

Chapter 2

Arrival and Service Processes

2.1 Introduction

For every queuing system we have to specify both the arrival and the service processes very clearly. The stochastic process that characterize arrival and service are key to explaining how queues build, fluctuate and dissipate. They are some of the key characteristics that determine the performance of a queueing system. If the times between arrivals of items are generally short and the service times are long it is clear that the queue will build up. Most important is how those times vary, i.e. the distribution of the inter-arrival times and service times. We will present the arrival and service processes that commonly occur in most queuing systems. However before we do that let us briefly review some important probability axioms and other aspects associated with probability of discrete event since a matrix approach of representing probability distributions will be used in this chapter a brief review of

2.2 Review of Probability for Discrete Random Variables and Matrices

We define a discrete state space $\mathcal{Z} = \{a_0, a_1, a_2, a_3, a_4, \dots\}$ as a countable sequence, where $a_i \neq a_j$, $i \neq j$. Throughout this book we will consider special discrete state space where $\mathcal{Z} = \{0, 1, 2, 3, 4, \dots\}$. Consider any random collection of discrete variables we call the set $\mathcal{Z}_s = \{a_1, a_2, \dots, a_N\}$. Let $\mathcal{Z}_s \subseteq \mathcal{Z}$. As an example \mathcal{Z}_s could be the possible outcome of tossing a six-sided die, i.e. $\mathcal{Z}_s = \{1, 2, 3, 4, 5, 6\}$, or the number of heads showing up when three unbiased coins are tossed in which case $\mathcal{Z}_s = \{0, 1, 2, 3\}$, etc. Let \mathcal{A} be a random variable that is in the set \mathcal{Z} and $P(\mathcal{A})$ be the probability of the event \mathcal{A} . We define p_i as the probability that \mathcal{A} assumes the value $i \in \mathcal{Z}$, then we write

1. $P(\mathcal{A} = i) = p_i, i \in \mathcal{Z}$,
2. $0 \leq p_i \leq 1$,
3. $\sum_{i \in \mathcal{Z}} p_i = 1$.

The expectation of a random variable can be written as

1. $E[\mathcal{A}]$ as the expected value of \mathcal{A} .
2. For the example above we can write $E[\mathcal{A}] = \sum_{i \in \mathcal{Z}} i p_i = \sum_{i > 0} \sum_{k=i}^{\infty} p_k$, and
3. in general we have $E[g(\mathcal{A})] = \sum_{i \in \mathcal{Z}} g(i) p_i$.

2.2.1 The z transform

For $|z| \leq 1$, we define the z -transform or the probability generating function (*pgf*) of \mathcal{A} as

$$A(z) = \sum_{i \in \mathcal{Z}} z^i p_i.$$

We have

$$E[\mathcal{A}] = \frac{dA(z)}{dz} \Big|_{z \rightarrow 1} = \sum_{i \geq 1} i p_i,$$

and in general we have the n^{th} factorial moment of \mathcal{A} written as

$$E_f^n[\mathcal{A}] = \frac{d^n A(z)}{dz^n} \Big|_{z \rightarrow 1} = \sum_{i \geq n} \frac{i!}{(i-n)!} p_i.$$

2.2.2 Bivariate Cases

Consider two random variables \mathcal{A} and \mathcal{B} , with $\mathcal{A} \in \mathcal{Z}$ and $\mathcal{B} \in \mathcal{Z}$, and let $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$ represent the union and intersection, respectively, of \mathcal{A} and \mathcal{B} , we have

1. $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$, where $P(\mathcal{A} \cap \mathcal{B})$ is the probability of both event \mathcal{A} and \mathcal{B} occurring.
2. $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A})$, where $P(\mathcal{B}|\mathcal{A})$ is the conditional probability that \mathcal{B} occurs given that \mathcal{A} has occurred.
3. $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$, if \mathcal{A} and \mathcal{B} are independent
4. $P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})}$, $P(\mathcal{A}) > 0$.
5. Let us write $p_{i,j} = \Pr\{\mathcal{A} = i, \mathcal{B} = j\}$ then we have $P(z_1, z_2) = \sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{Z}} z_1^i z_2^j p_{i,j}$, $|z_k| \leq 1$, $k = 1, 2$.

All these results can be written for multivariate cases of 3 or more random variables. For example, consider n random variables $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ and let $p_{i_1, i_2, \dots, i_n} = \Pr\{\mathcal{A}_1 = i_1, \mathcal{A}_2 = i_2, \dots, \mathcal{A}_n = i_n\}$, then we can write

$$P(z_1, z_2, \dots, z_n) = \sum_{i_1 \in \mathcal{Z}} \sum_{i_2 \in \mathcal{Z}} \cdots \sum_{i_n \in \mathcal{Z}} z_1^{i_1} z_2^{i_2} \cdots z_n^{i_n} p_{i_1, i_2, \dots, i_n}, \quad |z_k| \leq 1, \forall k \geq 1.$$

2.2.3 Some very Common Discrete Distributions

2.2.3.1 Bernoulli Process

The Bernoulli process is one of the most important processes in probability and especially when carrying out discrete time analysis of queues. Consider a random variable or an event \mathcal{A} that has only two possible outcomes, success (S) or failure (F). Let $P(\mathcal{A} = S) = p$ and $P(\mathcal{A} = F) = q = 1 - p$, then this process is called a Bernoulli process, if each trial is independent of the previous one and the outcomes have the same probability.

The z -transform or probability generating function (pgf) associated with Bernoulli distribution $B(z)$ is stated as

$$B(z) = q + pz, \quad |z| \leq 1.$$

As an example, consider taking well shuffled deck of playing cards (52 of them). If we draw a card from the deck at random and we consider drawing a King as success then

$$P(\mathcal{A} = S) = p = \frac{4}{52} = \frac{1}{13},$$

and of course

$$P(\mathcal{A} = F) = \frac{48}{52} = \frac{12}{13} = 1 - \frac{1}{13}.$$

If we replace the drawn card into the deck, reshuffle the deck and draw again at random, the probability p remains the same. If we take this example further and say customers arrive at every time interval according to a Bernoulli process with probability p , then we can determine, probabilistically how the number of customers grow in the system at any time.

2.2.3.2 Geometric Distribution

Geometric distribution is derived from the Bernoulli process. It captures how many repetitive tries it takes for a success to occur. For example, let us consider the example case of drawing a card from the deck as given above. Suppose we draw

from the deck and replace the drawn card each time back to the deck and we want to know the probability of how long it takes to draw the first King (first success). If we were successful the first time then the probability that it took us one draw to be successful is p . The probability that it took us two draws to be successful will imply that the first draw was unsuccessful while the second was successful. The probability that it took two draws for the first success is qp . Let the \mathcal{X} be the number of attempts for first success, then it follows that

$$P(\mathcal{X} = i) = q^{i-1}p, i \geq 1.$$

The pgf associated with the geometric distribution, $G(z)$, is written as

$$G(z) = \sum_{i=1}^{\infty} q^{i-1}p z^i = \frac{pz}{1-qz}, |z| \leq 1.$$

In some other disciplines or books people study, instead, the number of failures before a success. In that case we speak of a number $\tilde{\mathcal{X}} = \mathcal{X} - 1$. Hence we have

$$P(\tilde{\mathcal{X}} = i) = q^i p, i \geq 0.$$

In this case the pgf, $\widetilde{G}(z)$ is written as

$$\widetilde{G}(z) = \sum_{i=0}^{\infty} q^i p z^i = \frac{p}{1-qz}, |z| \leq 1.$$

We will be using the first definition in this book, because our interest is in queueing systems for which both inter-arrival and service times are always at least one time unit.

