.

### 1.2.1 Properties of $z$-Transforms

$z$-transform possesses some interesting properties which greatly facilitate the evaluation of parameters of a random variable. If $X$ and $Y$ are two independent random variables with respective probability mass functions $f_{k}$ and $f_{g}$, and their corresponding transforms $F(z)$ and $G(z)$ exist, then we have the two following properties:

## (a) Linearity property

$$
\begin{equation*}
a f_{k}+b g_{k} \Leftrightarrow a F(z)+b G(z) \tag{1.44}
\end{equation*}
$$

This follows directly from the definition of $z$-transform, which is a linear operation.

## (b) Convolution property

If we define another random variable $H=X+Y$ with a probability mass function $h_{k}$, then the $z$-transform $H(z)$ of $h_{k}$ is given by

$$
\begin{equation*}
H(z)=F(z) \cdot G(z) \tag{1.45}
\end{equation*}
$$

The expression can be proved as follows:

$$
\begin{aligned}
H(z) & =\sum_{k=0}^{\infty} h_{k} z^{k} \\
& =\sum_{k=0}^{\infty} \sum_{i=0}^{k} f_{i} g_{k-i} z^{k} \\
& =\sum_{i=0}^{\infty} \sum_{k=i}^{\infty} f_{i} g_{k-i} z^{k} \\
& =\sum_{i=0}^{\infty} f_{i} z^{i} \sum_{k=i}^{\infty} g_{k-i} z^{k-i} \\
& =F(z) \cdot G(z)
\end{aligned}
$$

The interchange of summary signs can be viewed from the following:

Index $i$

| Index | $f_{0} g_{0}$ |  |
| :--- | :--- | :--- |
| $k$ | $f_{0} g_{1}$ | $f_{1} g_{0}$ |
| $\downarrow$ | $f_{0} g_{2}$ | $f_{1} g_{1}$ |

Table 1.2 Some $z$-transform pairs

| Sequence | $z$-transform |
| :--- | :--- |
| $u_{k}=1 k=0,1,2 \ldots$ | $1 /(1-z)$ |
| $u_{k-a}$ | $z^{a} /(1-z)$ |
| $A a^{k}$ | $A /(1-a z)$ |
| $k a^{k}$ | $a z /(1-a z)^{2}$ |
| $(k+1) a^{k}$ | $1 /(1-a z)^{2}$ |
| $a / k!$ | $a e^{z}$ |

(c) Final values and expectation
(i)

$$
\begin{equation*}
\left.F(z)\right|_{z=1}=1 \tag{1.46}
\end{equation*}
$$

(ii)

$$
\begin{equation*}
E[X]=\left.\frac{d}{d z} F(z)\right|_{z=1} \tag{1.47}
\end{equation*}
$$

(iii)

$$
\begin{equation*}
E\left[X^{2}\right]=\left.\frac{d^{2}}{d z^{2}} F(z)\right|_{z=1}+\left.\frac{d}{d z} F(z)\right|_{z=1} \tag{1.48}
\end{equation*}
$$

Table 1.2 summarizes some of the $z$-transform pairs that are useful in our subsequent treatments of queueing theory.

Example 1.10
Let us find the $z$-transforms for Binomial, Geometric and Poisson distributions and then calculate the expected values, second moments and variances for these distributions.
(i) Binomial distribution:

$$
\begin{aligned}
B_{X}(z) & =\sum_{k=0}^{n}\binom{n}{k} p^{k}(1-p)^{n-k} z^{k} \\
& =(1-p+p z)^{n} \\
\frac{d}{d z} B_{X}(z) & =n p(1-p+p z)^{n-1}
\end{aligned}
$$

therefore

$$
E[X]=\left.\frac{d}{d z} B_{X}(z)\right|_{z=1}=n p
$$

and

$$
\begin{aligned}
\frac{d^{2}}{d z^{2}} B_{X}(z) & =n p(n-1) p(1-p+p z)^{n-2} \\
E\left[X^{2}\right] & =n(n-1) p^{2}+n p \\
\sigma^{2} & =E\left[X^{2}\right]-E^{2}[X] \\
& =n p(1-p)
\end{aligned}
$$

(ii) Geometric distribution:

$$
\begin{gathered}
G(z)=\sum_{k=1}^{\infty} p(1-p)^{k-1} z^{k}=\frac{p z}{1-(1-p) z} \\
E[X]=\frac{p}{1-(1-p) z}+\left.\frac{p z(1-p)}{(1-(1-p) z)^{2}}\right|_{z=1}=\frac{1}{p} \\
\left.\frac{d^{2}}{d z^{2}} G(z)\right|_{z=1}=2\left(\frac{1}{p^{2}}-\frac{1}{p}\right) \\
\sigma^{2}=\frac{1}{p^{2}}-\frac{1}{p}
\end{gathered}
$$

(iii) Poisson distribution:

$$
\begin{gathered}
G(z)=\sum_{k=0}^{\infty} \frac{(\lambda t)^{k}}{k!} e^{-\lambda t} z^{k}=e^{-\lambda t} e^{+\lambda t z}=e^{-\lambda t(1-z)} \\
E[X]=\left.\frac{d}{d z} G(z)\right|_{z=1}=\lambda t e^{-\lambda t(1-z)}=\lambda t \\
\left.\frac{d^{2}}{d z^{2}} G(z)\right|_{z=1}=(\lambda t)^{2} \\
\sigma^{2}=E[X]^{2}-E^{2}[X]=\lambda t
\end{gathered}
$$

Table 1.3 summarizes the $z$-transform expressions for those probability mass functions discussed in Section 1.2.3.

