1

Introduction

Stochastic (also referred to as random) models play an important role in our day-to-day activities. Just look around your living or working space and see the items, are needed on a daily basis. Dairy products, appliances, electronic devices, to name a few among several dozens of items, are manufactured/grown/cultivated and shipped from the factories to nearby stores for consumers. How are these delivered? A supply chain plays a major part in all of these, and as is well-known most things involved in this chain are random. How do we study randomness? Through probability. Whether someone likes probability or is allergic to it, it is there to rescue us and provide a good quality of life. While this book is being written, the whole planet has been affected by Covid-19 disrupting almost everyone in one form or the other. The containers used in shipping items via cargo ships are scarce, resulting in prohibitively high costs; trucks needed to haul away the filled containers after dropping the empty ones are in short supply, making all the cargo ships anchor close to the ports where they arrive. According to the news, the bottleneck is expected to last well into 2022. The point of this discussion is to illustrate the havoc played by uncertainty and why it is important to have several scenarios analyzed through stochastic modeling.

In this chapter, we present key concepts and results on probability, random variables, renewal processes including Poisson, and matrix analysis. For full details we refer the reader to the books cited in their respective sections.

The chapter is organized as follows. In section 1.1, probability concepts are reviewed. The basic concepts related to renewal processes are reviewed in section 1.2, and finally the matrix concepts, including Kronecker products and Kronecker sums, which play an important role in matrix-analytic methods (MAM), are reviewed in section 1.3.

1.1. Probability concepts

Even though probability theory was used to describe the experiences connected with games of chance and the calculation of certain probabilities, the main purpose is to discover the general rules and to construct satisfactory theoretical models for problems under study. Most phenomenon in our life are random and probability modeling is vital to understand and take appropriate actions. A brief history of probability is given below for those interested to know about it.

The history of probability is in essence separate from the history of statistics, although statistics relies on probability as the foundation of quantitative inference. Probability theory is believed to have been started by two famous French mathematicians, Blaise Pascal (1623–1662) and Pierre de Fermat (1601–1665), primarily in games of chance (until the 19th century its motivation remained mainly in this field). Later on a number of well-known mathematicians such as Jacob Bernoulli (1654–1705), Nicholas Bernoulli (1687–1759), Abraham de Moivre (1667–1754) and Pierre Simon de Laplace (1749–1827) developed the theory in a much more general setup. During the 19th century the French school (working on the foundations laid by Laplace) and the Russian school were very influential in the development of probability theory we see now. In the French school, to name a few, the main contributors were Simon Poisson (1781–1840), Augustin Cauchy (1789–1857), Jules Bienayme (1796–1878), Joseph Bertrand (1822–1900) and Henri Poincare (1854–1912). From the Russian school, V. Ya. Buniakovsky (1801–1889), P.L. Chebyshev (1821–1894), A.A. Markov (1856–1922), A.M. Liapunov (1857–1918) and A.N. Kolmogorov (1903–1987) were very significant contributors to probability theory. It was Kolmogorov who, in 1933, introduced the now generally accepted axiomatic approach suitable to probability theory and random processes. The books on probability theory by Paul Levy, Herald Cramer, B.V. Gnedenko and A.N. Kolmogorov, Michael Loeve and William Feller provided an impetus to the development of modern probability theory.

There are three major definitions of probability, namely, *axiomatic*, *frequency* and *classical*. Each one has its own merits and demerits. The axiomatic approach is mainly used in developing the mathematical theory of probability. The frequency approach gives an intuitive notion of probability. However, the computation of probability in practice is based on the classical approach.

Suppose that S is a sample space (i.e. S is the set of all possible outcomes of an experiment) and Ω to be the set of all possible subsets of S. For example, consider the experiment of throwing a six-sided die. It can readily be seen that $S = \{1, 2, 3, 4, 5, 6\}$, and $\Omega = \Big\{\{\emptyset\}, \{(1)\}, \cdots, \{6\}, \{1, 2\}, \cdots, \{5, 6\}, \cdots, S\Big\}$, where $\emptyset$ is null or an empty set. As an example, $\emptyset$ will include outcomes that are impossible such as seeing a number 7 or a negative number. Note that the cardinality

of Ω for this example is $2^6 = 64$. In general, if the sample space S has a finite number, say, N, of elements, then the cardinality of Ω is 2^N.

DEFINITION 1.1.– *Probability is a real valued (set) function defined on Ω corresponding to the sample space, S. That is, the probability of an event A, denoted by $P(A)$, is a real number such that the following axioms (a set of rules) are satisfied.*

1) $P(A) \geq 0$,

2) $P(S) = 1$,

3) If $A_i \cap A_j = \emptyset$, *for all* $i \neq j$, *then* $P(\bigcup A_i) = \sum P(A_i)$.

REMARK 1.1.– Note that axiom (3) implies that if A and B are mutually exclusive then $P(AUB) = P(A) + P(B)$.

REMARK 1.2.– The axiomatic approach is used mainly in developing the mathematical theory of probability. The axiomatic approach does not tell us how to compute the probability of an event. One can assign probabilities to the events in S arbitrarily as long as the axioms are not violated. Thus, there are an infinite number of ways of doing this. Because of this a different method based on frequency approach was developed.

The probabilities of events of interest are computed only based on the sample space and with no other prior information related to the events. Sometimes it is convenient to compute certain unconditional probabilities by first conditioning on some event, whose probability is easy to find. For example, suppose we draw two cards without replacement from a pack of 52 playing cards. What is the probability that the second card drawn will be an ace of spade?

Conditional probabilities also play an important role in stochastic models, one of the areas that deals with solving real-world problems. For example, consider a communication network that transmits messages from one node to another node. Arriving messages in sending node are stored in a finite buffer and transmitted on a first-come first-served basis. Note that not all messages are going to be admitted because of the finite capacity of the buffer. The computation of the probability of an admitted message finding exactly, say, k messages requires the use of conditional probability. Even waiting time problems have to use conditional probabilities to compute the distribution (and its measures) of the time spent in the system by an arriving customer. Later on, when we deal with Markov chains, conditional probability and conditional expectations play an important role.

Conditional probability is one of the most important concepts in probability and statistics. Although A. de Moivre explicitly introduced the concepts of independence and conditional probability, it was Kolmogorov who studied the notion of conditional probability in the general form. We are often interested in computing probabilities of

certain events when some partial information concerning the results of experiments is given. Also in some calculation of probabilities, it is often convenient to compute them by conditioning on certain events. In probability theory, the models under study are usually described by specifying the appropriate conditional probabilities or conditional distributions. The main topics, such as Bayesian inference, estimation theory, tests of hypotheses and decision theory, in statistics use several notions of conditioning.

DEFINITION 1.2.– *The conditional probability of B given A is defined as:*

$$P(B|A) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(A) > 0.$$

REMARK 1.3.– In the above definition of conditional probability, the requirement that $P(A) > 0$ is very natural, since the conditional probability is defined on the assumption that A has occurred. However, for theoretical purposes $P(B|A)$ is not defined if $P(A) = 0$. But for all practical examples we do not have to worry about this case.

REMARK 1.4.– Suppose that $P(A) > 0$ and $P(B) > 0$. Then one of the following will occur. Either

1) $P(B|A) < P(B)$; in this case we say that A carries negative information about B; or

2) $P(B|A) > P(B)$; in this case we say that A carries positive information about B; or

3) $P(B|A) = P(B)$; in this case we say that A does not contain any information about B.

REMARK 1.5.– It is very easy to show that if A carries negative (positive or no) information about B, then B also carries negative (positive or no) information about A. The concepts of positive and negative information in conditional probability were first introduced by K.L. Chung (1942).

From the discussion of the notion of conditional probability we see that all general theorems on probabilities are also valid for conditional probabilities.

Law of Total Probability: Suppose that $A_i, 1 \leq i \leq n$, are n mutually exclusive and collectively exhaustive events in S. That is, $A_i \cap A_j = \emptyset$, for $i \neq j$, and $\bigcup A_i = S$. Then for any event B in S, we have:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

DEFINITION 1.3.– *Two non-empty events A and B are said to be independent if the occurrence (or non-occurrence) of A does not affect the occurrence (non-occurrence) of B.*

REMARK 1.6.– The following are trivial pairs of independent events: A and $\emptyset$, A and S, S and $\emptyset$.

REMARK 1.7.– The above notion of independence is referred to as pairwise independence. If only two events are considered at any given time then we will freely use either one. However, when more than two events are considered, we need to specify the type of independence such as pairwise or mutually independent (see below).

The notion of independence plays an important role in probability, statistics and other applied areas that rely on probability and statistics. Sometimes the assumption of independence is misused and one has to pay a price for this. An example of this would be the following. In computer communications area, one of the important quantities of interest is the time taken for the token (a tag given to customers for sending packets of messages from one node to another) to return to the starting point. The times between successive arrivals of the token to a particular node are obviously not independent. However, in practice people do assume independence more for the sake of tracking the model mathematically. Depending on the situation this assumption may be a serious one.

DEFINITION 1.4.– *Mathematically non-empty events A and B are independent if any one of the following conditions holds good. Also note that one implies the other two.*

1) $P(B|A) = P(B)$.

2) $P(A \cap B) = P(A)P(B)$.

3) $P(A|B) = P(A)$.