2.2.3.3 Binomial Distribution

If on the other hand the question of interest to us is given that we are allowed to draw \mathcal{Y} times from the deck (with replacement), how many successes do we expect to have? Let \mathcal{X} be the number of draws and \mathcal{Y} the number of successes out of the \mathcal{X} draws ($\mathcal{Y} \leq \mathcal{X}$) then

$$P(\mathcal{Y} = j | \mathcal{X} = n) = \binom{n}{j} p^j q^{n-j}, 0 \leq j \leq n.$$

The pgf associated with the Binomial distribution, $\hat{B}(z)$, is written as

$$\hat{B}(z) = \sum_{j=0}^n \binom{n}{j} p^j q^{n-j} z^j = (pz + q)^n = (B(z))^n, |z| \leq 1.$$

The Poisson distribution can be derived as a limiting distribution of the Binomial under certain conditions.

2.2.3.4 Negative Binomial Distribution

Another distribution that is very related to the Bernoulli process is the negative binomial distribution. Here the interest is in knowing how many draws are needed to have exactly a particular number of successes. Let \mathcal{Z} be the number of draws required for the \mathcal{Y}^{th} success to occur. We have

$$P(\mathcal{Y} = j | \mathcal{Z} = n,) = \binom{n-1}{j-1} p^j q^{n-j}, \quad 1 \leq j \leq n.$$

It is clear that when $j = 1$ we have that the Negative Binomial is a special case of the Geometric distribution. The *pgf* associated with the Negative Binomial distribution, $NB(z)$, is written as

$$NB(z) = \sum_{j=1}^{n-1} \binom{n-1}{j-1} p^j q^{n-j} z^j = pz(pz + q)^{n-1} = pz(B(z))^{n-1}, \quad |z| \leq 1.$$

This is the discrete analog of the Erlang distribution.

2.2.3.5 Mixture of Geometric Distributions

Finally we present another distribution that is also related to the Bernoulli process and Geometric distribution – a mixture of geometric distributions. Consider n geometric distributions with parameters (p_k, q_k) , $k = 1, 2, \dots, n$. Suppose we are given a probability distribution $(\theta_1, \theta_2, \dots, \theta_n)$ with $\sum_{i=1}^n \theta_i = 1$. There is a distribution called a mixture of geometric distributions. If the number of draws for first success as \mathcal{X} is of the mixture of geometric distribution, then

$$P(\mathcal{X} = i) = \sum_{k=1}^n \theta_k q_k^{i-1} p_k, \quad k \geq 1.$$

This is the discrete equivalent of the hyper-exponential distribution.

For detailed exposition to these distributions, the reader is referred to basic probability books. Most of these distributions will be covered in detail later with regards to queueing theory. Specifically how they arise in discrete time queueing models will be further discussed later.

2.2.4 Brief Summary of Required Material from Matrices

In this section we briefly present a few properties of a square matrix that will be needed in this chapter. Generally a matrix D of dimension $m \times n$ can be written as

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}. \text{ Consider}$$

1. a non-negative finite square matrix A of dimension n , i.e. each element $a_{i,j}$ of this matrix has the property that $0 \leq a_{i,j} < \infty$,
2. an n row vector $\mathbf{c} = [c_1, c_2, \cdots, c_n]$, and
3. an n column vector $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$.
4. A column vector $\mathbf{1} = [1 \ 1 \ \cdots \ 1]^T$, where A^T is the transpose of a matrix A
5. $\mathbf{e}_j(n)$ which is an n column vector of zeros in all locations except at location j where there is a 1.
6. An identity matrix I . We will usually assume the dimension of I is obvious and when it is not we write it as $I(n)$, i.e. of dimension n .

We have the following properties of matrices that will be used in this book

1.

$$A^2 = \begin{bmatrix} \sum_{j=1}^n a_{1,j}a_{j,1} & \sum_{j=1}^n a_{1,j}a_{j,2} & \cdots & \sum_{j=1}^n a_{1,j}a_{j,n} \\ \sum_{j=1}^n a_{2,j}a_{j,1} & \sum_{j=1}^n a_{2,j}a_{j,2} & \cdots & \sum_{j=1}^n a_{2,j}a_{j,n} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{j=1}^n a_{n,j}a_{j,1} & \sum_{j=1}^n a_{n,j}a_{j,2} & \cdots & \sum_{j=1}^n a_{n,j}a_{j,n} \end{bmatrix},$$

2. From the above we can infer A^k , $k = 1, 2, 3, \cdots$,
3. By definition $A^0 = I$, the identity matrix, where $I_{i,j}$ are the elements of I and $I_{j,j} = 1$ with $I_{j,k} = 0$, for $j \neq k$.
- 4.

$$\mathbf{cA} = \left[\sum_{i=1}^n c_i a_{i,1}, \sum_{i=1}^n c_i a_{i,2}, \cdots \sum_{i=1}^n c_i a_{i,n} \right],$$

5.

$$\mathbf{Ab} = \begin{bmatrix} \sum_{j=1}^n a_{1,j}b_j \\ \sum_{j=1}^n a_{2,j}b_j \\ \vdots \\ \sum_{j=1}^n a_{n,j}b_j \end{bmatrix},$$

6. The matrix A has n eigenvalues λ_i , $i = 1, 2, \dots, n$ and associated left eigenvectors \mathbf{a}_i such that

$$\lambda_i \mathbf{a}_i = \mathbf{a}_i A.$$

7. The spectral radius of A denoted by $sp(A)$ is given as

$$sp(A) = \max\{|\lambda_i|, i = 1, 2, \dots, n\}.$$

Next we give some simple but very interesting examples of matrix representations of discrete distributions.

2.2.5 Examples of Simple Representations of Discrete Distributions Using Matrices

We will consider three simple examples of representations.

1. Consider a variable θ , $0 \leq \theta \leq 1$, and an integer $n < \infty$. Let $b(n, \theta, i) =$

$$\binom{n}{i} \theta^i (1 - \theta)^{n-i}, 0 \leq i \leq n. \text{ Define}$$

- an n row vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_n]$, where $\alpha_i = b(n, \theta, i)$,

- a column vector $\mathbf{t} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and

- a matrix $T = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$.

If we define a random variable $\mathcal{A} = \{0, 1, 2, \dots, n\}$, for which we have

$$Pr\{\mathcal{A} = 0\} = 1 - \boldsymbol{\alpha} \mathbf{1},$$

$$Pr\{\mathcal{A} = i\} = \boldsymbol{\alpha} T^{i-1} \mathbf{t}, i = 1, 2, \dots, n.$$

It is immediately obvious that what we have done is written the Binomial distribution in a matrix form (be it inefficiently). As an example,

$$Pr\{\mathcal{A} = j\} = \binom{n}{j} (1 - \theta)^j \theta^{n-j}.$$

2. Consider an n row vector $\alpha = [1, 0, 0, \dots, 0]$, an n column vector $\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$

and an n square matrix $T = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$. If we define a random variable $\mathcal{A} = \{0, 1, 2, \dots, n\}$, for which we have

$$Pr\{\mathcal{A} = i\} = \alpha T^{i-1} \mathbf{t}, \quad i = 1, 2, \dots, n,$$

then one sees that what we have is

$$Pr\{A = n\} = 1, \text{ and } Pr\{\mathcal{A} = j\} = 0, j \neq n.$$

This is a constant random variable.