Table $1.3 z$-transforms for some of the discrete random variables

| Random variable | $z$-transform |
| :--- | :--- |
| Bernoulli | $G(z)=q+p z$ |
| Binomial | $G(z)=(q+p z)^{n}$ |
| Geometric | $G(z)=p z /(1-q z)$ |
| Poisson | $G(z)=e^{-\lambda(1-z)}$ |



Figure 1.6 A famous legendary puzzle

## Example 1.11

This is a famous legendary puzzle. According to the legend, a routine morning exercise for Shaolin monks is to move a pile of iron rings from one corner (Point A) of the courtyard to another (Point C) using only a intermediate point (Point B) as a resting point (Figure 1.6). During the move, a larger ring cannot be placed on top of a smaller one at the resting point. Determine the number of moves required if there are $k$ rings in the pile.

## Solution

To calculate the number of moves $\left(m_{k}\right)$ required, we first move the top $(k-1)$ rings to Point B and then move the last ring to Point C , and finally move the $(k-1)$ rings from Point B to Point C to complete the exercise. Denote its $z$-transform as $M(z)=\sum_{k=0}^{\infty} m_{k} z^{k}$ and $m_{0}=0$, then from the above-mentioned recursive approach we have

$$
\begin{aligned}
& m_{k}=m_{k-1}+1+m_{k-1} \quad k \geq 1 \\
& m_{k}=2 m_{k-1}+1
\end{aligned}
$$

Multiplying the equation by $z^{k}$ and summing it from zero to infinity, we have

$$
\begin{aligned}
& \sum_{k=1}^{\infty} m_{k} z^{k}=2 \sum_{k=1}^{\infty} m_{k-1} z^{k}+\sum_{k=1}^{\infty} z^{k} \\
& M(z)-m_{0}=2 z M(z)+\frac{z}{1-z}
\end{aligned}
$$

and

$$
M(z)=\frac{z}{(1-z)(1-2 z)}
$$

To find the inverse of this expression, we do a partial fraction expansion:

$$
M(z)=\frac{1}{1-2 z}+\frac{-1}{1-z}=(2-1) z+\left(2^{2}-1\right) z^{2}+\left(2^{3}-1\right) z^{3}+\ldots
$$

Therefore, we have $m_{k}=2^{k}-1$

Example 1.12
Another well-known puzzle is the Fibonacci numbers $\{1,1,2,3,5,8,13,21$, $\ldots\}$, which occur frequently in studies of population grow. This sequence of numbers is defined by the following recursive equation, with the initial two numbers as $f_{0}=f_{1}=1$ :

$$
f_{k}=f_{k-1}+f_{k-2} \quad k \geq 2
$$

Find an explicit expression for $f_{k}$.

## Solution

First multiply the above equations by $z^{k}$ and sum it to infinity, so we have

$$
\begin{gathered}
\sum_{k=2}^{\infty} f_{k} z^{k}=\sum_{k=2}^{\infty} f_{k-1} z^{k}+\sum_{k=2}^{\infty} f_{k-2} z^{k} \\
F(z)-f_{1} z-f_{0}=z\left(F(z)-f_{0}\right)+z^{2} F(z) \\
F(z)=\frac{-1}{z^{2}+z-1}
\end{gathered}
$$

Again, by doing a partial fraction expression, we have

$$
\begin{aligned}
F(z) & =\frac{1}{\sqrt{5} z_{1}\left[1-\left(z / z_{1}\right)\right]}-\frac{1}{\sqrt{5} z_{2}\left[1-\left(z / z_{2}\right)\right]} \\
& =\frac{1}{\sqrt{5} z_{1}}\left(1+\frac{z}{z_{1}}+\ldots\right)-\frac{1}{\sqrt{5} z_{2}}\left(1+\frac{z}{z_{2}}+\ldots\right)
\end{aligned}
$$

where

$$
z_{1}=\frac{-1+\sqrt{5}}{2} \quad \text { and } \quad z_{2}=\frac{-1-\sqrt{5}}{2}
$$

Therefore, picking up the $k$ term, we have

$$
f_{k}=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{k+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{k+1}\right]
$$

### 1.3 LAPLACE TRANSFORMS

Similar to $z$-transform, a continuous function $f(t)$ can be transformed into a new complex function to facilitate arithmetic manipulations. This transformation operation is called the Laplace transformation, named after the great French mathematician Pierre Simon Marquis De Laplace, and is defined as

$$
\begin{equation*}
F(s)=L[f(t)]=\int_{-\infty}^{\infty} f(t) e^{-s t} d t \tag{1.49}
\end{equation*}
$$

where $s$ is a complex variable with real part $\sigma$ and imaginary part $j \omega$; i.e. $s=\sigma+j \omega$ and $j=\sqrt{-1}$. In the context of probability theory, all the density functions are defined only for the real-time axis, hence the 'two-sided' Laplace transform can be written as

$$
\begin{equation*}
F(s)=L[f(t)]=\int_{0^{-}}^{\infty} f(t) e^{-s t} d t \tag{1.50}
\end{equation*}
$$

with the lower limit of the integration written as $0^{-}$to include any discontinuity at $t=0$. This Laplace transform will exist so long as $f(t)$ grows no faster than an exponential, i.e.:

$$
|f(t)| \leq M e^{\alpha t}
$$

for all $t \geq 0$ and some positive constants M and $\alpha$. The original function $f(t)$ is called the inverse transform or inverse of $F(s)$, and is written as

$$
f(t)=L^{-1}[F(s)]
$$

The Laplace transformation is particularly useful in solving differential equations and corresponding initial value problems. In the context of queueing theory, it provides us with an easy way of finding performance measures of a queueing system in terms of Laplace transforms. However, students should note that it is at times extremely difficult, if not impossible, to invert these Laplace transform expressions.