DEFINITION 1.5.– *Events $A_i, 1 \leq i \leq n$, in n in S are said to be mutually independent if for every choice of r distinct indices $i_1, \cdots, i_r$ from among $1, 2, \cdots, n$ and for every $r = 2, 3, \cdots, n$, we have:*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_r}).$$

1.1.1. *Random variables*

In probability and statistics, most of the times the quantities that are of interest are not the outcomes of an experiment under study but rather the values associated with the outcome of the experiment. For example, when n items from the output of the process are inspected the quality control inspector is concerned about the total

number of defective out of the n chosen and the corresponding probabilities rather than the way those defective, if any, were selected. In this section, we review the important concept of a random variable and the probability functions associated with it.

DEFINITION 1.6.– *A random variable is a real-valued function defined on the sample space S into the set of real numbers.*

A random variable is discrete if it takes only discrete values. Examples of such a random variable are: (1) number of defective out of a sample of *n* items chosen; (2) number of cycles before the failure of a pressure regulator that controls downstream water pressure; (3) number of phone calls arriving at an office during a day; (4) number of misprints on a printed page; (5) number of messages arriving at a communication node during an hour; (6) number of industrial accidents in a given week; (7) number of molds made in a month; and (8) number of packages delivered in a week.

A random variable that takes values in an interval is said to be a continuous random variable. Some examples of a continuous random variable are: (1) lifetime of a light bulb; (2) time taken to complete job by a machine; (3) length of a phone call; (4) length of service time at a teller; (5) time between successive arrivals of messages; and (6) delay time in receiving a token in a transmission system.

1.1.2. *Discrete probability functions*

The study of random variables is done through the probability functions associated with them. For a discrete random variable X, the function $f(x)$, defined as $f(x) = P(X = x)$, is called the probability mass function (PMF) of X.

DEFINITION 1.7.– *A given function $f(x)$ is a PMF of a (discrete) random variable if and only if the following conditions are satisfied:*

1) $f(x) \geq 0$, *for all* x, *and*

2) $\sum_x f(x) = 1$.

Some well-known probability mass functions in the context of stochastic modeling used in this book are listed below and for others we refer the reader to any textbook on probability and statistics.

1) *Uniform*:

$$f(x) = \begin{cases} \dfrac{1}{N}, & x = a_1, \cdots, a_N, \\ 0, & \text{elsewhere.} \end{cases}$$

2) *Poisson*:

$$f(x) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, \cdots, \\ \\ 0,\ \text{elsewhere.} \end{cases}$$

3) *Geometric*:

$$f(x) = \begin{cases} p\,(1-p)^x,\ x = 0, 1, \cdots, \\ \\ 0,\ \text{elsewhere.} \end{cases}$$

1.1.3. *Probability generating function*

The probability generating function (PGF) is key in deriving and proving results in stochastic models. So, we review it here.

DEFINITION 1.8.– *Suppose that X is a discrete random variable with $P(X = k) = a_k$. Then the PGF is defined as:*

$$a(z) = \sum_k z^k a_k,\ |z| \leq 1. \qquad [1.1]$$

1.1.4. *Continuous probability functions*

In the continuous case, the probability function similar to PMF of a discrete random variable is called the probability density function (PDF). This function is continuous on its domain except possibly at a finite number of points and the definition is given below. For a continuous random variable X taking values in the interval $(a, b), -\infty < a < b < \infty$, the function $f(x)$ defined as:

$$f(x)dx = P(x < X < x + dx),\ x \in (a, b),$$

is called the probability density function of X.

DEFINITION 1.9.– *The function $f(x)$ is PDF of a continuous random variable X on (a, b) if and only if $f(x)$ satisfies the following conditions:*

1) $f(x) \geq 0$, for all x, and

2) $\int_a^b f(x) = 1$.

Some well-known probability density functions used in this book in the context of stochastic modeling are listed below and for others we refer the reader to any textbook on probability and statistics.

1) *Uniform*:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, \ a \le x \le b, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

2) *Exponential*:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \ x \ge 0, \lambda > 0, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

3) *Erlang of order* m:

$$f(x) = \begin{cases} \dfrac{\lambda^m}{(m-1)!} x^{m-1} e^{-\lambda x}, \ x \ge 0, \lambda > 0, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

4) *Hyperexponential of order* m *with mixing probability vector* $\boldsymbol{p}$ *and the parameter vector* $\boldsymbol{\lambda}$:

$$f(x) = \begin{cases} \sum\limits_{k=1}^{m} p_j \lambda_j e^{-\lambda_j x}, \ x \ge 0, \lambda_j > 0, 1 \le j \le m, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

5) *Gamma*:

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \ \alpha > 0, \beta > 0, x \ge 0, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

6) *Weibull*:

$$f(x) = \begin{cases} \dfrac{\alpha}{\beta} \left(\dfrac{x}{\beta} \right)^{\alpha-1} e^{-(x/\beta)^\alpha}, \ \alpha > 0, \beta > 0, x \ge 0, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

7) *Beta*:

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \ \alpha > 0, \beta > 0, 0 \le x \le 1, \\ \\ 0, \ \text{elsewhere.} \end{cases}$$

[Note: In the above $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.]

Whether a random variable, X, is discrete or continuous or a combination of both, the (cumulative) distribution function of X is defined as:

DEFINITION 1.10.– *The (cumulative) distribution function (CDF) of a random variable X, denoted as, say, $F(x)$, is defined as:*

$$F(x) = P(X \leq x),\ -\infty < x < \infty. \quad [1.2]$$

REMARK 1.8.– Note that the CDF is a right-continuous and non-decreasing function, which tends to 0 as $x \to -\infty$ and goes to 1 as $x \to \infty$.

1.1.5. *Laplace transform and Laplace-Stieltjes transform*

In this section, we briefly discuss the Laplace transform (LT) that plays an important role in stochastic modeling.

Since we are focusing on queueing and related topics in this two-volume book, we assume that the underlying random variables are all non-negative.

DEFINITION 1.11.– *Suppose that $f(x)$ is the PDF of a non-negative random variable, X. The Laplace transform (LT) of $f(x)$ (or equivalently of X) is defined as:*

$$f^*(s) = \int_0^\infty e^{-s\,x} f(x)dx,\ Re(s) \geq 0. \quad [1.3]$$

DEFINITION 1.12.– *Suppose that $F(x)$ is the CDF of a non-negative random variable, X. The Laplace-Stieltjes transform (LST) of $F(x)$ (or equivalently of X) is defined as:*

$$F^*(s) = \int_0^\infty e^{-s\,x} dF(x),\ Re(s) \geq 0. \quad [1.4]$$

DEFINITION 1.13.– *Suppose that $\bar{F}(x) = P(X > x)$ is the tail probability function of a non-negative random variable, X. The Laplace-Stieltjes transform (LST) of $\bar{F}(x)$ is defined as:*

$$\bar{F}^*(s) = \int_0^\infty e^{-s\,x} d\bar{F}(x),\ Re(s) \geq 0. \quad [1.5]$$

REMARK 1.9.– Since probabilistic interpretation plays a key role in stochastic modeling, we give one such an interpretation for the LST of a continuous random variable X with PDF given by $f(.)$. Suppose that a catastrophic event occurs and that it could affect (or kill) the lifetime, say, X, of a random component. Suppose that X has the PDF given by $f(.)$ and that Y, with an exponential distribution with parameter s, is used to model the catastrophic event. Then the LST of $f(x)$ gives the probability that the lifetime of the component is not affected during its lifetime by the catastrophic event.

RESULT 1.1.– Since $F(t) = \int_0^t f(x)dx$, the LST of $F(t)$ is same as the LT of $f(t)$. That is:

$$F^*(s) = \int_0^\infty e^{-sx} dF(x) = \int_0^\infty e^{-sx} f(x)dx, \;\; Re(s) \geq 0. \qquad [1.6]$$

RESULT 1.2.– The LT of $F(t)$ and the LT of $\bar{F}(t)$ are given by:

$$\hat{F}^*(s) = \int_0^\infty e^{-sx} F(x)dx = \int_0^\infty e^{-sx} \int_0^x f(y)dydx = \frac{f^*(s)}{s}, \; Re(s) > 0,$$
$$\hat{\bar{F}}^*(s) = \int_0^\infty e^{-sx} \int_x^\infty f(y)dydx = \frac{1 - f^*(s)}{s}, \; Re(s) > 0, \qquad [1.7]$$

RESULT 1.3.– (Abelian theorem)

For some $a > 0$, $\lim_{t\to\infty} \frac{F(t)}{t^a} = \frac{b}{\Gamma(a+1)} \;\Rightarrow\; \lim_{s\to 0+} s^a \, F^*(s) = b.$

RESULT 1.4.– (Tauberian theorem)

For some $a > 0$, $\lim_{s\to 0+} s^a \, F^*(s) = b \;\Rightarrow\; \lim_{t\to\infty} \frac{F(t)}{t^a} = \frac{b}{\Gamma(a+1)}.$

1.1.6. *Measures of a random variable*

So far, we discussed how a random variable is studied through the probability functions associated with it. Most of the time the information conveyed in the probability functions can be effectively summarized by their general shapes and locations of certain parameters. For a great many distributions these characteristics can be fully described by a small number of numerical quantities (also referred to as measures) that are unique to the distributions under study. Also in many instances one is not interested in studying the random variable in detail but only to get some idea about the nature of the random variable itself. The following example should motivate the need for this section.