3. Consider

- a variable θ ,
- an n row vector $\alpha = [1, 0, 0, \dots, 0]$,

- an n column vector $\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \theta \end{bmatrix}$ and

- an n square matrix $T = \begin{bmatrix} 1-\theta & \theta & 0 & \dots & 0 \\ 0 & 1-\theta & \theta & \dots & 0 \\ 0 & 0 & 1-\theta & \theta & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1-\theta \end{bmatrix}$.

If we define a random variable $A = \{0, 1, 2, \dots, n\}$, for which we have

$$Pr\{\mathcal{A} = i\} = \alpha T^{i-1} \mathbf{t}, \quad i = 1, 2, \dots, n.$$

One sees that this random variable has the negative binomial distribution. For example,

$$Pr\{\mathcal{A} = j\} = \binom{n-1}{j-1} (1-\theta)^{j-1} \theta^{n-j}.$$

These are just simple examples of representation of some well known discrete distributions using the matrix approach. The key here is that we have demonstrated, through examples, that at least some well-known discrete distributions can be represented using matrix form. This approach becomes very useful later when we discuss some very important and basic queueing models.

2.2.6 Matrix Representation

Consider a square matrix T of dimension $1 < m < \infty$ with the following properties.

- its elements T_{ij} satisfy the condition: $0 \leq T_{ij} < \infty$
- there is constant κ such that the spectral radius $\rho = sp(\kappa T) < 1$. Usually $0 < \kappa \leq 1$.

Theorem 2.1 Then there exists a row vector \mathbf{a} , with elements $a_i > 0$ and $\mathbf{a}\mathbf{1} \leq 1$, and a column vector \mathbf{w} , with elements $w_i > 0$, such that

$$0 \leq \mathbf{a}T^{k-1}\mathbf{w} \leq 1, \quad \forall k = 1, 2, \dots$$

and

$$0 \leq \mathbf{a}(I - T)^{-1}\mathbf{w} \leq 1.$$

Proof: If $0 \leq T_{ij} < \infty$ and $sp(T) < 1$, then it is known from matrix analysis that $0 \leq (T^n)_{ij} < \infty$, $\forall k \geq 0$, hence it is possible to find two vectors \mathbf{a} and \mathbf{w} such that

$$0 \leq \mathbf{a}T^{k-1}\mathbf{w} \leq 1, \text{ and } \sum_{k=1}^{\infty} \mathbf{a}T^{k-1}\mathbf{w} \leq 1..$$

Definition: Let \mathcal{A} be a discrete random variable, and p_k be defined as $p_k = Pr\{\mathcal{A} = k\}$, then

$$p_k = \mathbf{a}T^{k-1}\mathbf{w}, \quad k = 1, 2, 3, \dots$$

$$p_0 = 1 - \mathbf{a}(I - T)^{-1}\mathbf{w}.$$

These matrix representations will be used later in Section 2.5.

2.3 Arrival and Service Processes

These two processes can probably be discussed separately. But because the distributions used for service processes can also be used for the inter-arrival times, we will combine them in the same section.

Consider a situation where packets arrive to a router at discrete time points. Let J_n be the arrival time of the n th packet. Further let A_n be the inter-arrival time between the n th and $(n-1)$ th packet, i.e. $A_n = J_n - J_{n-1}$, $n = 1, 2, \dots$. We let J_0 be the start of our system and assume an imaginary zeroth packet so that $A_1 = J_1 - J_0$. If A_1, A_2, A_3, \dots are independent random variables then the time points J_1, J_2, J_3, \dots are said to form renewal points. If also $A_n = A$, $\forall n$, then we say the inter-arrival times have identical distributions. We present the distributions commonly assumed by A . We will not be concerned at this point about A_1 since it may have a special behaviour. We assume it is a well behaved interval and not a residual. Dealing with the residuals at this point would distract from the main discussion in this chapter.

Let t_n^0 be the start time of the service time of the n th packet and let t_n^f be the completion time of its service time duration, i.e. $S_n = t_n^f - t_n^0$. If $S_n = S$, $\forall n$, then we say that the service times have identical distributions. We can also consider a system that is saturated with a queue of packets needing service. Let C_n be the service completion time of the n th packet. Then $S_n = C_n - C_{n-1}$.

First let us briefly introduce the renewal process, which is related to some of these arrival and service processes which are independent.

2.4 Renewal Process

Let X_n be the inter-event time between the $(n-1)$ th and n th events such as arrivals. In our context X_n could be A_n or S_n . If $Y_0 = 0$ represents the start of our observing the system, then $Y_n = \sum_{i=0}^n X_i$ is the time of occurrence of the n th event. For example, $J_n = \sum_{i=0}^n A_i$ is the arrival time of the n th customer. We shall assume for simplicity that $X_0 = 0$ and X_1, X_2, \dots are independent. Later we may also assume that they are identically distributed variables (note that it is more general to assume that X_0 has a value and it has a different distribution than the rest, but that is not very critical for our purpose here). If they are independent and identically distributed we say they are *iid*.

If we consider the duration of X_n as how long the n th inter-event lasts, then Y_n is the point in time when the n th process starts and then lasts for X_n time units. The time Y_n is the regeneration point or the renewal epoch. We assume that $P\{X_n > 0\} > 0$, $\forall n > 1$, and we let \mathbf{p}_n be the collection $\mathbf{p}_n = (p_1^{(n)}, p_2^{(n)}, \dots)$, where $p_i^{(n)} = \Pr\{X_n = i\}$, $\forall n \geq 1$. We define the z -transform of this inter-event times as

$$p_n^*(z) = \sum_{j=1}^{\infty} p_j^{(n)} z^j, \quad |z| \leq 1, \quad \forall n. \quad (2.1)$$

Let the mean number of events per unit time be given by μ then the mean interevent time

$$E[X_n] = \left. \frac{dp_n^*(z)}{dz} \right|_{z \rightarrow 1} = \mu^{-1} > 0, \quad \forall n. \quad (2.2)$$

Since X_1, X_2, \dots are independent we can obtain the distribution of Y_n as the convolution sum of X_1, X_2, \dots as follows

$$\Pr\{Y_n\} = \Pr\{X_1 * X_2 * \dots * X_n\}. \quad (2.3)$$

Letting $q_j^{(n)} = \Pr\{Y_n = j\}$ and $q_n^*(z) = \sum_{j=1}^{\infty} q_j^{(n)} z^j$, $|z| \leq 1$ we have

$$q_n^*(z) = p_1^*(z) p_2^*(z) \cdots p_n^*(z). \quad (2.4)$$

2.4.1 Example:

$$\Pr\{Y_2 = j\} = \sum_{v=1}^{j-1} \Pr\{X_1 = v\} \Pr\{X_2 = j - v\}.$$

If $X_1 = X_2 = X$, then $p_1^*(z) = p_2^*(z) = p^*(z)$ and hence we have

$$q_2^*(z) = (p^*(z))^2.$$

Further let

$$p_1 = 0.1, p_2 = 0.3, p_3 = 0.6, p_j = 0, j \geq 4,$$

then

$$p^*(z) = 0.1z + 0.3z^2 + 0.6z^3$$

and

$$q_2^*(z) = 0.01z^2 + 0.06z^3 + 0.21z^4 + 0.36z^5 + 0.36z^6.$$

From this we obtain

$$q_1^{(2)} = 0, q_2^{(2)} = 0.01, q_3^{(2)} = 0.06, q_4^{(2)} = 0.21, q_5^{(2)} = 0.36, q_6^{(2)} = 0.36, q_j^{(2)} = 0, \forall j \geq 7.$$

In this case, for example, X could be the length of time it takes to process a job. If we have two jobs in a system the total time it takes to process the two is given by Y_2 . It is easy to show that if we have K jobs then the z -transform of the total time to process all of them will be $(p(z))^K$.