### 1.3.1 Properties of the Laplace Transform

The Laplace transform enjoys many of the same properties as the ztransform as applied to probability theory. If $X$ and $Y$ are two independent continuous random variables with density functions $f_{X}(x)$ and $f_{Y}(y)$, respectively and their corresponding Laplace transforms exist, then their properties are:
(i) Uniqueness property

$$
\begin{equation*}
f_{X}(\tau)=f_{Y}(\tau) \quad \text { implies } \quad F_{X}(s)=F_{Y}(s) \tag{1.51}
\end{equation*}
$$

(ii) Linearity property

$$
\begin{equation*}
a f_{X}(x)+b f_{Y}(y) \Rightarrow a F_{X}(s)+b F_{Y}(s) \tag{1.52}
\end{equation*}
$$

(iii) Convolution property

$$
\text { If } Z=X+Y \text {, then }
$$

$$
\begin{align*}
F_{Z}(s) & =L\left[f_{z}(z)\right]=L\left[f_{X+Y}(x+y)\right] \\
& =F_{X}(s) \cdot F_{Y}(s) \tag{1.53}
\end{align*}
$$

(iv) Expectation property

$$
\begin{gather*}
E[X]=-\left.\frac{d}{d s} F_{X}(s)\right|_{s=0} \quad \text { and } \quad E\left[X^{2}\right]=\left.\frac{d^{2}}{d s^{2}} F_{X}(s)\right|_{s=0}  \tag{1.54}\\
E\left[X^{n}\right]=\left.(-1)^{n} \frac{d^{n}}{d s^{n}} F_{X}(s)\right|_{s=0} \tag{1.55}
\end{gather*}
$$

(v) Differentiation property

$$
\begin{gather*}
L\left[f_{X}^{\prime}(x)\right]=s F_{X}(s)-f_{X}(0)  \tag{1.56}\\
L\left[f_{X}^{\prime \prime}(x)\right]=s^{2} F_{X}(s)-s f_{X}(0)-f_{X}^{\prime}(0) \tag{1.57}
\end{gather*}
$$

Table 1.4 shows some of the Laplace transform pairs which are useful in our subsequent discussions on queueing theory.

Table 1.4 Some Laplace transform pairs

| Function | Laplace transform |
| :--- | :--- |
| $\delta(t)$ unit impulse | 1 |
| $\delta(t-a)$ | $e^{-a s}$ |
| 1 unit step | $1 / s$ |
| $t$ | $1 / s^{2}$ |
| $t^{n-1} /(n-1)!$ | $1 / s^{n}$ |
| Ae $e^{a t}$ | $A /(s-a)$ |
| $t e^{a t}$ | $1 /(s-a)^{2}$ |
| $t^{n-1} e^{a t /}(n-1)!$ | $1 /(s-a)^{n} n=1,2, \ldots$ |

## Example 1.13

Derive the Laplace transforms for the exponential and $k$-stage Erlang probability density functions, and then calculate their means and variances.
(i) exponential distribution

$$
\begin{gathered}
F(s)=\int_{0}^{\infty} e^{-s s} \lambda e^{-\lambda x} d x=\left[-\frac{\lambda}{s+\lambda} e^{-(s+\lambda) x}\right]_{0}^{\infty}=\frac{\lambda}{s+\lambda} \\
F(s)=\int_{0}^{\infty} e^{-s x} \lambda e^{-\lambda x} d x=\left[-\frac{\lambda}{s+\lambda} e^{-(s+\lambda) x}\right]_{0}^{\infty}=\frac{\lambda}{s+\lambda} \\
E[X]=-\left.\frac{d}{d s} F(s)\right|_{s=0}=\frac{1}{\lambda} \\
E\left[X^{2}\right]=\left.\frac{d^{2}}{d s^{2}} F(S)\right|_{s=0}=\frac{2}{\lambda^{2}} \\
\sigma^{2}=E\left[X^{2}\right]-E^{2}[X]=\frac{1}{\lambda^{2}} \\
C=\frac{\sigma}{E[X]}=1
\end{gathered}
$$

(ii) $k$-stage Erlang distribution

$$
\begin{aligned}
F(s) & =\int_{0}^{\infty} e^{-s s} \frac{\lambda^{k} x^{k-1}}{(k-1)!} e^{-\lambda x} d x=\frac{\lambda^{k}}{(k-1)!} \int_{0}^{\infty} x^{k-1} e^{-(s+\lambda) x} d x \\
& =\frac{\lambda^{k}}{(s+\lambda)^{k}(k-1)!} \int_{0}^{\infty}\{(s+\lambda) x\}^{k-1} e^{-(s+\lambda) x} d(s+\lambda) x
\end{aligned}
$$