Suppose that one is planning to buy a new model car. Among all other criteria for buying a new car, let us assume that person gives priority for a car that gives good mileage. The MPG (miles per gallon) of a new model car is a random variable (why?). But the person is not interested in knowing the probability distribution of this random variable, only the average MPG. Of course, the average MPG depends on a number of variables, such as the size of the car, power of the engine and type of transmission. However, this measure will give the person a smaller set of cars to pick from. The key point here is how one or more measures of a random variable is used in practice. There are several other instances, which will be seen throughout the book.

Some commonly used measures are: (1) nth raw moment, especially the first (also referred to as expected value) and second moments; (2) standard deviation; and (3) percentiles.

The definitions of these measures are given below.

DEFINITION 1.14.– *The* nth raw moment*: of a random variable is defined as:*

$$E(X^n) = \begin{cases} \sum_{x \in S} x^n f(x), \ X \text{ is discrete,} \\ \\ \int_0^\infty x^n f(x) dx, \ X \text{ is continuous.} \end{cases} \quad [1.8]$$

DEFINITION 1.15.– *The standard deviation of a random variable* X *is defined as* Standard deviation (σ_X)*:*

$$\sigma_X = \sqrt{E(X^2) - [E(X)]^2} \quad [1.9]$$

DEFINITION 1.16.– *The* $100p$*th percentile of a random variable* X *is defined as:* $100p$th Percentile*: The* $100p$*th percentile of a random variable* X*, denoted by* x_p*, for* $0 < p < 1$*, is the solution to the equation:*

$$F(x_p) = P(X \leq x_p) = p. \quad [1.10]$$

1.2. Renewal process

A renewal process is a sequence of random variables, $\{X_n : n \geq 1\}$, having a common probability (mass or density, depending on whether the random variables are discrete or continuous) function. Suppose that we are dealing with continuous random variables and let $f(x)$ and $F(x)$, respectively, denote the common PDF and CDF of $X_1, X_2, \cdots$. Below, we summarize some key observations and results, and refer the reader to Cox (1962) and Karlin and Taylor (1975) for more details.

DEFINITION 1.17.– *If* X_1 *also has the same PDF as the rest of* X_n*, the renewal process* $\{X_n : n \geq 0\}$ *is referred to as an ordinary renewal process.*

DEFINITION 1.18.– *If* X_1 *has a different PDF than the rest of* X_n*, the renewal process* $\{X_n : n \geq 0\}$ *is referred to as a modified renewal process.*

DEFINITION 1.19.– *If* X_1 *has the PDF given by* $\mu \bar{F}(x)$*, where* μ^{-1} *is the mean of the random variable with CDF given by* $F(x)$*, the renewal process* $\{X_n : n \geq 0\}$ *is referred to as a stationary (or equilibrium) renewal process.*

REMARK 1.10.– Note that for a Poisson process (see section 1.2.3), the ordinary renewal process and the stationary renewal process are identical.

1.2.1. *Renewal function*

Suppose that $N(t)$ denotes the number of renewals in $(0, t]$ corresponding to an ordinary renewal process.

DEFINITION 1.20.– *The renewal function, $M(t)$, is defined as the expected number of renewals in t units of time. That is, $M(t) = E(N(t))$.*

Suppose that $S_n = \sum_{i=1}^{n} X_i$. Then, we have the following key results.

RESULT 1.5.– $N(t) \geq n \Longleftrightarrow S_n \leq t$.

RESULT 1.6.– Result 1.5 implies that:

$$P(N(t) \geq n) = P(S_n \leq t) = F^{(n)}(t), \qquad [1.11]$$

where $F^{(n)}(.)$ is the n-fold convolution of $F(.)$.

RESULT 1.7.– The renewal function is given by:

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} nP(N(t) = n) = \sum_{n=1}^{\infty} F^{(n)}(t),\ t \geq 0. \qquad [1.12]$$

One of the most celebrated equations in renewal theory is the famous renewal equation. This is obtained by conditioning on the first renewal.

RESULT 1.8.– The renewal equation corresponding to the renewal process $\{X_n\}$ is given by:

$$M(t) = F(t) + \int_0^t M(t - x)dF(x),\ t \geq 0, \qquad [1.13]$$

whose solution is given by:

$$M(t) = F(t) + \int_0^t F(t - x)dM(x),\ t \geq 0. \qquad [1.14]$$

REMARK 1.11.– It is worth pointing out how the solution for equation [1.13] is obtained as the function given in equation [1.14] by replacing the functions $M(.)$ and $F(.)$ inside the integral in equation [1.13] with $F(.)$ and $M(.)$, respectively.

Suppose that $G(t, z) = \sum_{n=0}^{\infty} z^n P(N(t) = n),\ t \geq 0, |z| \leq 1$, denotes the probability generating function of $N(t)$ and $g^*(s, z)$ is the LST of $G(t, z)$. That is, $g^*(s, z)$ is the joint transform.

RESULT 1.9.– It is easy to verify that, for $|z| \leq 1, Re(s) > 0$:

$$g^*(s,z) = \begin{cases} \dfrac{1-f^*(s)}{s[1-zf^*(s)]}, \text{ ordinary renewal process,} \\ \dfrac{1-zf^*(s)+zf_1^*(s)-f_1^*(s)}{s[1-zf^*(s)]}, \text{ modified renewal process,} \\ \dfrac{1}{s}+\dfrac{\mu(z-1)[1-f^*(s)]}{s^2[1-zf^*(s)]}, \text{ stationary renewal process.} \end{cases} \quad [1.15]$$

RESULT 1.10.– The Laplace transform, $M^*(s)$, of $M(t)$ under various renewal processes is:

$$M^*(s) = \begin{cases} \dfrac{f^*(s)}{s[1-f^*(s)]}, \text{ ordinary renewal process,} \\ \dfrac{f_1^*(s)}{s[1-f^*(s)]}, \text{ modified renewal process,} \\ \dfrac{\mu}{s^2}, \text{ stationary renewal process.} \end{cases} \quad [1.16]$$

REMARK 1.12.– We note that for the stationary renewal process, $M(t) = \dfrac{t}{\mu'}$, where $\mu' = \dfrac{1}{\mu}$ is the mean of the underlying random variable.

RESULT 1.11.– In the case of an ordinary renewal process, asymptotically the renewal function approaches $\dfrac{t}{\mu'}$. That is:

$$\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{\mu'} = \mu. \quad [1.17]$$

Suppose that $m(t)$ denotes the renewal density. That is, $m(t) = M'(t)$. Note that $m(t)dt$ gives the expected number of renewals in $(t, t+dt)$.

RESULT 1.12.– Suppose that $m^*(s)$ denotes the LT of $m(t)$. It is easy to verify that:

$$m^*(s) = \begin{cases} \dfrac{f^*(s)}{1-f^*(s)}, \text{ ordinary renewal process,} \\ \dfrac{f_1^*(s)}{1-f^*(s)}, \text{ modified renewal process.} \end{cases} \quad [1.18]$$

RESULT 1.13.– In the case of an ordinary renewal process, asymptotically the renewal density, $m_o(t)$, approaches $\frac{1}{\mu'}$. That is:

$$\lim_{t\to\infty} m_o(t) = \frac{1}{\mu'} = \mu. \qquad [1.19]$$

RESULT 1.14.– Suppose that $m_1^*(s)$ is the LT of the density, $m_1(t)$, of the modified renewal process. Then using equation [1.18] we get:

$$m_1^*(s) = f_1^*(s) + m_1^*(s)f^*(s), \qquad [1.20]$$

from which we get the following integral equation. This equation plays an important role in stochastic modeling.

$$m_1(t) = f_1(t) + \int_0^t m_1(t-x)f(x)dx, \qquad [1.21]$$

whose solution can be verified to be:

$$m_1(t) = f_1(t) + \int_0^t f_1(t-x)m_o(x)dx. \qquad [1.22]$$

RESULT 1.15.– The integral equation for the ordinary renewal process can be obtained with a similar argument that leads to result 1.8 or simply differentiating equation [1.13] and is given by:

$$m_o(t) = f(t) + \int_0^t m_o(t-x)f(x)dx, \qquad [1.23]$$

whose solution can be verified to be

$$m_o(t) = f(t) + \int_0^t f(t-x)m_o(x)dx. \qquad [1.24]$$

REMARK 1.13.– Noting that $f(x)dx$ gives the probability that in $(x, x+dx)$ a (first) renewal occurs, the integral equation given in equation [1.23] can be given a nice probabilistic interpretation: the LHS gives the probability of a renewal in a small time interval near t and the RHS is the sum of a probability of a (first) renewal near t and the probability that there is a renewal near $t-x$, which is followed by an inter-renewal time of duration $x, 0 < x < t$.

RESULT 1.16.– Suppose that $r(t)$ is defined as the forward recurrence time (also referred to as residual lifetime). That is, $r(t) = t - S_{N(t)}$. If $f_r(x,t)$ denotes the probability density function of $r(t)$, then it is known that (see, e.g. Cox (1962))

$$f_r(t,x) = f(t+x) + \int_0^t h(t-u)f(u+x)du. \qquad [1.25]$$

RESULT 1.17.– The limiting density function of $r(t)$ is given by $\bar{F}(x)/\mu$ as $t \to \infty$ and the LST of the limiting density is given by $\dfrac{1 - f^*(s)}{s\mu}$.