2.4.2 Number of renewals

Let Z_n be the number of renewals in $[0, n]$ then

$$Z_n = \max\{m \geq 0 \mid Y_m \leq n\}. \quad (2.5)$$

Hence,

$$P\{Z_n \geq j\} = P\{Y_j \leq n\} = \sum_{v=1}^n P\{Y_j = v\}. \quad (2.6)$$

If we let $m_n = E[Z_n]$ be the expected number of renewals in $[0, n]$, it is straightforward to show that

$$m_n = \sum_{j=1}^n \sum_{v=j}^n P\{Y_j = v\} \quad (2.7)$$

m_n is called the *renewal function*. It is a well known elementary renewal theorem that the mean renewal rate μ is given by

$$\mu = \lim_{n \rightarrow \infty} \frac{m_n}{n}. \quad (2.8)$$

For a detailed treatment of renewal process, the reader is referred to Wolff (1989). Here we only give a skeletal proof.

In what follows, we present distributions that are commonly used to describe arrival and service processes. Some of them are of the renewal types.

2.5 Special Arrival and Service Processes in Discrete Time

2.5.1 Bernoulli Process

The Bernoulli process in general terms was presented in a previous section of the chapter. Let us consider it in the context of a queueing system. If time is slotted, i.e. if we consider discrete time of equal intervals, and assume the arrival of packets at each interval could only be singly with probability $p > 0$ and we have $q = 1 - p$ and are equally likely at each time. Then at any interval between time t_n and t_{n+1} the probability of one packet arrival is p and no packet arrival is q .

The idea can be extended to service completion probability. Suppose a job (packet) is receiving service (being processed) and at each time when it is in service that service could end with probability p or continue with probability q . Then the service process is based on a Bernoulli process.

2.5.2 Geometric Distribution

Geometric distribution is the most commonly used discrete time inter-arrival or service time distribution. Its attraction in queuing theory is its lack of *memory property*.

Consider a random variable X which has only two possible outcomes - *success* or *failure*, represented by the state space $\{0, 1\}$, i.e. 0 is failure and 1 is success. Let q be the probability that an outcome is a failure, i.e. $Pr\{X = 0\} = q$ and $Pr\{X = 1\} = p = 1 - q$ be the probability that it is a success.

The mean number of successes in one trial is given as

$$E[X] = 0 \times q + 1 \times p = p.$$

Let $\theta^*(z)$ be the z -transform of this random variable and given as

$$\theta^*(z) = q + pz.$$

Also we have

$$E[X] = \frac{d(q + pz)}{dz} \Big|_{z \rightarrow 1} = p. \quad (2.9)$$

Suppose we carry out an experiment which has only two possible outcomes 0 or 1 and each experiment is independent of the previous and they all have the same outcomes X , we say this is a Bernoulli process, and $\theta(z)$ is the z -transform of this Bernoulli process. Further, let τ be a random variable that represents the time (number of trials) by which the first success occurs, where all the trials are independent of each other. It is simple to show that if the first success occurs at the τ^{th} trial then the first $\tau - 1$ trials must have been failures. The random variable τ has a geometric distribution and

$$Pr\{\tau = n\} = q^{n-1}p, \quad n \geq 1. \quad (2.10)$$

The mean interval for a success is

$$E[X] = \sum_{n=1}^{\infty} nq^{n-1}p = p^{-1}. \quad (2.11)$$

We chose to impose the condition that $n \geq 1$ because in the context of discrete time queueing systems our inter-arrival times and service times have to be at least one unit of time long, respectively. Let the z -transform of this geometric distribution be $T^*(z)$, we have

$$T^*(z) = pz(1 - qz)^{-1}, \quad |z| \leq 1. \quad (2.12)$$

We can also obtain the mean time to success from the z -transform as

$$E[X] = \frac{T^*(z)}{dz} \Big|_{z \rightarrow 1} = p^{-1}. \quad (2.13)$$

In the context of arrival process, T is the inter-arrival time. Success implies an arrival. Hence, at any time, the probability of an arrival is p and no arrival is q . This is known as the Bernoulli process.

Let Z_m be the number of arrivals in the time interval $[0, m]$, with $\Pr\{Z_m = j\} = a_{m,j}$, $j \geq 0$, $m \geq 1$ and $a_m^*(z) = \sum_{j=0}^m a_{m,j} z^j$. Since the outcome of time trial follows a Bernoulli process then we have

$$a_m^*(z) = (q + pz)^m, \quad m \geq 1, |z| \leq 1. \quad (2.14)$$

The distribution of Z_m is given by the Binomial distribution as follows:

$$\Pr\{Z_m = i\} = a_{m,i} = \binom{m}{i} q^{m-i} p^i, \quad 0 \leq i \leq m \quad (2.15)$$

The variable $a_{m,i}$ is simply the coefficient of z^i in the term $a_m^*(z)$. The mean number of successes in m trials is given as

$$E[Z_m] = \frac{a_m^*(z)}{dz} \Big|_{z \rightarrow 1} = mp. \quad (2.16)$$

In the context of service times, T is the service time of a customer. Success implies the completion of a service. Hence, at any time the probability of completing an ongoing service is p and no service completion is q .

This is what is known as the lack of memory property.

2.5.2.1 Lack of Memory Property:

This lack of memory property is a feature that makes geometric distribution very appealing for use in discrete stochastic modelling. It is shown as follows:

$$\Pr\{T = n + 1 | T > n\} = \frac{\Pr\{T = n + 1\}}{\Pr\{T > n\}} = \frac{q^n p}{\sum_{m=n+1}^{\infty} q^{m-1} p} = \frac{q^n p}{q^n} = p \quad (2.17)$$

which implies that the duration of the remaining portion of a service time is independent of how long the service had been going on, i.e. *lack of memory*.

The mean inter-arrival time or mean service time is given by p^{-1} .

Throughout the rest of this book when we say a distribution is geometric with parameter p we imply that it is governed by a Bernoulli process that has the success probability of p .

While there are several discrete distributions of interest to us in queueing theory, we find that most of them can be studied under the general structure of what is known as the Phase type distribution.

2.5.3 Phase Type Distribution

Phase type distributions are getting to be very commonly used these days after Neuts (1981) made them very popular and easily accessible. They are often referred to as the PH distribution. The PH distribution has become very popular in stochastic modelling because it allows numerical tractability of some difficult problems and in addition several distributions encountered in queueing seem to resemble the PH distribution. In fact, Johnson and Taaffe (1989) have shown that most of the commonly occurring distributions can be approximated by the phase type distributions using moment matching approach based on three moments. The approach is based on using mixtures of two Erlang distributions - not necessarily of common order. They seem to obtain very good fit for most of the cases which they studied. Other works of fitting phase-type distributions include those of Asmussen and Nerman (1991), Bobbio and Telek (1992), and Bobbio and Cumani (1992). The data fitting works by most of these authors are for the continuous PH distributions. There are several other works by Telek and his team for the fitting discrete PH (Bobbio et al (2004)).

Phase type distributions are distributions of the time until absorption in an absorbing Markov chain. If after an absorption the chain is restarted, then it represents the distribution of a renewal process.

Consider an $(n_t + 1)$ absorbing discrete time Markov chain (DTMC) with state space $\{0, 1, 2, \dots, n_t\}$ and let state 0 be the absorbing state. Let T be an m -dimension sub-stochastic matrix with entries

$$T = \begin{bmatrix} T_{1,1} & \cdots & T_{1,n_t} \\ \vdots & \vdots & \vdots \\ T_{n_t,1} & \cdots & T_{n_t,n_t} \end{bmatrix}$$

and also let

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_{n_t}], \text{ and } \boldsymbol{\alpha} \mathbf{1} \leq 1.$$

An example of a PH with $n = 2$ states is shown in Fig. 2.1.