Table 1.5 Laplace transforms for some probability functions

| Random variable | Laplace transform |
| :--- | :--- |
| Uniform $a<x<b$ | $F(s)=e^{-s(a+b)} / s(b-a)$ |
| Exponent | $F(s)=\lambda / s+\lambda$ |
| Gamma | $F(s)=\lambda^{\alpha} /(s+\lambda)^{\alpha}$ |
| Erlang-k | $F(s)=\lambda^{k} /(s+\lambda)^{k}$ |

The last integration term is recognized as the gamma function and is equal to $(k-1)$ ! Hence we have

$$
F(s)=\left(\frac{\lambda}{s+\lambda}\right)^{k}
$$

Table 1.5 gives the Laplace transforms for those continuous random variables discussed in Section 1.1.3.

## Example 1.14

Consider a counting process whose behavior is governed by the following two differential-difference equations:

$$
\begin{aligned}
& \frac{d}{d t} P_{k}(t)=-\lambda P_{k}(t)+\lambda P_{k-1}(t) \quad k>0 \\
& \frac{d}{d t} P_{0}(t)=-\lambda P_{0}(t)
\end{aligned}
$$

Where $P_{k}(t)$ is the probability of having $k$ arrivals within a time interval $(0, t)$ and $\lambda$ is a constant, show that $P_{k}(t)$ is Poisson distributed.

Let us define the Laplace transform of $P_{k}(t)$ and $P_{0}(t)$ as

$$
\begin{aligned}
& F_{k}(s)=\int_{0}^{\infty} e^{-s t} P_{k}(t) d t \\
& F_{0}(s)=\int_{0}^{\infty} e^{-s t} P_{0}(t) d t
\end{aligned}
$$

From the properties of Laplace Transform, we know

$$
\begin{aligned}
& L\left[P_{k}^{\prime}(t)\right]=s F_{k}(s)-P_{k}(0) \\
& L\left[P_{0}^{\prime}(t)\right]=s F_{0}(s)-P_{0}(0)
\end{aligned}
$$

Substituting them into the differential-difference equations, we obtain

$$
\begin{aligned}
& F_{0}(s)=\frac{P_{0}(0)}{s+\lambda} \\
& F_{k}(s)=\frac{P_{k}(0)+\lambda L_{k-1}(s)}{s+\lambda}
\end{aligned}
$$

If we assume that the arrival process begins at time $t=0$, then $P_{0}(0)=1$ and $P_{k}(0)=0$, and we have

$$
\begin{aligned}
F_{0}(s) & =\frac{1}{s+\lambda} \\
F_{k}(s) & =\frac{\lambda}{s+\lambda} F_{k-1}(s)=\left(\frac{\lambda}{s+\lambda}\right)^{k} F_{0}(s) \\
& =\frac{\lambda^{k}}{(s+\lambda)^{k+1}}
\end{aligned}
$$

Inverting the two transforms, we obtain the probability mass functions:

$$
\begin{aligned}
& P_{0}(0)=e^{-\lambda t} \\
& P_{k}(t)=\frac{(\lambda t)^{k}}{k!} e^{-\lambda t}
\end{aligned}
$$

### 1.4 MATRIX OPERATIONS

In Chapter 8, with the introduction of Markov-modulated arrival models, we will be moving away from the familiar Laplace ( $z$-transform) solutions to a new approach of solving queueing systems, called matrix-geometric solutions. This particular approach to solving queueing systems was pioneered by Marcel F Neuts. It takes advantage of the similar structure presented in many interesting stochastic models and formulates their solutions in terms of the solution of a nonlinear matrix equation.

### 1.4.1 Matrix Basics

A matrix is a $m \times n$ rectangular array of real (or complex) numbers enclosed in parentheses, as shown below:

$$
\tilde{A}=\left(a_{i j}\right)=\left(\begin{array}{cccc}
a_{11} & a_{12} & \cdots & a_{1 n} \\
a_{21} & a_{22} & \cdots & a_{2 n} \\
\vdots & & & \vdots \\
a_{m 1} & a_{m 2} & \cdots & a_{m n}
\end{array}\right)
$$

where $a_{i j}$ 's are the elements (or components) of the matrix. A $m \times 1$ matrix is a column vector and a $1 \times n$ matrix is a row vector. In the sequel, we denote matrices by capital letters with a tilde $(\sim)$ on top, such as $\tilde{A}, \tilde{B} \& \tilde{C}$, column vectors by small letters with a tilde, such as $\tilde{f} \& \tilde{g}$, and row vectors by small Greek letters, such as $\tilde{\pi} \& \tilde{v}$.

A matrix whose elements are all zero is called the null matrix and denoted by $\tilde{0}$. A diagonal matrix $(\tilde{\Lambda})$ is a square matrix whose entries other than those in the diagonal positions are all zero, as shown below:

$$
\begin{aligned}
\tilde{\Lambda} & =\operatorname{diag}\left(a_{11}, a_{12}, \ldots, a_{n n}\right) \\
& =\left(\begin{array}{cccc}
a_{11} & 0 & \ldots & 0 \\
0 & a_{12} & \ldots & 0 \\
\ldots & & \ddots & \\
0 & 0 & \ldots & a_{n n}
\end{array}\right)
\end{aligned}
$$

If the diagonal entries are all equal to one then we have the identity matrix ( $\tilde{I}$ ).