DEFINITION 1.21.– Directly Riemann integrable (DRI) functions*: A real-valued function defined on* $(0, \infty)$ *is said to be DRI if and only the partial (Riemann) sums taken over the intervals partitioned on* $(0, \infty)$ *are finite and both converge to the same limit.*

RESULT 1.18.– A function satisfying any of the following conditions guarantees the function to be DRI (note that these are not necessary!).

1) The function is non-negative, continuous and has a finite support.

2) The function is non-negative, continuous, bounded such that the upper Riemann sum is bounded.

3) The function is non-negative, monotone non-increasing and Riemann integrable.

4) The function is non-negative and bounded above by a DRI function.

RESULT 1.19.– *Key Renewal theorem*: Suppose that $k(t)$ is a DRI function. Then, we have:

$$\begin{aligned} &\lim_{t\to\infty} \int_0^t k(t-x)dM(x) = \lim_{t\to\infty} k * M(t) \\ &= \lim_{t\to\infty} M * k(t) = \frac{1}{\mu'}\int_0^\infty k(t)dt. \end{aligned} \quad [1.26]$$

1.2.2. *Terminating renewal process*

In the previous section, we talked about renewal process, which does not terminate. That is, $F(\infty) = 1$. However, there are times when we need to study terminating, also referred to as transient, renewal process. That is, we can have a situation wherein $F(\infty) < 1$. In this case, the integral equation for renewals as well as the key renewal theorem need to be applied differently.

DEFINITION 1.22.– *A renewal process,* $\{X_n : n \geq 1\}$*, is said to be a terminating renewal process if its common CDF,* $F(.)$*, is such that* $F(\infty) < 1$.

RESULT 1.20.– For a terminating renewal process, $\{X_n : n \geq 1\}$, with CDF, $F(.)$, the associated counting process, $N(t)$, is such that $N = N(\infty) < \infty$, almost surely. That is, the number of renewals, N, during $[0, \infty]$ is finite almost surely. Furthermore,

$$P(N = k) = F^{k-1}(\infty)(1 - F(\infty)),\ k \geq 1. \quad [1.27]$$

RESULT 1.21.– The renewal function, $M(t)$, for the terminating renewal process is such that we have:

$$M(\infty) = \frac{1}{1 - F(\infty)}. \quad [1.28]$$

RESULT 1.22.– For any DRI function, $k(t)$ with $k(\infty) = \lim_{t\to\infty} k(t)$ exists, we have:

$$\lim_{t\to\infty} k * M(t) = \frac{k(\infty)}{1 - F(\infty)}. \quad [1.29]$$

RESULT 1.23.– If the terminating renewal process, $\{X_n : n \geq 1\}$, is a delayed one (i.e. the initial one has a different distribution function, say, $H(.)$) then, for any DRI function, $k(t)$ with $k(\infty) = \lim_{t\to\infty} k(t)$ exists, we have:

$$\lim_{t\to\infty} k * M(t) = \frac{k(\infty)H(\infty)}{1 - F(\infty)}. \quad [1.30]$$

1.2.3. *Poisson process*

One of the most celebrated stochastic processes is the Poisson process. That is, a renewal process whose inter-arrival times follow an exponential distribution. We will briefly summarize some key results related to Poisson processes for our needs here. Any additional ones needed will be mentioned in appropriate places.

RESULT 1.24.– The PDF, CDF, and the LST are:

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t},\ t \geq 0, \\ F(t) &= 1 - e^{-\lambda t},\ t \geq 0, \\ f^*(s) &= \frac{\lambda}{s + \lambda}. \end{aligned} \quad [1.31]$$

RESULT 1.25.– The exponential distribution possesses the famous memoryless property:

$$P(X > t + t_1 | X > t_1) = P(X > t),\ t_1, t \geq 0. \quad [1.32]$$

RESULT 1.26.– The counting process, $N(t)$, denoting the number of Poisson arrivals by time t has the following properties:

1) $P[N(t) = k] = e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ k = 0, 1, 2, \cdots$.

2) Possess independent increments. That is, $N(t)$ is independent of $N(t+s) - N(t)$.

3) Possess stationary property. That is, the counting process depends only on the lag as opposed to the actual time point. That is, $N(t+s)-N(t)$ has the same statistical property as $N(u+s) - N(u)$.

RESULT 1.27.– The renewal function, $M(t)$, for the Poisson process with parameter λ is obviously a linear function. That is, $M(t) = \lambda\, t$, for $t \geq 0$. One can see this by looking at result 1.17 for the current case noting that $f^*(s) = \dfrac{\lambda}{\lambda + s}$ yields $M^*(s) = \dfrac{\lambda}{s^2}$ implying $M(t) = \lambda\, t$.

RESULT 1.28.– We have:

$$P(N(t) \geq n) = P(S_n \leq t) = \sum_{i=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^i}{i!},\ n\ \geq\ 0, \qquad [1.33]$$

from which the density function of S_n is obtained as:

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!},\ n \geq 0,\ t \geq 0. \qquad [1.34]$$

REMARK 1.14.– The result in equation [1.34] is intuitively obvious. In order for the nth renewal to occur in $(t, t+dt)$, $n-1$ renewals should have occurred by time t and in the small interval a renewal occurs. Note that this PDF corresponds to the celebrated Erlang random variable, which will be used extensively in this two-volume book.

REMARK 1.15.– The above density function given in equation [1.34] is that of Erlang (a gamma family).

RESULT 1.29.– The superposition of two or more Poisson processes is again a Poisson process.

RESULT 1.30.– Given that exactly one Poisson event has occurred by time t, the distribution of the time of the occurrence of this event is uniform on $[0, t]$. That is:

$$\begin{aligned} P(X_1 < s | N(t) = 1) &= \frac{P(X_1 < s, N(t) = 1)}{P(N(t) = 1)} \\ &= \frac{P(N(s) = 1, N(t-s) = 0)}{P(N(t) = 1)} = \frac{s}{t}. \end{aligned} \qquad [1.35]$$

RESULT 1.31.– Given that $N(t) = n$, the joint distribution of the arrival times, $S_1, \cdots, S_n$, is that of the joint distribution of the order statistics of n uniformly

distributed random variables on $(0,t)$. That is, for $0 < u_1 < \cdots < u_n < t$, the conditional density is given by:

$$f(u_1, \cdots, u_n | N(t) = n)$$
$$= \frac{P(N(u_1) = 1, N(u_2 - u_1) = 1, \cdots, N(u_n - u_{n-1}) = 1, N(t - u_n) = 0)}{P(N(t) = n)}$$
$$= \frac{n!}{t^n}.$$

1.3. Matrix analysis

Matrix theory plays an important role in many areas such as business, economics, statistics, engineering, finance, stochastic modeling and other applied fields. Also, it is fairly easy to introduce this subject at the undergraduate level for the students to get familiarize with the concepts as well as to apply them to advanced fields such as Markov chains and queues.

In this section, we briefly summarize some of the key results and properties of matrices that are crucial to MAM. For details we refer to books such as Dhrymes (2013); Marcus and Minc (1964); Graham (1981); Seneta (2006); and Steeb and Hardy (2011).

1.3.1. *Basics*

DEFINITION 1.23.– *A matrix A is an array of elements arranged in rows and columns. It is defined through the dimension and the (i,j)th entries. Thus, an $m \times n$ matrix A is defined as:*

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,j} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,j} & \cdots & a_{m,n} \end{pmatrix}. \quad [1.36]$$

REMARK 1.16.– When the dimension of a matrix needs to be displayed, we will do so by labeling it as subscripts. Thus, an $m \times n$ matrix A will be denoted as $A_{m \times n}$. Also, uppercase letters will be used to denote a matrix and the corresponding lowercase letter for its elements.

DEFINITION 1.24.– *The transpose of $A_{m \times n} = (a_{i,j})$ is the $n \times m$ matrix obtained by interchanging the rows and columns. The transpose notation used in this book will be A^T. Thus, $A^T = (a_{j,i})$.*

REMARK 1.17.– If $m = n$, we say that A is a square matrix and will be denoted by A_n when the dimension needs to be displayed.

DEFINITION 1.25.– *A square matrix A is symmetric if and only if $A = A^T$.*

DEFINITION 1.26.– *A square matrix A is said to be a diagonal matrix if and only if $a_{i,j} = 0$, for all $i \neq j$. In this case, it is sometimes convenient to display as $A = \Delta\{a_{1,1}, \cdots, a_{m,m}\}$.*

REMARK 1.18.– A diagonal matrix such that all its diagonal entries are 1 is called an identity matrix and will be denoted by $I_m = \Delta\{1, \cdots, 1\}$.

DEFINITION 1.27.– *A square matrix A is said to be reducible if after rearranging the rows and the columns (i.e. some permutations of rows and columns) it can be written as:*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}. \qquad [1.37]$$

If A cannot be written as above, then it is said to be irreducible.

RESULT 1.32.– A is irreducible $\Leftrightarrow A - \Delta(A)$ is irreducible.