In this context, α_i is the probability that the system starts from a transient state i , $1 \leq i \leq n_t$, and T_{ij} is the probability of transition from a transient state i to a transient state j . We say the phase type distribution is characterized by $(\boldsymbol{\alpha}, T)$ of dimension n_t . We also define α_0 such that $\boldsymbol{\alpha} \mathbf{1} + \alpha_0 = 1$.

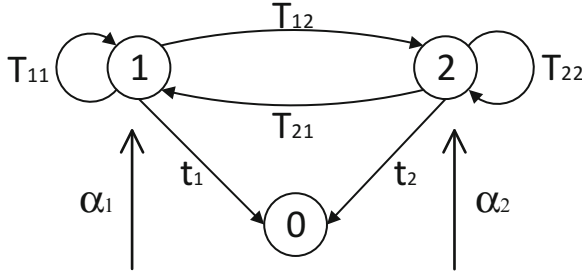


Fig. 2.1 A Two State PH Distribution Diagram

The transition matrix P of this absorbing Markov chain is given as

$$P = \begin{bmatrix} 1 & 0 \\ \mathbf{t} & T \end{bmatrix} \quad (2.18)$$

where $\mathbf{t} = \mathbf{1} - T\mathbf{1}$.

The phase type distribution with parameters $\boldsymbol{\alpha}$ and T is usually written as PH distribution with representation $(\boldsymbol{\alpha}, T)$. The matrix T and vector $\boldsymbol{\alpha}$ satisfy the following conditions:

Conditions:

1. Every element of T is between 0 and 1, i.e. $0 \leq T_{ij} \leq 1$.
2. At least for one row i of T we have $\sum_{j=1}^{n_t} T_{ij} \mathbf{1} < 1$.
3. The matrix $T + \mathbf{t}\boldsymbol{\alpha}$ is irreducible.
4. $\boldsymbol{\alpha}\mathbf{1} \leq 1$ and $\alpha_0 = 1 - \boldsymbol{\alpha}\mathbf{1}$.

Throughout this book, whenever we write a PH distribution $(\boldsymbol{\alpha}, T)$ there is always a bolded lower case column vector associated with it; in this case \mathbf{t} . As another example, if we have a PH distribution $(\boldsymbol{\beta}, S)$ then there is a column vector \mathbf{s} which is given as $\mathbf{s} = \mathbf{1} - S\mathbf{1}$.

If we now define p_i as the probability that the time to absorption into state $n_t + 1$ is i , then we have

$$p_0 = \alpha_0, \quad (2.19)$$

$$p_i = \boldsymbol{\alpha} T^{i-1} \mathbf{t}, \quad i \geq 1. \quad (2.20)$$

Let $p^*(z)$ be the z -transform of this PH distribution, then

$$p^*(z) = \alpha_0 + z\boldsymbol{\alpha}(I - zT)^{-1}\mathbf{t}, \quad |z| \leq 1. \quad (2.21)$$

Then n th factorial moment of the time to absorption is given as

$$\mu'_n = n! \boldsymbol{\alpha} T^{n-1} (I - T)^{-n} \mathbf{1}. \quad (2.22)$$

Specifically the mean time to absorption is

$$\mu'_1 = E[X] = \boldsymbol{\alpha} (I - T)^{-1} \mathbf{1}. \quad (2.23)$$

We will show later in the study of the phase renewal process that also

$$\mu'_1 = E[X] = (\boldsymbol{\pi} \mathbf{t})^{-1}, \quad (2.24)$$

where

$$\boldsymbol{\pi} = \boldsymbol{\pi} (T + \mathbf{t} \boldsymbol{\alpha}), \quad \boldsymbol{\pi} \mathbf{1} = 1.$$

Example:

Consider a phase type distribution with representation $(\boldsymbol{\alpha}, T)$ given as

$$T = \begin{bmatrix} 0.1 & 0.2 & 0.05 \\ 0.3 & 0.15 & 0.1 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}, \quad \boldsymbol{\alpha} = [0.3 \ 0.5 \ 0.2], \quad \alpha_0 = 0.$$

For this $\alpha_0 = 0$ and $\mathbf{t} = [0.65 \ 0.45 \ 0.2]^T$. We have

$$p_1 = 0.46, p_2 = 0.2658, p_3 = 0.1346, p_4 = 0.0686, p_5 = 0.0349,$$

$$p_6 = 0.0178, p_7 = 0.009, p_8 = 0.0046, p_9 = 0.0023, p_{10} = 0.0012,$$

$$p_{11} = 0.00060758, p_{12} = 0.0003093, \dots$$

Alternatively we may report our results as the complement of the cumulative distribution, i.e.

$$P_k = Pr\{X \geq k\} = 1 - Pr\{X \leq k\} = \boldsymbol{\alpha} T^{k-1} (I - T)^{-1} \mathbf{t}.$$

This is given as

$$P_1 = 1.0, P_2 = 0.54, P_3 = 0.2743, P_4 = 0.1397, P_5 = 0.0711, P_6 = 0.0362, \dots$$

2.5.3.1 Two very important closure properties of phase type distributions:

Consider two discrete random variables X and Y that have phase type distributions with representations (α, T) and (β, S) .

1. **Sum:** Their sum $Z = X + Y$ has a phase type distribution with representation (δ, D) with

$$D = \begin{bmatrix} T & \mathbf{t}\beta \\ \mathbf{0} & S \end{bmatrix}, \quad \delta = [\alpha \quad \alpha_0\beta].$$

2. **Mixture:** Their mixture with $[\theta_1, \theta_2]$, $0 \leq \theta_i \leq 1$, $i = 1, 2$, mixing density $(\theta_1 + \theta_2 = 1)$ has a phase type distribution with representation (δ, D) with

$$D = \begin{bmatrix} T & \mathbf{0} \\ \mathbf{0} & S \end{bmatrix}, \quad \delta = [\theta_1\alpha \quad \theta_2\beta].$$

3. **Minimum:** Their minimum $W = \min(X, Y)$ has a phase type distribution with representation (δ, D) with

$$D = T \otimes S, \quad \delta = [\alpha \otimes \beta].$$

4. **Maximum:** Their maximum $U = \max(X, Y)$ has a phase type distribution with representation (δ, D) with

$$D = \begin{bmatrix} T \otimes S & T \otimes \mathbf{t} & S \otimes \mathbf{t} & S \otimes \mathbf{s} \\ \mathbf{0} & T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \delta = [\alpha \otimes \beta, \alpha_0\beta, \alpha\beta_0, \mathbf{0}].$$

These results can be extended directly to the case of more than two PH distributions.

2.5.3.2 Minimal coefficient of variation of a discrete PH distribution

The coefficient of variation (cv) of a discrete PH distribution has a different behaviour compared to its continuous counterpart. For example, for some integer $K < \infty$ a random variable X with

$$Pr\{X = k\} = \begin{cases} 1, & k = K, \\ 0, & k \neq K, \end{cases}, \quad (2.25)$$

can be represented by the discrete PH distribution. This PH distribution has a cv of zero. This type of case with cv of zero is not encountered in the continuous PH. This information can sometimes be used to an advantage when trying to fit a dataset to a

discrete PH distribution. In general for the discrete PH the coefficient of variation is a function of its mean.