The transpose $\tilde{A}^{T}$ of a $m \times n$ matrix $\tilde{A}=\left(a_{i j}\right)$ is the $n \times m$ matrix obtained by interchanging the rows and columns of $\tilde{A}$, that is

$$
\tilde{A}^{T}=\left(a_{j i}\right)=\left(\begin{array}{cccc}
a_{11} & a_{21} & \cdots & a_{m 1} \\
a_{12} & a_{22} & \cdots & a_{m 2} \\
\cdots & & & \vdots \\
a_{1 n} & a_{2 n} & \cdots & a_{m n}
\end{array}\right)
$$

The inverse of an $n$-rowed square matrix $\tilde{A}$ is denoted by $\tilde{A}^{-1}$ and is an $n-$ rowed square matrix that satisfies the following expression:

$$
\tilde{A} \tilde{A}^{-1}=\tilde{A}^{-1} \tilde{A}=\tilde{I}
$$

$\tilde{A}^{-1}$ exists (and is then unique) if and only if $A$ is non-singular, i.e. if and only if the determinant of $A$ is not zero, $\tilde{A} \neq 0$. In general, the inverse of $\tilde{A}$ is given by

$$
\tilde{A}^{-1}=\frac{1}{\operatorname{det} \tilde{A}}\left(\begin{array}{cccc}
A_{11} & A_{21} & \ldots & A_{n 1} \\
A_{12} & A_{22} & & \\
\ldots & & & \ldots \\
A_{1 n} & A_{2 n} & & A_{n n}
\end{array}\right)
$$

where $A_{i j}$ is the cofactor of $a_{i j}$ in $\tilde{A}$. The cofactor of $a_{i j}$ is the product of $(-1)^{i+j}$ and the determinant formed by deleting the $i$ th row and the $j$ th column from the $\operatorname{det} \tilde{A}$. For a $2 \times 2$ matrix $\tilde{A}$, the inverse is given by

$$
\tilde{A}=\left(\begin{array}{ll}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{array}\right) \quad \text { and } \quad \tilde{A}^{-1}=\frac{1}{a_{11} a_{22}-a_{21} a_{12}}\left(\begin{array}{cc}
a_{22} & -a_{12} \\
-a_{21} & a_{11}
\end{array}\right)
$$

We summarize some of the properties of matrixes that are useful in manipulating them. In the following expressions, $\alpha$ and $\beta$ are numbers:
(i) $\quad \alpha(\tilde{A}+\tilde{B})=\alpha \tilde{A}+\alpha \tilde{B} \quad$ and $\quad(\alpha+\beta) \tilde{A}=\alpha \tilde{A}+\beta \tilde{A}$
(ii) $\quad(\alpha \tilde{A}) \tilde{B}=\alpha(\tilde{A} \tilde{B})=\tilde{A}(\alpha \tilde{B}) \quad$ and $\quad \tilde{A}(\tilde{B} \tilde{C})=(\tilde{A} \tilde{B}) \tilde{C}$
(iii) $(\tilde{A}+\tilde{B}) \tilde{C}=\tilde{A} \tilde{C}+\tilde{B} \tilde{C} \quad$ and $\tilde{C}(\tilde{A}+\tilde{B})=\tilde{C} \tilde{A}+\tilde{C} \tilde{B}$
(iv) $\tilde{A} \tilde{B} \neq \tilde{B} \tilde{A}$ in general
(v) $\tilde{A} \tilde{B}=\tilde{0}$ does not necessarily imply $\tilde{A}=\tilde{0}$ or $\tilde{B}=\tilde{0}$
(vi) $(\tilde{A}+\tilde{B})^{T}=\tilde{A}^{T}+\tilde{B}^{T} \quad$ and $\quad\left(\tilde{A}^{T}\right)^{T}=\tilde{A}$
(vii) $(\tilde{A} \tilde{B})^{T}=\tilde{B}^{T} \tilde{A}^{T} \quad$ and $\quad \operatorname{det} \tilde{A}=\operatorname{det} \tilde{A}^{T}$
(viii) $\left(\tilde{A}^{-1}\right)^{-1}=\tilde{A} \quad$ and $\quad(\tilde{A} \tilde{B})^{-1}=\tilde{B}^{-1} \tilde{A}^{-1}$
(ix) $\quad\left(\tilde{A}^{-1}\right)^{T}=\left(\tilde{A}^{T}\right)^{-1} \quad$ and $\quad\left(\tilde{A}^{2}\right)^{-1}=\left(\tilde{A}^{-1}\right)^{2}$

### 1.4.2 Eigenvalues, Eigenvectors and Spectral Representation

An eigenvalue (or characteristic value) of an $n \times n$ square matrix $\tilde{A}=\left(a_{i j}\right)$ is a real or complex scalar $\lambda$ satisfying the following vector equation for some non-zero (column) vector $\tilde{x}$ of dimension $n \times 1$. The vector $\tilde{x}$ is known as the eigenvector, or more specifically the column (or right) eigenvector:

$$
\begin{equation*}
\tilde{A} \tilde{x}=\lambda \tilde{x} \tag{1.58}
\end{equation*}
$$

This equation can be rewritten as $(\tilde{A}-\lambda \tilde{I}) \tilde{x}=0$ and has a non-zero solution $\tilde{x}$ only if $(\tilde{A}-\lambda \tilde{I})$ is singular; that is to say that any eigenvalue must satisfy $\operatorname{det}(\tilde{A}-\lambda \tilde{I})=0$. This equation, $\operatorname{det}(\tilde{A}-\lambda \tilde{I})=0$, is a polynomial of degree n in $\lambda$ and has exactly n real or complex roots, including multiplicity. Therefore,

A has n eigenvalues $\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}$ with the corresponding eigenvectors $\tilde{x}_{1}$, $\tilde{x}_{2}, \ldots, \tilde{x}_{n}$. The polynomial is known as the characteristic polynomial of A and the set of eigenvalues is called the spectrum of $\tilde{A}$.