DEFINITION 1.28.– *The basic matrix operations are: (1) addition of matrices of the same dimensions. Thus, $A_{m\times n} + B_{m\times n} = C_{m\times n}$ with $c_{i,j} = a_{i,j} + b_{i,j}$; (2) scalar multiplication of a matrix: $dA = (d\, a_{i,j})$; (3) multiplication of two matrices requires that the number columns of the left side matrix equals the number of rows on the right. Thus, the matrix product, $A_{m\times n}B_{q\times r}$ makes sense if and only if $n = q$. Similarly, $B_{q\times r}A_{m\times n}$ makes sense if and only if $r = m$. The product $A_{m\times n}B_{n\times r}$ yields a matrix $C_{m\times r} = (c_{i,j})$, where $c_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j}$.*

DEFINITION 1.29.– *Hadamard (or Schur) product of two matrices, $A_{m\times n}$ and $B_{m\times n}$, denoted as $A \circ B$, is defined $A \circ B = B \circ A = (a_{i,j}b_{i,j})$. That is, one takes element-wise products.*

DEFINITION 1.30.– *A square matrix A_m is said to be stable if and only if the following three conditions are satisfied:*

1) $a_{i,i} < 0$, for all $1 \leq i \leq m$;

2) $a_{i,j} \geq 0$, for all $1 \leq i, j \leq m$;

3) $\sum_{j=1}^{m} a_{i,j} \leq 0$, for all $1 \leq i \leq m$, and at least for one i, $\sum_{j=1}^{m} a_{i,j} < 0$.

DEFINITION 1.31.– *A square matrix A_m is said to be semi-stable if and only if the following three conditions are satisfied:*

1) $a_{i,i} \leq 0$, *for all* $1 \leq i \leq m$;

2) $a_{i,j} \geq 0$, *for all* $1 \leq i, j \leq m$;

3) $\sum_{j=1}^{m} a_{i,j} \leq 0$, *for all* $1 \leq i \leq m$.

REMARK 1.19.– Note that a stable matrix is always semi-stable but the converse is not true.

DEFINITION 1.32.– *The rank of A is defined as the maximum number of linearly independent rows (or columns) of A.*

REMARK 1.20.– The rank of $A_{m\times n}$ cannot exceed the minimum of m and n.

DEFINITION 1.33.– *A square matrix A_m is said to have an inverse or simply be non-singular, if and only if there exists a matrix B_m such that $AB = BA = I_m$. We will denote by A^{-1} the inverse of A.*

REMARK 1.21.– The rank of A_m is m if and only if A is non-singular. In this case, it is also said that A has full rank.

DEFINITION 1.34.– *A square matrix A_m is said to be a Toeplitz matrix if $a_{i,j} = a_{i+r,j+r}$, for all i, j, and r. That is, $a_{i,j}$ depends only on the difference $|i - j|$.*

REMARK 1.22.– It is worth pointing out that while we do not take the approach of discussing MAM in the context of Toeplitz and asymptotically Toeplitz matrices. However, if one is interested to know about that approach we recommend the books by Bini et al. (2005) and Dudin et al. (2020).

REMARK 1.23.– The variance-covariance matrix seen in regression analysis is a classical example of a Toeplitz matrix. In general, Toeplitz matrix is of the form:

$$A = (a_{i,j}) = \begin{pmatrix} a_0 & a_1 & \cdots & a_{j-1} & \cdots & a_{m-1} \\ a_{-1} & a_0 & \cdots & a_{j-2} & \cdots & a_{m-2} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{1-m} & a_{2-m} & \cdots & a_{j-m} & \cdots & a_0 \end{pmatrix} \quad [1.38]$$

DEFINITION 1.35.– *A matrix A is non-negative if and only if all its elements are non-negative.*

REMARK 1.24.– If A is a non-negative and an irreducible square matrix, then there exists a positive integer, say, n^* such that $A^n > 0,\ n \geq n^*$.

DEFINITION 1.36.– *The trace, $tr(A)$, of a square matrix A is defined as the sum of the diagonal elements. That is, $tr(A) = \sum_{i=1}^{m} a_{i,i}$.*

REMARK 1.25.– If A and B are square matrices, then $tr(A+B) = tr(A) + tr(B)$ and $tr(AB) = tr(BA)$.

DEFINITION 1.37.– *The determinant, denoted by $|A|$, of a square matrix A is defined as:*

$$|A| = \sum_{\{j_1,\cdots,j_m\}} (-1)^r a_{1,j_1}\, a_{2,j_2}, \cdots, a_{m,j_m}, \qquad [1.39]$$

where the summation is taken over (distinct) permutations of the integers $\{1, 2, \cdots, m\}$ and r is taken to be 0 or 1 depending on whether the number of transpositions needed to bring the permutation $\{j_1, \cdots, j_m\}$ back to the natural order of $(1, 2, \cdots, m)$ is even or odd.

REMARK 1.26.– For a square matrix A_m, we have $|A^T| = |A|$. Further, if $B = dA$, then $|B| = d^m|A|$.

DEFINITION 1.38.– *Suppose that we obtain B_{m-1} from A_m by deleting the ith row and the jth column. The quantity $(-1)^{i+j}|B_{m-1}|$ is called the cofactor of the element $a_{i,j}$.*

DEFINITION 1.39.– *Suppose that we obtain B_{m-1} from A_m by deleting the ith row and the jth column. The quantity $(-1)^{i+j}|B_{m-1}|$ is called the cofactor of the element $a_{i,j}$.*

DEFINITION 1.40.– *For a square matrix A, the adjoint of A, denoted usually by $adj(A)$, is a square matrix whose (i,j)th element is the cofactor, say, $a^*_{j,i}$, of the (j,i)th element of A. Or equivalently, the $adj(A) = (A^*)^T$, where $A^* = (a^*_{i,j})$.*

REMARK 1.27.– In terms of cofactors, the determinant of A is obtained as:

$$|A| = \sum_{j=1}^{m} a_{i,j} a^*_{i,j} = \sum_{i=1}^{m} a_{i,j} a^*_{i,j}. \qquad [1.40]$$

REMARK 1.28.– In terms of cofactors, the inverse of a non-singular matrix A is obtained as:

$$A^{-1} = \frac{1}{|A|} adj(A). \qquad [1.41]$$

RESULT 1.33.– If $A_{m\times n}$ and $B_{n\times m}$, then we have:

$$(I_m + A\,B)^{-1} \text{ exists } \Leftrightarrow (I_n + B\,A)^{-1} \text{ exists.}$$

The following result is a matrix-analog of the binomial theorem (when the matrices A and B commute). Since to the author's knowledge a proof of this result is not seen in the literature, a proof is given.

RESULT 1.34.– Suppose that A and B are square matrices of order m. Define the matrix polynomial $f(z)$ as $f(z) = (A + zB)^n$. Then, we have:

$$f(z) = \sum_{k=0}^{n} z^k F_{k,n},$$

where the square matrices, $F_{k,n}$, of order m, are recursively computed (in that order) as follows:

$$F_{0,0} = I_m, \;\; F_{0,1} = A, \;\; F_{1,1} = B,$$

$$F_{0,r} = A\, F_{0,r-1}, \;\; 2 \leq r \leq n,$$

$$F_{i,r} = A\, F_{i,r-1} + B\, F_{i-1,r-1}, \;\; 1 \leq i \leq r-1, \;\; 2 \leq r \leq n,$$

$$F_{r,r} = B\, F_{r-1,r-1}, \;\; 2 \leq r \leq n.$$

PROOF.– Suppose we define an upper triangular matrix of K order $(n+1)m$ as:

$$K = \begin{bmatrix} A & B & & & & \\ & A & B & & & \\ & & A & B & & \\ & & & \ddots & & \\ & & & & A & B \\ & & & & & A \end{bmatrix}.$$

Verify that K^r is also upper triangular and is of the form:

$$K^r = \begin{bmatrix} F_{0,r} & F_{1,r} & F_{2,r} & F_{3,r} & F_{4,r} & \cdots & F_{n,r} \\ & F_{0,r} & F_{1,r} & F_{2,r} & F_{3,r} & \cdots & F_{n-1,r} \\ & & F_{0,r} & F_{1,r} & F_{2,r} & \cdots & F_{n-2,r} \\ & & & \ddots & \ddots & \ddots & \ddots \\ & & & & & F_{0,r} & F_{1,r} \\ & & & & & & F_{0,r} \end{bmatrix}, \; 2 \leq r \leq n.$$

Note that in K^r, the matrices $F_{k,r} = 0,\ k \geq r+1$. The recursive equations can easily be derived. For example, looking at $K^r = K\,K^{r-1}$ and applying the matrix multiplication rule one sees the recursive equations. The stated result on the expansion of $f(z)$ will follow immediately once we show that the matrices $F_{k,r}$ are the coefficients of $z^k,\ 0 \leq k \leq r$, in the matrix expansion of $(A + z\,B)^r$. To see this, we apply the induction method. First note that the result is true for $r = 2$ by observing $(A + z\,B)^2 = (A + z\,B)(A + z\,B) = A^2 + z(AB + BA) + z^2B^2 = F_{0,2} + zF_{1,2} + z^2F_{2,2}$.