A general inequality for the minimal cv of a discrete PH distribution was obtained by Telek (2000) as follows. Consider the discrete PH distribution (α, T) of order n_t . Its mean is given by μ' . Let us write μ' as $\mu' = \lfloor \mu' \rfloor + \langle \mu' \rangle$ where $\lfloor \mu' \rfloor$ is the integer part of μ' and $\langle \mu' \rangle$ is the fractional part with $0 \leq \langle \mu' \rangle < 1$. Telek (2000) proved that the cv of this discrete PH distribution written as $cv(\alpha, T)$ satisfies the inequality

$$cv(\alpha, T) \geq \begin{cases} \frac{\langle \mu' \rangle (1 - \langle \mu' \rangle)}{(\mu')^2}, & \mu' < n_t, \\ \frac{1}{n_t} - \frac{1}{\mu'}, & \mu' \geq n_t. \end{cases} \quad (2.26)$$

The proof for this can be found in Telek Telek (2000).

Throughout this book we will assume that $\alpha_0 = 0$, since we are dealing with queueing systems for which we do not allow inter-arrival times or service times to be zero.

2.5.3.3 Examples of special phase type distributions

Some special cases of discrete phase type distributions include:

1. **Geometric distribution** with $\alpha = 1$, $T = q$, $\mathbf{t} = p = 1 - q$. Then $p_0 = 0$, and $p_i = q^{i-1}p$, $i \geq 1$ and $n_t = 1$. For this distribution, the mean $\mu' = 1/p$ and the $cv(1, q) = 1 - 1/p$
2. **Negative binomial distribution** with

$$\alpha = [1, 0, 0, \dots, 0], \text{ and } T = \begin{bmatrix} q & p & & & \\ & q & p & & \\ & & \ddots & \ddots & \\ & & & q & p \\ & & & & q \end{bmatrix},$$

and n_t is the number of successes we are looking for occurring at the i^{th} trial. It is easy to show that

$$p_i = \alpha T^{i-1} \mathbf{t} = \binom{i-1}{n_t-1} p^{n_t} q^{i-n_t}, \quad i \geq n_t.$$

3. **Mixed Geometric distribution** with

$$\alpha = [\theta_1, \theta_2, \dots, \theta_n], \quad 0 \leq \theta_i \leq 1, \quad i = 1, 2, \dots, n \text{ and } T = \begin{bmatrix} q_1 & & & \\ & q_2 & & \\ & & \ddots & \\ & & & q_n \end{bmatrix},$$

with $\sum_{i=1}^n \theta_i = 1$ and $0 < q_i < 1, \forall i$.

4. **Multiple-time-scaled PH Distribution:** Consider a random variable X with PH distribution $(\boldsymbol{\alpha}, T)$ of dimension n . Now consider another random variable Y such that

$$Pr\{X = j\} = Pr\{Y = kj\}, j = 1, 2, \dots; k = 1, 2, \dots,$$

with

$$E[Y] = kE[X].$$

We find that Y has a PH distribution $(\boldsymbol{\beta}, S)$ of dimension kn and its parameters are given as

$$\boldsymbol{\beta} = \boldsymbol{\alpha} \otimes \mathbf{e}_{k-1}^T(k), \quad S = \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \dots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix},$$

where $S_{ii} = \begin{bmatrix} \mathbf{0} & I_{k-1} \\ T_{ii} & \mathbf{0}^T \end{bmatrix}$ and $S_{ij} = T_{ij} \mathbf{e}_k(k) \otimes \mathbf{e}_k(k)^T, i \neq j$.

5. **General discrete distribution with finite support** can be represented by a discrete phase type distribution with $\boldsymbol{\alpha} = [1, 0, 0, \dots, 0]$ and $0 < t_{ij} \leq 1, j = i + 1$, and $t_{ij} = 0, j \neq i + 1$ where n_t is the length of the support. For example, for a constant inter-arrival time with value of $n_t = 4$ we have $\boldsymbol{\alpha} = [1, 0, 0, 0]$ and

$$T = \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & \end{bmatrix}.$$

Note that the general distribution with finite support can also be represented as a phase type distribution with $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_{n_t}]$ and $t_{ij} = 1, i = j - 1$, and $t_{ij} = 0, i \neq j - 1$. For the example of constant inter-arrival time with value of $n_t = 4$ we have $\boldsymbol{\alpha} = [0, 0, 0, 1]$ and

$$T = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The case of general discrete distribution will be discussed in detail later.

2.5.3.4 Analogy between PH and Geometric distributions

Essentially discrete phase type distribution is simply a matrix version of the geometric distribution. The geometric distribution has parameters p for success (or arrival) and q for failure (no arrival). The discrete phase type with representation $(\boldsymbol{\alpha}, T)$ has $\mathbf{t}\boldsymbol{\alpha}$ for success (arrival) and T for failure (no arrival). So if we consider this in the context of a Bernoulli process we have the z -transform of this process as $\theta^*(z) = T + z\mathbf{t}\boldsymbol{\alpha}$. Next we discuss the Phase renewal process.

2.5.3.5 Phase Renewal Process:

Consider an inter-event time X which is described by a phase type distribution with the representation $(\boldsymbol{\alpha}, T)$. The matrix T records transitions with no event occurring and the matrix $\mathbf{t}\boldsymbol{\alpha}$ records the occurrence of an event and the re-start of the renewal process. If we consider the interval $(0, n)$ and define the number of renewals in this interval as $N(n)$ and the phase of the PH distribution at phase $J(n)$ at time n , and define

$$P_{i,j}(k, n) = \Pr\{N(n) = k, J(n) = j | N(0) = 0, J(0) = i\},$$

and the associated matrix $P(k, n)$ such that $(P(k, n))_{i,j} = P_{i,j}(k, n)$, we have

$$P(0, n+1) = TP(0, n), \quad n \geq 0 \quad (2.27)$$

$$P(k, n+1) = TP(k, n) + (\mathbf{t}\boldsymbol{\alpha})P(k-1, n), \quad k = 1, 2, \dots; n \geq 0. \quad (2.28)$$

Define

$$P^*(z, n) = \sum_{k=0}^n z^k P(k, n), \quad n \geq 0.$$

We have

$$P^*(z, n) = (T + z(\mathbf{t}\boldsymbol{\alpha}))^n. \quad (2.29)$$

This is analogous to the $(p + zq)^m$ for the z -transform of the Binomial distribution presented earlier. It is immediately clear that the matrix $T^* = T + \mathbf{t}\boldsymbol{\alpha}$ is a stochastic matrix that represent the transition matrix of the phase process associated with this process.

Secondly, we have T as the matrix analogue of p in the Bernoulli process while $\mathbf{t}\boldsymbol{\alpha}$ is the analogue of q in the Bernoulli process. Hence for one time epoch $T + z(\mathbf{t}\boldsymbol{\alpha})$ is the z -transform of an arrival. This phase renewal process can be found in state i in the long run with probability π_i where $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_{n_t}]$ and it is given by

$$\boldsymbol{\pi} = \boldsymbol{\pi}(T + \mathbf{t}\boldsymbol{\alpha}), \quad (2.30)$$

and

$$\boldsymbol{\pi}\mathbf{1} = 1. \quad (2.31)$$

Hence the average number of arrivals in one time unit is

$$E[Z_1] = \boldsymbol{\pi}(0 \times T + 1 \times (\mathbf{t}\boldsymbol{\alpha}))\mathbf{1} = \boldsymbol{\pi}(\mathbf{t}\boldsymbol{\alpha})\mathbf{1} = \boldsymbol{\pi}\mathbf{t}. \quad (2.32)$$

This is the arrival rate and its inverse is the mean inter-arrival time of the corresponding phase type distribution, as pointed out earlier on.