Similarly, the row (or left) eigenvectors are the solutions of the following vector equation:

$$
\begin{equation*}
\tilde{\pi} \tilde{A}=\lambda \tilde{\pi} \tag{1.59}
\end{equation*}
$$

and everything that is said about column eigenvectors is also true for row eigenvectors.

Here, we summarize some of the properties of eigenvalues and eigenvectors:
(i) The sum of the eigenvalues of $\tilde{A}$ is equal to the sum of the diagonal entries of $\tilde{A}$. The sum of the diagonal entries of $\tilde{A}$ is called the trace of $\tilde{A}$.

$$
\begin{equation*}
\operatorname{tr}(\tilde{A})=\sum_{i} \lambda_{i} \tag{1.60}
\end{equation*}
$$

(ii) If A has eigenvalues $\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}$, then $\lambda_{1}^{k}, \lambda_{2}^{k}, \ldots, \lambda_{n}^{k}$ are eigenvectors of $\tilde{A}^{k}$, and we have

$$
\begin{equation*}
\operatorname{tr}\left(\tilde{A}^{k}\right)=\sum_{i} \lambda_{i}^{k} \quad k=1,2, \ldots \tag{1.61}
\end{equation*}
$$

(iii) If $\tilde{A}$ is a non-singular matrix with eigenvalues $\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}$, then $\left.\lambda_{1}^{-1}>, \lambda_{2}^{-1}\right), \ldots, \lambda_{n}^{-1}$ are eigenvectors of $\tilde{A}^{-1}$. Moreover, any eigenvector of $\tilde{A}$ is an eigenvector of $\tilde{A}^{-1}$.
(iv) $\tilde{A}$ and $\tilde{A}^{T}$ do not necessarily have the same eigenvectors. However, if $\tilde{A}^{T} \tilde{x}=\lambda \tilde{x}$ then $\tilde{x}^{T} \tilde{A}=\lambda \tilde{x}^{T}$, and the row vector $\tilde{x}^{T}$ is called a left eigenvector of $\tilde{A}$.

It should be pointed out that eigenvalues are in general relatively difficult to compute, except for certain special cases.

If the eigenvalues $\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}$ of a matrix $\tilde{A}$ are all distinct, then the corresponding eigenvectors $\tilde{x}_{1}, \tilde{x}_{2}, \ldots, \tilde{x}_{n}$ are linearly independent, and we can express $\tilde{A}$ as

$$
\begin{equation*}
\tilde{A}=\tilde{N} \tilde{\Lambda} \tilde{N}^{-1} \tag{1.62}
\end{equation*}
$$

where $\tilde{\Lambda}=\operatorname{diag}\left(\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}\right), \tilde{N}=\left[\tilde{x}_{1}, \tilde{x}_{2}, \ldots, \tilde{x}_{n}\right]$ whose $i$ th column is $\tilde{x}_{i}, \tilde{N}^{-1}$ is the inverse of $\tilde{N}$, and is given by

$$
\tilde{N}^{-1}=\left[\begin{array}{l}
\tilde{\pi}_{1} \\
\tilde{\pi}_{2} \\
\ldots \\
\tilde{\pi}_{n}
\end{array}\right]
$$

By induction, it can be shown that $\tilde{A}^{k}=\tilde{N} \tilde{\Lambda}^{k} \tilde{N}^{-1}$.
If we define $\tilde{B}_{k}$ to be the matrix obtained by multiplying the column vector $\tilde{x}_{k}$ with the row vector $\tilde{\pi}_{k}$, then we have

$$
\begin{align*}
\tilde{B}_{k} & =\tilde{x}_{k} \tilde{\pi}_{k} \\
& =\left(\begin{array}{ccc}
x_{k}(1) \pi_{k}(1) & \ldots & x_{k}(1) \pi_{k}(n) \\
\ldots & & \ldots \\
x_{k}(n) \pi_{k}(1) & \ldots & x_{k}(n) \pi_{k}(n)
\end{array}\right) \tag{1.63}
\end{align*}
$$

It can be shown that

$$
\begin{align*}
\tilde{A} & =\tilde{N} \tilde{\Lambda} \tilde{N}^{-1} \\
& =\lambda_{1} \tilde{B}_{1}+\lambda_{2} \tilde{B}_{2}+\ldots+\lambda_{n} \tilde{B}_{n} \tag{1.64}
\end{align*}
$$

and

$$
\begin{equation*}
\tilde{A}^{k}=\lambda_{11}^{k} \tilde{B}_{1}+\lambda_{2}^{k} \tilde{B}_{2}+\ldots+\lambda_{n}^{k} \tilde{B}_{n} \tag{1.65}
\end{equation*}
$$

The expression of $\tilde{A}$ in terms of its eigenvalues and the matrices $\tilde{B}_{k}$ is called the spectral representation of $\tilde{A}$.