Assume now that the result is true for $(A + z\,B)^i,\ i = 1, \cdots, r-1$. Thus, using the fact that $F_{k,r-1}$ is the coefficient of $z^k,\ k = 0, \cdots, r-1$, in the expansion of $(A + z\,B)^{r-1}$, we have:

$$(A + z\,B)^r = (A + z\,B)\sum_{k=0}^{r-1} z^k F_{k,r-1},$$

which on expanding gives:

$$(A + z\,B)^r = A\,F_{0,r-1} + \sum_{k=1}^{r-1} z^k\,[A\,F_{k,r-1} + B\,F_{k-1,r-1}] + z^r B\,F_{r-1,r-1},$$

which from the recursive equations shows that the result is true for $i = r$. This completes the proof. □

RESULT 1.35.– If A and B commute, then we have:

$$f(z) = \sum_{k=0}^{n} \binom{n}{k} z^k\,A^{n-k}\,B^k. \qquad [1.42]$$

PROOF.– First note that $F_{k,n}$ has $\binom{n}{k}$ terms involving the products of the powers of A and B. When A and B commute, all different ways of producing the product of A^{n-k} and B^k will result in the same product. □

1.3.2. *Eigenvalues and eigenvectors*

DEFINITION 1.41.– *Given a square matrix A_m, the associated characteristic equation is defined as the mth degree polynomial obtained as $det(\xi I - A)$.*

DEFINITION 1.42.– *Given a square matrix A, the roots of its characteristic equation are called eigenvalues or characteristic roots. That is, ξ is said to be an eigenvalue of A, if it satisfies:*

$$det(\xi I - A) = 0. \qquad [1.43]$$

DEFINITION 1.43.– *Given a square matrix A, the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are, respectively, called the left and right eigenvectors corresponding to the eigenvalue, ξ, if :*

$$\boldsymbol{u}\, A = \xi \boldsymbol{u}, \quad \text{and} \quad A\, \boldsymbol{v} = \xi \boldsymbol{v}. \qquad [1.44]$$

RESULT 1.36.– Suppose that A is a square matrix and that ξ is an eigenvalue of A. Then for any matrix $B = a_1\, I + a_2\, A$, where a_1 and $a_2 \neq 0$ are scalars, then an eigenvalue of B is of the form $a_1 + a_2\, \xi$. The (left or right) eigenvectors of A and B are identical.

RESULT 1.37.– Suppose that A is non-singular, then the eigenvalues of A are non-zero. Further, the eigenvalues of A^{-1} are reciprocals of the eigenvalues of A.

RESULT 1.38.– The eigenstructure, namely, the eigenvalues and their corresponding eigenvectors, of A and A^T are the same.

RESULT 1.39.– Suppose that $f(.)$ is a polynomial. Then, if ξ is an eigenvalue of A, then $f(\xi)$ is an eigenvalue of $f(A)$.

DEFINITION 1.44.– *Two square matrices, A and B, are said to be similar if there exists a non-singular matrix, say, P, such that $P^{-1}AP = B$. The matrix P is referred to as a similarity transformation matrix.*

RESULT 1.40.– Similar matrices have identical eigenstructures.

DEFINITION 1.45.– *A square matrix A is said to be diagonalizable if A is similar to a diagonal matrix.*

RESULT 1.41.– A square matrix is diagonalizable if and only if there exists a linearly independent set of eigenvectors and they are used as columns of P in $P^{-1}AP = \Delta$, where Δ is a diagonal matrix.

RESULT 1.42.– Eigenvectors corresponding to distinct eigenvalues of A are linearly independent.

DEFINITION 1.46.– *Algebraic multiplicity of an eigenvalue corresponds to the multiplicity of the eigenvalue (or root).*

DEFINITION 1.47.– *Geometric multiplicity of an eigenvalue corresponds to the dimension of the subspace of the eigenvector of the eigenvalue.*

RESULT 1.43.– Algebraic multiplicity of an eigenvalue is always greater than or equal to the geometric multiplicity of the eigenvalue.

RESULT 1.44.– A is diagonalizable $\Leftrightarrow$ algebraic multiplicity of ξ = geometric multiplicity of ξ, for every ξ of A.

DEFINITION 1.48.– *A square matrix A is orthogonal if $A^T = A^{-1}$. That is, $A\,A^T = A^T\,A = I$ implying that the columns of A are linearly independent and the inner products are 0 or 1.*

RESULT 1.45.– Suppose that A and B are square matrices. Then AB and BA have identical eigenstructure.

RESULT 1.46.– Suppose that A_m is a square matrix. Then there exists a B_m such that $\sum_{i,j} |a_{i,j} - b_{i,j}| < \epsilon$ and B has distinct eigenvalues.

RESULT 1.47.– Suppose that A is a non-negative square matrix. Suppose that $\boldsymbol{u} \geq \boldsymbol{v}$. Then $A\,\boldsymbol{u} \geq A\,\boldsymbol{v}$.

RESULT 1.48.– Suppose that A is a positive square matrix. Then, for all $\boldsymbol{u} \geq \mathbf{0}$, we have $A\,\boldsymbol{u} \geq \mathbf{0}$.

RESULT 1.49.– Suppose that A is a positive square matrix. Then the maximum eigenvalue of A is positive.

RESULT 1.50.– *Perron-Frobenius Result*: Suppose that A is a non-negative square matrix of dimension n. Then we have the following:

1) There exists an eigenvalue, say, η, such that $\eta > 0$.

2) There exists strictly positive left and right eigenvectors associated with η.

3) Suppose that $\boldsymbol{u}$ and $\boldsymbol{v}$ are, respectively, left and right eigenvectors of A corresponding to η, that is, $\boldsymbol{u}A = \eta\boldsymbol{u}$ and $A\boldsymbol{v} = \eta\boldsymbol{v}$. Then, we can normalize them so as to have $\boldsymbol{u}\boldsymbol{v} = 1$.

4) η is the maximal eigenvalue. That is, if r is any other eigenvalue, then $|r| < \eta$.

5) Eigenvectors associated with η are unique up to a constant multiple.

6) If B is any other matrix of dimension n with $0 \leq B \leq A$ and if ξ is an eigenvalue of B, then $|\xi| \leq \eta$. Further, if $|\xi| = \eta$, then $B = A$.

7) η is a simple root of the characteristic equation of A.

8) The maximal eigenvalue (or spectral radius) , η, lies between the minimum and maximum of the row sums of A.

9) If A is irreducible, then a weaker condition holds good: $|r| \leq \eta$.

RESULT 1.51.– Suppose that A is an irreducible stable matrix. Let $\theta \geq \max_i |a_{i,i}|$. Then, the matrix $B = I + \frac{1}{\theta}A$ is irreducible and non-negative. Writing $A = \theta(B - I)$, and noting that all eigenvalues of B are less than 1, from result 1.36 we infer that all eigenvalues of A are of the from $\xi = \theta(\lambda - 1)$, where λ is an eigenvalue of B. Thus, all eigenvalues of A have strictly negative real parts. This also implies that the stable matrices are non-singular.

REMARK 1.29.– We can also say that a square matrix, A, is a stable matrix if all its eigenvalues have strictly negative real parts. Similarly, a square matrix, A, is said to be semi-stable if all its eigenvalues have non-positive real parts. These are important observations and will be used in later chapters.

RESULT 1.52.– Suppose that $f(x)$ is a polynomial in x. If ξ is an eigenvalue of a matrix A_m, then $f(\xi)$ is an eigenvalue of $f(A)$.

REMARK 1.30.– The spectral radius (or the maximal eigenvalue) of a non-negative matrix plays an important role in stochastic modeling. We propose (based on our experience) using Elsner's algorithm to compute the spectral radius. Suppose that A is an irreducible non-negative matrix with η as its spectral radius. If A is not irreducible (which occurs commonly in many applications), then one can identify the principal submatrix of A with the spectral radius and then apply Elsner's algorithm to this submatrix. Elsner's algorithm is easy to implement and also converges fast. Necessary steps of the algorithm are given below.

Let $\boldsymbol{u^{(n)}}$ be positive vectors normalized as $\boldsymbol{u^{(n)}e} = 1, n \geq 0$.

Define $S_j^{(n)} = (\boldsymbol{u^{(n)}}A)_j / u_j^{(n)}$, $1 \leq j \leq m$. Find ν_n and μ_n such that:

$$S_{\nu_n}^{(n)} \leq S_j^{(n)} \leq S_{\mu_n}^{(n)},\ 1 \leq j \leq m.$$

Define, for $0 < \alpha < 1$,

$$d_n = \frac{S_{\nu_n}^{(n)} - A_{\nu_n,\nu_n}}{S_{\nu_n}^{(n)} - B_{\nu_n,\nu_n} + \alpha(S_{\mu_n}^{(n)} - S_{\nu_n}^{(n)})}.$$

Next iterate value is obtained as:

$$u_j^{(n+1)} = \begin{cases} \frac{u_{\nu_n}^{(n)}}{[1-(1-d_n)u_{\nu_n}^{(n)}]},\ j \neq \nu_n, \\ \frac{d_n u_{\nu_n}^{(n)}}{[1-(1-d_n)u_{\nu_n}^{(n)}]},\ j = \nu_n, \end{cases}$$

Elsner has proved that for all $n \geq 0$:

$$S_{\nu_n}^{(n)} \leq \zeta \leq S_{\mu_n}^{(n)},$$

$$\lim_{n\to\infty} S_{\mu_n}^{(n)} = \lim_{n\to\infty} S_{\nu_n}^{(n)} = Sp(B), \boldsymbol{u^{(n)}} \to \boldsymbol{u},$$

[Note: $\boldsymbol{u}B = \zeta\boldsymbol{u}$.]