Define r_n as the probability of a renewal at time n , then we have

$$r_n = \boldsymbol{\alpha}(T + \mathbf{t}\boldsymbol{\alpha})^{k-1}\mathbf{t}, \quad k \geq 1. \quad (2.33)$$

Keep in mind that $\boldsymbol{\pi}$ is the solution to the equations

$$\boldsymbol{\pi} = \boldsymbol{\pi}(T + \mathbf{t}\boldsymbol{\alpha}), \quad \boldsymbol{\pi}\mathbf{1} = 1.$$

This $\boldsymbol{\pi}$ represents the probability vector of the PH renewal process being found in a particular phase in the long term. If we observe this phase process at arbitrary times given that it has been running for a while, then the remaining time before an event (or for that matter the elapsed time since an event) also has a phase type distribution with representation $(\boldsymbol{\pi}, T)$. It was shown in Latouche and Ramaswami (1999) that

$$\boldsymbol{\pi} = (\boldsymbol{\alpha}(I - T)^{-1}\mathbf{1})^{-1}\boldsymbol{\alpha}(I - T)^{-1}. \quad (2.34)$$

Both the inter-arrival and service times can be represented by the phase type distributions. Generally, phase type distributions are associated with cases in which the number of phases are finite. However, recently, the case with infinite number of phases has started to receive attention.

2.5.4 The infinite phase distribution (IPH)

The IPH is a discrete phase type distribution with infinite number of phases. It is still represented as $(\boldsymbol{\alpha}, T)$, except that now we have the number of phases $n_t = \infty$. The IPH was introduced by Shi and Liu (1998). One very important requirement is that the matrix T be irreducible and $T\mathbf{1} \leq \mathbf{1}$, with at least one row being strictly less than 1.

If we now define p_i as the probability that the time to absorption into a state we label as $*$ is i , then we have

$$\begin{aligned} p_0 &= \alpha_0 \\ p_i &= \boldsymbol{\alpha}T^{i-1}\mathbf{t}, \quad i \geq 1 \end{aligned}$$

Every other measure carried out for the PH can be easily derived for the IPH. However, we have to be cautious in many instances. For example, the inverse of $I - T$ may not be unique and as such we have to define it as appropriate for the situation under consideration. Secondly computing p_i above requires special techniques at times depending on the structure of the matrix T . The rectangular iteration was proposed by Shi et al (1996) for such computations.

2.5.5 General Inter-event Times

General types of distributions, other than of the phase types, can be used to describe both inter-arrival and service times. In continuous times it is well known that general distributions encountered in queueing systems can be approximated by continuous time PH distributions. This is also true for discrete distributions. However, discrete distributions have an added advantage in that if the distribution has a finite support then it can be represented exactly by discrete PH. We proceed to show how this is true by using a general inter-event time X with finite support and a general distribution given as

$$Pr\{X = j\} = a_j, j = 1, 2, \dots, n_t < \infty.$$

There are at least two exact PH representations for this distribution, one based on remaining time and the other on elapsed time.

2.5.5.1 Remaining Time Representation

Consider a PH distribution (α, T) of dimension n_t . Let

$$\alpha = [a_1, a_2, \dots, a_{n_t}], \quad (2.35)$$

and

$$T_{i,j} = \begin{cases} 1, & j = i - 1 \\ 0, & \text{otherwise,} \end{cases} \quad (2.36)$$

Then the distribution of this PH is given as

$$\alpha T^{k-1} \mathbf{t} = a_k; k = 1, 2, \dots, n_t. \quad (2.37)$$

For detailed discussion on this see Alfa (2004).

In general, even if the support is not finite we can represent the general inter-event times with the IPH, by letting $n_t \rightarrow \infty$. For example, consider the inter-event times \mathcal{A} which assume values in the set $\{1, 2, 3, \dots\}$ with $a_i = Pr\{\mathcal{A} = i\}$, $i = 1, 2, 3, \dots$. It is easy to represent this distribution as an IPH with $\alpha = [a_1, a_2, \dots]$ and $T = \begin{bmatrix} 0 & 0 \\ I_\infty & 0 \end{bmatrix}$.

2.5.5.2 Elapsed Time Representation

Consider a PH distribution $(\boldsymbol{\alpha}, T)$ of dimension n_t . Let

$$\boldsymbol{\alpha} = [1, 0, \dots, 0], \quad (2.38)$$

and

$$T_{i,j} = \begin{cases} \tilde{a}_i, & j = i + 1 \\ 0, & \text{otherwise,} \end{cases} \quad (2.39)$$

where

$$\tilde{a}_i = \frac{u_i}{u_{i-1}}, \quad u_i = 1 - \sum_{v=1}^i a_v, \quad u_0 = 1, \quad \tilde{a}_{n_t} = 0.$$

Then the distribution of this PH is given as

$$\boldsymbol{\alpha} T^{k-1} \mathbf{t} = a_k; \quad k = 1, 2, \dots, n_t. \quad (2.40)$$

For detailed discussion on this see Alfa (2004). Similarly we can use IPH to represent this distribution using elapsed time by allowing $n_t \rightarrow \infty$.

2.5.6 Markovian Arrival Process

All the inter-event times discussed so far are of the renewal types and are assumed to be independent and identically distributed (*iid*). However, in telecommunication queueing systems and most other traffic queueing systems for that matter, inter-arrival times are usually correlated. So the assumption of independent inter-arrival times is not valid in some instances.

Earlier Neuts (1979, 1992) and Lucantoni (1991) presented the Markovian arrival process (MAP) which can handle correlated arrivals and is also tractable mathematically. In what follows, we first describe the single arrival MAP and then briefly present the batch MAP.

Define two sub-stochastic matrices D_0 and D_1 , both of the dimensions n . The elements $(D_0)_{ij}$ refer to transition from state i to state j without an (event) arrival because the transitions are all within the n_t transient states. The elements $(D_1)_{ij}$ refer to transition from state i into the absorbing state 0 with an instantaneous restart from the transient state j with an (event) arrival during the absorption. We note that the phase from which an absorption occurred and the one from which the next process starts are connected and hence this captures the correlation between inter-arrival times. The matrix $D = D_0 + D_1$ is a stochastic matrix, and we assume it is

irreducible. Note that $D\mathbf{1} = \mathbf{1}$. If we define $\{(N_n, J_n), n \geq 0\}$ as the total number of arrivals and the phase of the MAP at time n , then the transition matrix representing this system is

$$P = \begin{bmatrix} D_0 & D_1 & & & \\ & D_0 & D_1 & & \\ & & D_0 & D_1 & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{bmatrix}. \quad (2.41)$$

Consider the discrete-time Markov renewal process embedded at the arrival epochs and with transition probabilities defined by the sequence of matrices

$$Q(k) = [D_0]^{k-1} D_1, \quad k \geq 1. \quad (2.42)$$

The MAP is a discrete-time point process generated by the transition epochs of that Markov renewal process.

Once more let N_m be the number of arrivals at time epochs $1, 2, \dots, m$, and J_m the state of the Markov process at time m . Let $P_{r,s}(n, m) = \Pr\{N_m = n, J_m = s \mid N_0 = 0, J_0 = r\}$ be the (r, s) entry of a matrix $P(n, m)$. The matrices $P(n, m)$ satisfy the following discrete Chapman-Kolmogorov difference equations:

$$P(n, m+1) = P(n, m)D_0 + P(n-1, m)D_1, \quad n \geq 1, m \geq 0 \quad (2.43)$$

$$P(0, m+1) = P(0, m)D_0 \quad (2.44)$$

$$P(0, 0) = I \quad (2.45)$$

where I is the identity matrix and $P(u, v) = 0$ for $u \geq v+1$.