### 1.4.3 Matrix Calculus

Let us consider the following set of ordinary differential equations with constant coefficients and given initial conditions:

$$
\begin{align*}
& \frac{d}{d t} x_{1}(t)=a_{11} x_{1}(t)+a_{11} x_{2}(t)+\ldots+a_{1 n} x_{n}(t) \\
& \vdots  \tag{1.66}\\
& \frac{d}{d t} x_{n}(t)=a_{n 1} x_{1}(t)+a_{n 2} x_{2}(t)+\ldots+a_{n n} x_{n}(t)
\end{align*}
$$

In matrix notation, we have

$$
\begin{equation*}
\tilde{x}(t)^{\prime}=\tilde{A} \tilde{x}(t) \tag{1.67}
\end{equation*}
$$

where $\tilde{x}(t)$ is a $n \times 1$ vector whose components $x_{i}(t)$ are functions of an independent variable $t$, and $x(t)^{\prime}$ denotes the vector whose components are the derivatives $d x_{i} / d t$. There are two ways of solving this vector equation:
(i) First let us assume that $\tilde{x}(t)=e^{\lambda t} \tilde{p}$, where $\tilde{P}$ is a scalar vector and substitute it in Equation (1.67), then we have

$$
\lambda e^{\lambda_{t}} \tilde{p}=\tilde{A}\left(e^{\lambda_{t}} \tilde{p}\right)
$$

Since $e^{\lambda_{t}} \neq 0$, it follows that $\lambda$ and $\tilde{p}$ must satisfy $\tilde{A} \tilde{p}=\tilde{\lambda} \tilde{p}$; therefore, if $\lambda_{i}$ is an eigenvector of A and $\tilde{p}_{i}$ is a corresponding eigenvector, then $e^{\lambda_{i}} \tilde{p}_{i}$ is a solution. The general solution is given by

$$
\begin{equation*}
\tilde{x}(t)=\sum_{i=1}^{n} \alpha_{i} e^{\lambda_{i}} \tilde{p}_{i} \tag{1.68}
\end{equation*}
$$

where $\alpha_{i}$ is the constant chosen to satisfy the initial condition of Equation (1.67).
(ii) The second method is to define the matrix exponential $e^{\tilde{A} t}$ through the convergent power series as

$$
\begin{equation*}
\exp \{\tilde{A} t\}=\sum_{k=0}^{\infty} \frac{(\tilde{A} t)^{k}}{k!}=I+\tilde{A} t+\frac{(\tilde{A} t)^{2}}{2!}+\ldots+\frac{(\tilde{A} t)^{k}}{k!} \tag{1.69}
\end{equation*}
$$

By differentiating the expression with respect to $t$ directly, we have

$$
\begin{aligned}
\frac{d}{d t}\left(e^{\tilde{A} t}\right) & =\tilde{A}+\tilde{A}^{2} t+\frac{\tilde{A}^{3} t^{2}}{2!}+\ldots \\
& =\tilde{A}\left(\tilde{I}+\tilde{A} t+\frac{\tilde{A}^{2} t^{2}}{2!}+\ldots\right)=\tilde{A} e^{\tilde{A} t}
\end{aligned}
$$

Therefore, $e^{\tilde{A} t}$ is a solution to Equation (1.67) and is called the fundamental matrix for (1.67).

We summarize some of the useful properties of the matrix exponential below:
(i) $e^{\tilde{A}(s+t)}=e^{\tilde{A} s} e^{\tilde{A} t}$
(ii) $e^{\tilde{A} t}$ is never singular and its inverse is $e^{-\tilde{A} t}$
(iii) $e^{(\tilde{A}+\tilde{B}) t}=e^{\tilde{A} t} e^{\tilde{B} t}$ for all t , only if $\tilde{A} \tilde{B}=\tilde{B} \tilde{A}$
(iv) $\frac{d}{d t} e^{\tilde{A} t}=\tilde{A} e^{\tilde{A} t}=e^{\tilde{A} t} \tilde{A}$
(v) $(\tilde{I}-\tilde{A})^{-1}=\sum_{i=0}^{\infty} \tilde{A}^{i}=\tilde{I}+\tilde{A}+\tilde{A}^{2}+\ldots$
(vi) $e^{\tilde{A} t}=\tilde{N} e^{\tilde{\Lambda} t} \tilde{N}^{-1}$

$$
=e^{\lambda_{1} t} \tilde{B}_{1}+e^{\lambda_{2} t} \tilde{B}_{2}+\ldots+e^{\lambda_{n} t} \tilde{B}_{n}
$$

where

$$
e^{\tilde{\Lambda} t}=\left(\begin{array}{ccc}
e^{\lambda_{1} t} & \cdots & 0 \\
& \ddots & \\
0 & \cdots & e^{\lambda_{n} t}
\end{array}\right)
$$

and $\tilde{B}_{i}$ are as defined in Equation (1.63).
Now let us consider matrix functions. The following are examples:

$$
\tilde{A}(t)=\left(\begin{array}{cc}
t & 0 \\
t^{2} & 4 t
\end{array}\right) \text { and } \tilde{A}(\theta)=\left(\begin{array}{cc}
\sin \theta & \cos \theta \\
0 & -\sin \theta
\end{array}\right)
$$