RESULT 1.53.– If a square matrix A_m has all its eigenvalues less than 1 in modulus, then we have:

$$(I - A)^{-n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} A^k, \;\; n \geq 1. \qquad [1.45]$$

1.3.3. *Partitioned matrices*

Partitioned matrices play a vital role in the computational aspects of stochastic models. So, we will give a few key results and additional ones can be generated out of these; as well, one can consult the references mentioned for more details.

RESULT 1.54.– If A_m and B_n are non-singular and if $C_{m\times n}$ and $D_{n\times m}$ are two matrices, then:

$$[A + CBD]^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}D)^{-1}DA^{-1}.$$

RESULT 1.55.– Suppose that A_m is non-singular and that $\boldsymbol{a}$ and $\boldsymbol{b}$ are, respectively, column and row vectors of dimension m. Then, we have for any scalar c,

$$[A + c\,\boldsymbol{a}\,\boldsymbol{b}]^{-1} = A^{-1} - \frac{c}{1 + \boldsymbol{b}A^{-1}\,\boldsymbol{a}} A^{-1}\boldsymbol{a}\,\boldsymbol{b}\,A^{-1}.$$

RESULT 1.56.– Assuming that A^{-1} and B^{-1} exist, we have:

$$[A + C\,B\,C^T]^{-1} = A^{-1} - A^{-1}\,C\,(B^{-1} + C^T\,A^{-1}\,C)^{-1}\,C^T\,A^{-1}.$$

RESULT 1.57.– Assuming the mentioned inverses exist, we have:

$$(I_m + A_{m\times n}B_{n\times m})^{-1} = I_m - A(I_n + B_{n\times m}A_{m\times n})^{-1}B.$$

RESULT 1.58.– If A is non-singular and is partitioned as:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad [1.46]$$

where A_{11} and A_{22} are non-singular matrices of orders, say, m_1 and m_2, then we have:

$$A^{-1} = \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}.$$

RESULT 1.59.– Suppose that the partitioned matrix A (see equation [1.46]) is non-singular but A_{11} and A_{22} are singular. Suppose that $C = (A_{11} + A_{12}A_{21})^{-1}$ and $D = (A_{22} - A_{21}C\,A_{12} - A_{21}C\,A_{12}A_{22})^{-1}$ exist. Then, the inverse of A is given by:

$$A^{-1} = \begin{pmatrix} E_1 & E_2 \\ E_3 & E_4 \end{pmatrix}, \text{ where } E_k = \begin{cases} C + C\,A_{12}(I + A_{22})D\,A_{21}\,C, & k = 1, \\ E_1\,A_{12} - C\,A_{12}(I + A_{22})\,D, & k = 2, \\ -D\,A_{21}C, & k = 3, \\ D - D\,A_{21}C\,A_{12}, & k = 4. \end{cases}$$

RESULT 1.60.– Assuming that A^{-1} and B^{-1} exist, we have:

$$[A + B\,C]^{-1} = A^{-1} - A^{-1}\,B\,(I + C\,A^{-1}\,B)^{-1}\,C\,A^{-1}.$$

RESULT 1.61.– Assuming that A^{-1} and B^{-1} exist, we have:

$$[I + A^{-1}]^{-1} = A(A + I)^{-1},$$

$$(A + BB^T)^{-1}B = A^{-1}\,B(I + B^T\,A^{-1}B)^{-1},$$

$$(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B = B(A + B)^{-1}A,$$

$$A - A(A + B)^{-1}A = B - B(A + B)^{-1}B,$$

$$(A^{-1} + B^{-1}) = A^{-1}(A + B)B^{-1},$$

$$(I + A\,B)^{-1} = I - A(I + BA)^{-1}B,$$

$$(I + A\,B)^{-1}A = A(I + AB)^{-1}.$$

RESULT 1.62.– Assuming that A^{-1} and B^{-1} exist, we have:

$$(A + B)^{-1} = A^{-1} + B^{-1} \;\Rightarrow\; A\,B^{-1}\,A = B\,A^{-1}\,B.$$

1.3.4. *Matrix differentiation*

Differentiation of vectors and matrices is needed in deriving expressions and proving results in stochastic modeling. Thus, in this section, we review the needed definitions, concepts and results. For more details, we refer the reader to the references mentioned in section 1.3.

DEFINITION 1.49.– *Suppose that* $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x}))$ *is a vector-valued differentiable function defined on* $\Re^n$ *to* $\Re^n$*. The partial derivative,* $\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}$*, is defined as:*

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

DEFINITION 1.50.– *Suppose that* $f : \Re^n \rightarrow \Re$ *is a real-valued differentiable function. Then, the partial derivative,* $\dfrac{\partial f}{\partial \boldsymbol{x}}$*, is defined as* $\dfrac{\partial f}{\partial \boldsymbol{x}} = \left(\dfrac{\partial f}{\partial x_1}, \cdots, \dfrac{\partial f}{\partial x_n}\right)$.

REMARK 1.31.– From the definition, we see that $\dfrac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}} = I_n$.

RESULT 1.63.– If the matrix A_n is independent of $\boldsymbol{x}$, then we have $\dfrac{\partial}{\partial \boldsymbol{x}} A\,\boldsymbol{x} = A$.

RESULT 1.64.– If the matrix $A_{m\times n}$ is independent of both $\boldsymbol{z}$, a column vector of dimension m, and $\boldsymbol{x}$, a column vector of dimension n, then we have:

$$\frac{\partial}{\partial \boldsymbol{z}}(\boldsymbol{z}^T A \boldsymbol{x}) = \boldsymbol{x}^T A^T \text{ and } \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{z}^T A \boldsymbol{x}) = \boldsymbol{z}^T A. \qquad [1.47]$$

RESULT 1.65.– If the matrix A_n is independent of $\boldsymbol{x}$, which is a column vector of dimension n, then we have $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T A \boldsymbol{x}) = \boldsymbol{x}^T (A + A^T)$.

RESULT 1.66.– Suppose that the matrix $A_n(x)$ is a function of x and that $A_n^{-1}(x)$ exists. Then, we have:

$$\frac{d}{dx} A_n^{-1}(x) = -A_n^{-1}(x) \frac{d}{dx} A_n(x)\, A_n^{-1}(x). \qquad [1.48]$$

RESULT 1.67.– Suppose that the matrix A and B are two matrices for which addition makes sense. Then, we have $\dfrac{d}{dx}[A + xB] = B$.

RESULT 1.68.– Suppose that the matrix operations shown below are valid. Then, we have the following:

$$\frac{d}{dx}A^T(B+xC)^{-1}D = -A^T(B+xC)^{-1}C\,(B+xC)^{-1}D. \quad [1.49]$$

RESULT 1.69.– Suppose that $f(A)$ and $g(A)$ are functions of A. Assume that the products shown below are well defined. Then, we have:

$$\frac{d}{dA}\{[f(A)]^T g(A)\} = \frac{d}{dA}f(A)g(A) + \frac{d}{dA}g(A)f(A). \quad [1.50]$$

RESULT 1.70.– Suppose that A and B are square matrices of dimension m. Suppose that $f(z) = (A+zB)^n$, where z is a scalar and n is a non-negative integer. Then, we have:

$$\frac{df}{dz} = \sum_{k=0}^{n-1}(A+zB)^k\,B\,(A+zB)^{n-1-k}. \quad [1.51]$$

1.3.5. *Exponential matrix*

Exponential matrices arise naturally in stochastic modeling. These matrices need to be computed at times and knowing the properties of an exponential matrix will greatly benefit in coding as well as to prove analytical results. We review a few aspects related to an exponential matrix.

DEFINITION 1.51.– *Suppose that A is a finite-dimensional (square) matrix of dimension m. The exponential of A is defined as the Picard series:*

$$e^A = \sum_{k=0}^{\infty}\frac{A^k}{k!}, \quad [1.52]$$

where $A^0 = I_m$.

REMARK 1.32.– Note that by definition e^A is always a non-negative matrix. If A is irreducible, then e^A will be a positive matrix.

RESULT 1.71.– Suppose that a is a scalar and that A_m is a square matrix. Then, we have:

$$e^{aI_m+A} = e^a\,e^A. \quad [1.53]$$

RESULT 1.72.– If A_m and B_m commute, then:

$$e^{(A+B)} = e^A\,e^B = e^B\,e^A. \quad [1.54]$$

RESULT 1.73.– Suppose that A^{-1} exists. Then, we have:

$$\int_0^t f(x)\, e^{A\,x}\, dx = f(t)\, e^{A\,t}\, A^{-1} - f(0)A^{-1} - A^{-1} \int_0^t e^{A\,x}\, f'(x)\, dx, \qquad [1.55]$$

where $f'(x)$ denotes the first derivative of $f(x)$.

RESULT 1.74.– Suppose that A^{-1} exists. Then, we have:

$$\int_0^\infty e^{-s\,x}\, e^{A\,x}\, dx = (sI - A)^{-1}, \;\; Re(s) \geq 0. \qquad [1.56]$$

RESULT 1.75.– Suppose that A_m is a square matrix with an eigenvalue ξ. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ denote, respectively, the left and right eigenvectors of A corresponding to ξ. Then, e^{ξ} is an eigenvalue of e^A, and $\boldsymbol{u}$ and $\boldsymbol{v}$ are, respectively, the left and right eigenvectors of e^A corresponding to e^{ξ}.