We can easily extend the calculus of scalar functions to matrix functions. In the following, $\tilde{A}(t)$ and $\tilde{B}(t)$ are matrix functions with independent variable $t$ and $\tilde{U}$ a matrix of real constants:
(i) $\frac{d}{d t} \tilde{U}=0$
(ii) $\frac{d}{d t}(\alpha \tilde{A}(t)+\beta \tilde{B}(t))=\alpha \frac{d}{d t} \tilde{A}(t)+\beta \frac{d}{d t} \tilde{B}(t)$
(iii) $\frac{d}{d t}(\tilde{A}(t) \tilde{B}(t))=\left(\frac{d}{d t} \tilde{A}(t)\right) \tilde{B}(t)+\tilde{A}(t)\left(\frac{d}{d t} \tilde{B}(t)\right)$
(iv) $\frac{d}{d t} \tilde{A}^{2}(t)=\left(\frac{d}{d t} \tilde{A}(t)\right) \tilde{A}(t)+\tilde{A}(t)\left(\frac{d}{d t} \tilde{A}(t)\right)$
(v) $\frac{d}{d t} \tilde{A}^{-1}(t)=-\tilde{A}^{-1}(t)\left(\frac{d}{d t} \tilde{A}(t)\right) \tilde{A}^{-1}(t)$
(vi) $\frac{d}{d t}(\tilde{A}(t))^{T}=\left(\frac{d}{d t} \tilde{A}(t)\right)^{T}$


Figure 1.7 Switches for Problem 6

## Problems

1. A pair of fair dice is rolled 10 times. What will be the probability that 'seven' will show at least once.
2. During Christmas, you are provided with two boxes $A$ and $B$ containing light bulbs from different vendors. Box A contains 1000 red bulbs of which $10 \%$ are defective while Box B contains 2000 blue bulbs of which $5 \%$ are defective.
(a) If I choose two bulbs from a randomly selected box, what is the probability that both bulbs are defective?
(b) If I choose two bulbs from a randomly selected box and find that both bulbs are defective, what is the probability that both came from Box A?
3. A coin is tossed an infinite number of times. Show that the probability that $k$ heads are observed at the $n$th tossing but not earlier equals $\binom{n-1}{k-1} \mathrm{p}^{k}(1-\mathrm{p})^{n-k}$, where $p=P\{\mathrm{H}\}$.
4. A coin with $P\{H\}=p$ and $P\{T\}=q=1-p$ is tossed $n$ times. Show that the probability of getting an even number of heads is $0.5\left[1+(q-p)^{n}\right]$.
5. Let $A, B$ and $C$ be the events that switches $a, b$ and $c$ are closed, respectively. Each switch may fail to close with probability q. Assume that the switches are independent and find the probability that a closed path exists between the terminals in the circuit shown for $q=0.5$.
6. The binary digits that are sent from a detector source generate bits 1 and 0 randomly with probabilities 0.6 and 0.4 , respectively.
(a) What is the probability that two 1 s and three 0 s will occur in a 5-digit sequence.
(b) What is the probability that at least three 1 s will occur in a 5digit sequence.


Figure 1.8 Communication network with 5 links
7. The binary input $\mathbf{X}$ to a channel takes on one of two values, 0 or 1 , with probabilities $3 / 4$ and $1 / 4$ respectively. Due to noise induced errors, the channel output $\mathbf{Y}$ may differ from $\mathbf{X}$. There will be no errors in $\mathbf{Y}$ with probabilities $3 / 4$ and $7 / 8$ when the input $\mathbf{X}$ is 1 or 0 , respectively. Find $P(\mathbf{Y}=1), P(\mathbf{Y}=0)$ and $P(\mathbf{X}=1 \mid$ $Y=1$ ).
8. A wireless sensor node will fail sooner or later due to battery exhaustion. If the failure rate is constant, the time to failure $T$ can be modelled as an exponentially distributed random variable. Suppose the wireless sensor node follow an exponential failure law in hours as $f_{T}(t)=\alpha u(t) e^{-\alpha t}$, where $u(t)$ is the unit step function and $\alpha>0$ is a parameter. Measurements show that for these sensors, the probability that $T$ exceeds $10^{4}$ hours is $\mathrm{e}^{-1}(\approx 0.368)$. Using the value of the parameter $\alpha$ determined, calculate the time $t_{0}$ such that the probability that $T$ is less than $t_{0}$ is 0.05 .
9. A communication network consists of five links that interconnect four routers, as shown below. The probability that each of the link is operational is 0.9 and independent. What is the probability of being able to transmit a message from router $A$ to router $B$ (assume that packets move forward in the direction of the destination)?
10. Two random variables $X$ and $Y$ take on the values $i$ and $2^{i}$ with probability $1 / 2^{i}(i=1,2, \ldots)$. Show that the probabilities sum to one. Find the expected value of $X$ and $Y$.
11. There are three identical cards, one is red on both sides, one is yellow on both sides and the last one is red on one side and yellow on the other side. A card is selected at random and is red on the upper side. What is the probability that the other side is yellow?


[^0]:    Queueing Modelling Fundamentals Second Edition Ng Chee-Hock and Soong Boon-Hee © 2008 John Wiley \& Sons, Ltd

[^1]:    ${ }^{1}$ For simplicity, we assume here that the random variables are non-negative.