PROOF.– Follows immediately on noting that:

$$\boldsymbol{u}A = \xi\boldsymbol{u} \;\; \Rightarrow \;\; \boldsymbol{u}A^n = \xi^n\boldsymbol{u}, \;\; n \geq 1, \qquad [1.57]$$

which immediately implies that:

$$\boldsymbol{u}e^A = e^{\xi}\boldsymbol{u}. \qquad [1.58]$$

Similarly, one can show that $e^A\, \boldsymbol{v} = e^{\xi}\boldsymbol{v}$. □

RESULT 1.76.– Differentiation of an exponential matrix yields the following:

$$\frac{d}{dx}e^{Ax} = Ae^{Ax} = e^{Ax}A.$$

RESULT 1.77.– Suppose that A is non-singular. Then, we have:

$$\int e^{At}dt = A^{-1}e^{At} = e^{At}A^{-1}.$$

RESULT 1.78.– The inverse of e^A always exists and is given by e^{-A}.

RESULT 1.79.– Suppose that $Q = (q_{i,j})$ is a semi-stable matrix (recall: $q_{i,i} < 0,\; q_{i,j} \geq 0$, and $\sum_j q_{i,j} = 0$, for all i). Also, let $\boldsymbol{\pi}$ denote the left eigenvector corresponding to the maximal eigenvalue, namely, 0. Then, we have:

$$\int_0^t e^{Q\,x}\, dx = t\,\boldsymbol{e}\,\boldsymbol{\pi} + [I - e^{Q\,t}][\boldsymbol{e}\,\boldsymbol{\pi} - Q]^{-1}, \; t \geq 0.$$

1.3.6. *Kronecker products and Kronecker sums*

In stochastic modeling, Kronecker products and Kronecker sums play an important role and simplify a number of steps in proofs. Thus, in this section, we review the basic ones needed for the book. It is highly recommended that readers go through the books referenced earlier.

DEFINITION 1.52.– *Suppose that $A = (a_{ij})$ is an $m \times n$ matrix and that $B = (b_{ij})$ is a $p \times q$ matrix. The Kronecker product of A and B, denoted by $A \otimes B$, is a matrix of dimension $mp \times nq$ and is given by:*

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}B & a_{n2}B & \cdots & a_{mn}B \end{pmatrix}.$$

RESULT 1.80.– Below, we assume that the matrix operations such as multiplications and additions are meaningful.

$$(A + B) \otimes C = (A \otimes C) + (B \otimes C),$$

$$A \otimes (B + C) = (A \otimes B) + (A \otimes C),$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C).$$

$$(A \otimes B)^T = (A^T \otimes B^T),$$

$$(A \otimes B)(C \otimes D) = (AC \otimes BD).$$

RESULT 1.81.– If A and B are non-singular matrices, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

REMARK 1.33.– In general, Kronecker product operation is not commutative. That is, in general $A \otimes B \neq B \otimes A$.

RESULT 1.82.– Suppose that A has eigenvalues $\xi_i,\ i = 1, 2, \cdots, m$, with the corresponding eigenvectors $\boldsymbol{u}_i$; B has eigenvalues $\gamma_j,\ j = 1, 2, \cdots, n$, with the corresponding eigenvectors $\boldsymbol{v}_j$. Then, the eigenvalues of $A \otimes B$ are given by $\xi_i \gamma_j,\ 1 \leq i \leq m,\ 1 \leq j \leq n$, with the corresponding eigenvectors given by $\boldsymbol{u}_i \otimes \boldsymbol{v}_j$.

DEFINITION 1.53.– *Suppose that A and B are two square matrices of dimension m and n, respectively. The Kronecker sum , denoted by, $A \oplus B$, is defined as:*

$$A \oplus B = A \otimes I_n + I_m \otimes B.$$

RESULT 1.83.– Suppose that A has eigenvalues $\xi_i, i = 1, 2, \cdots, m$, with the corresponding eigenvectors $\boldsymbol{u}_i$; B has eigenvalues $\gamma_j, j = 1, 2, \cdots, n$, with the corresponding eigenvectors $\boldsymbol{v}_j$. Then, the eigenvalues of $A \oplus B$ are given by $\xi_i + \gamma_j,\ 1 \leq i \leq m,\ 1 \leq j \leq n$, with the corresponding eigenvectors given by $\boldsymbol{u}_i \otimes \boldsymbol{v}_j$.

RESULT 1.84.– If A and B are two square matrices of dimension m and n, respectively, then:

$$e^{A \oplus B} = e^A \otimes e^B.$$

It is worth pointing out that the above equation is obtained by noticing the following:

$$(A \otimes I)^k = A^k \otimes I,\ k \geq 0,$$

$$e^{(A \otimes I)} = e^A \otimes I,$$

$$A \otimes B = (A \otimes I)(I \otimes B),$$

$$e^{A \otimes I + I \otimes B} = e^{A \otimes I} e^{I \otimes B}.$$

The last equality holds good on noting that $A \otimes I$ and $I \otimes B$ commute. That is, $(A \otimes I)(I \otimes B) = (I \otimes B)(A \otimes I)$.

RESULT 1.85.– If A and B are two square matrices of dimension m and n, respectively, then:

$$(A \otimes B)^r = A^r \otimes B^r,\ r \geq 0.$$

RESULT 1.86.– If A and B are stable matrices, then $A \oplus B$ is non-singular. The following result plays a key role in stochastic modeling.

$$\int_0^\infty e^{At} \otimes e^{Bt} dt = -(A \oplus B)^{-1}.$$

1.3.7. *Vectorization (or direct sums) of matrices*

Vectorization or simply the direct sums of matrices is a convenient tool in performing analysis of stochastic models among many other areas.

DEFINITION 1.54.– *Suppose that $A_{m \times n} = (a_{i,j})$ is given. Then, the direct (row) sum of A, denoted as $\tau(A)$, is a row vector of dimension mn by taking the rows of A and appending each row to the previous one. That is:*

$$\boldsymbol{\tau}(A) = (a_{1,1}, \cdots, a_{1,n}, a_{2,1}, \cdots, a_{2,n}, \cdots, a_{m,1}, \cdots, a_{m,n}).$$

DEFINITION 1.55.– *Suppose that $A_{m\times n} = (a_{i,j})$ is given. Then, the direct (column) sum of A, denoted as $\tau^c(A)$, is a column vector of dimension mn by taking the columns of A and appending each column to the previous one. That is:*

$$\boldsymbol{\tau}^c(A) = (a_{1,1}, \cdots, a_{m,1}, a_{1,2}, \cdots, a_{m,2}, \cdots, a_{1,n}, \cdots, a_{m,n})^T.$$

RESULT 1.87.– Suppose that both A and B are of dimension $m \times n$. Then, we have $\boldsymbol{\tau}(A+B) = \boldsymbol{\tau}(A) + \boldsymbol{\tau}(B)$ and $\boldsymbol{\tau}^c(A+B) = \boldsymbol{\tau}^c(A) + \boldsymbol{\tau}^c(B)$.

RESULT 1.88.– Suppose that both A and B are of dimension $m \times n$, and both C and D are of dimension $n \times r$. Then, we have:

$$\begin{aligned}\boldsymbol{\tau}((A+B)(C+D)) &= (\boldsymbol{\tau}(A) + \boldsymbol{\tau}(B))(I_m \otimes C)(I_m \otimes D)\\ &= (\boldsymbol{\tau}(C) + \boldsymbol{\tau}(D))(A^T \otimes I_r)(B^T \otimes I_r),\end{aligned}$$

$$\begin{aligned}\boldsymbol{\tau}^c((A+B)(C+D)) &= ((I_r \otimes A) + (I_r \otimes B))(\boldsymbol{\tau}^c(C) + \boldsymbol{\tau}^c(D))\\ &= ((C^T \otimes I_m) + (D^T \otimes I_m))(\boldsymbol{\tau}^c(A) + \boldsymbol{\tau}^c(B)).\end{aligned}$$

RESULT 1.89.– Suppose $A_{m\times n}$ and $B_{n\times r}$ are two rectangular matrices. Then:

$$\boldsymbol{\tau}(AB) = \boldsymbol{\tau}(B)(A^T \otimes I_r) = \boldsymbol{\tau}(A)(I_m \otimes B),$$

$$\boldsymbol{\tau}^c(AB) = (I_r \otimes A)\boldsymbol{\tau}^c(B) = (B \otimes I_m)\boldsymbol{\tau}^c(A).$$

RESULT 1.90.– Suppose $A_{m_1\times m_2}, B_{m_2\times m_3}$, and $C_{m_3\times m_4}$ are three rectangular matrices. Then:

$$\boldsymbol{\tau}(ABC) = \boldsymbol{\tau}(C)(B^T A^T \otimes I_{m_4}) = \boldsymbol{\tau}(B)(A^T \otimes C) = \boldsymbol{\tau}(A)(I_{m_1} \otimes BC),$$

$$\boldsymbol{\tau}^c(ABC) = (I_{m_1} \otimes AB)\boldsymbol{\tau}^c(C) = (C^T \otimes A)\boldsymbol{\tau}^c(B) = (C^T B^T \otimes I_{m_1})\boldsymbol{\tau}^c(A).$$