The matrix generating function

$$P^*(z, m) = \sum_{n=0}^m P(n, m) z^n, \quad |z| \leq 1 \quad (2.46)$$

is given by

$$P^*(z, m) = (D_0 + zD_1)^m, \quad m \geq 0 \quad (2.47)$$

If the stationary vector π of the Markov chain described by D satisfies the equation

$$\pi D = \pi, \quad (2.48)$$

and

$$\pi \mathbf{1} = 1 \quad (2.49)$$

then $\lambda' = \pi D_1 \mathbf{1}$ is the probability that, in the stationary version of the arrival process, there is an arrival at an arbitrary point. The parameter λ' is the expected number of arrivals at an arbitrary time epoch or the discrete arrival rate of the MAP.

It is clear that its k^{th} moments about zero are all the same for $k \geq 1$, because $\mathbf{1}^j \pi D_1 \mathbf{1} = \pi D_1 \mathbf{1} = \lambda'$, $\forall j \geq 1$, since its j^{th} moment about zero is given as $\sum_{k=0}^{\infty} k^j \pi D_k \mathbf{1}$. Hence its variance σ_X^2 is given as

$$\sigma_X^2 = \lambda'' - (\lambda')^2 = \lambda' - (\lambda')^2 = \lambda'(1 - \lambda'). \quad (2.50)$$

Its j^{th} autocorrelation factor for the number of arrivals, $ACF(j)$, is given as

$$ACF(j) = \frac{1}{\sigma_X^2} [\pi D_1 D^{j-1} D_1 \mathbf{1} - (\lambda')^2], \quad j \geq 1. \quad (2.51)$$

The autocorrelation between inter-arrival times is captured as follows. Let X_i be the i^{th} inter-arrival time then we can say that

$$Pr\{X_i = k\} = \pi D_0^{k-1} D_1 \mathbf{1}, \quad k \geq 1, \quad (2.52)$$

where $\pi = \pi(I - D_0)^{-1} D_1$, $\pi \mathbf{1} = 1$. Then $E[X_i]$, the expected value of X_i , $\forall i$, is given as

$$E[X_i] = \sum_{j=1}^{\infty} j \pi D_0^{j-1} D_1 \mathbf{1} = \pi(I - D_0)^{-1} \mathbf{1}. \quad (2.53)$$

The autocorrelation sequence r_k , $k = 1, 2, \dots$ for the inter-arrival times is thus given as

$$r_k = \frac{E[X_\ell X_k] - E[X_\ell]E[X_k]}{E[X_\ell^2] - (E[X_\ell])^2}, \quad \ell = 0, 1, 2, \dots. \quad (2.54)$$

Special Cases:

The simplest MAP is the Bernoulli process with $D_1 = q$ and $D_0 = p = 1 - q$. The discrete phase type distribution is also a MAP with $D_1 = \mathbf{t}\alpha$ and $D_0 = T$.

Another interesting example of MAP is the Markov modulated Bernoulli process (MMBP) which is controlled by a Markov chain which has a transition matrix P and rates given by the $\theta = \text{diag}(\theta_1, \theta_1, \dots, \theta_n)$, $0 \leq \theta_i \leq 1$, $i = 1, 2, \dots, n$. In this case, $D_0 = (I - \theta)P$ and $D_1 = \theta P$.

Several examples of discrete MAP can be found in Alfa and Neuts (1995), Alfa and Chakravathy (1994), Alfa et al (1995), Liu95, Park et al (1994) and Blondia (1992).

2.5.6.1 Platoon Arrival Process (PAP)

The PAP is a special case of MAP which occurs in several traffic situations mainly in telecommunications and vehicular types of traffic. It captures distinctly, in addition to correlation in arrival process, the bursts in traffic arrival process termed platoons here.

The PAP is an arrival process with two regimes of traffic. There is a platoon of traffic (group of packets) which has intra-platoon intervals of arrival times that are identical. The number of arrivals in a platoon is random. At the end of a platoon there is an inter-platoon interval between the end of one platoon and the start of the next platoon arrivals. The inter-platoon intervals are different from the intra-platoon intervals. A good example is if one observes the departure of packets from a router queue, one observes a platoon departure that consists of traffic departing as part of a busy period. The inter-platoon times are essentially the service times. The last packet that departs at the end of a busy period marks the end of a platoon. The next departure will be the head of a new platoon. The interval between the last packet of a platoon and the first packet of the next platoon is the inter-platoon time which in this case is the sum of the service time and the residual inter-arrival time of a packet into the router. If we consider this departure from the router as an input to another queueing system, then the arrival to this other queueing system is a PAP. This type of arrival pattern is also noticed at a signalized road traffic intersection, where departing vehicles up form some kind of platoon when the light turns green. This platoon eventually dissipates, depending on how long the green light lasts.

A PAP is described as follows. Let packets arrive in platoons of random sizes with probability mass function (*pmf*) given as $\{p_k, k \geq 1\}$, i.e. p_k is the probability that the number in a platoon is k . Time intervals between arrivals of packets in the same platoon denoted as intra-platoon inter-arrival times have a *pmf* denoted as $\{p_1(j), j \geq 1\}$. The time intervals between the arrival of the last packet in a platoon and the first packet of the next platoon is referred to as the inter-platoon inter-arrival times and has the *pmf* denoted by $\{p_2(j), j \geq 1\}$. Let the size of platoon be distributed according to the PH distribution (δ, F) . For this PH let us allow $\delta_0 = 1 - \delta \mathbf{1} \geq 0$, and define $\mathbf{f} = \mathbf{1} - F\mathbf{1}$. Here we have

$$p_k = \begin{cases} \delta_0, & k = 1 \\ \delta F^{k-1} \mathbf{f}, & k = 1, 2, \dots \end{cases} \quad (2.55)$$

The platooned arrival process (PAP) is a discrete-time Markov renewal process whose transition probabilities are described by the sequence of matrices

$$f(j) = \begin{bmatrix} \delta_0 p_2(j) & \delta p_2(j) \\ \mathbf{f} p_1(j) & F p_1(j) \end{bmatrix}, \quad j \geq 1. \quad (2.56)$$

If we let the intra-platoon and inter-platoon inter-arrival times assume PH distribution then we actually end up with a PAP that is a special case of a MAP.

Let (α_1, T_1) be a PH distribution that describes intra-platoon times and (α_2, T_2) a PH distribution describing the inter-platoon times. Then the PAP is described by two matrices D_0 and D_1 , with

$$D_0 = \begin{bmatrix} T_2 & \mathbf{0} \\ \mathbf{0} & I \otimes T_1 \end{bmatrix}, \quad (2.57)$$

and

$$D_1 = \begin{bmatrix} \delta_0 \mathbf{t}_2 \alpha_2 & \delta \otimes \mathbf{t}_2 \alpha_1 \\ \mathbf{f} \otimes \mathbf{t}_1 \alpha_2 & F \otimes \mathbf{t}_1 \alpha_1 \end{bmatrix}. \quad (2.58)$$

Note that $\mathbf{f} = \mathbf{1} - F\mathbf{1}$, $\mathbf{t}_k = \mathbf{1} - T_k\mathbf{1}$, $k = 1, 2$.

Next we explain the elements of the matrices D_0 and D_1 .

For the matrix D_0 we have

- the matrix T_2 captures the phase transitions during an inter-platoon arrival time, and
- the matrix $I \otimes T_1$ captures the transitions during an inter-arrival time within a platoon.

For the matrix D_1 we have

- the matrix $\delta_0 \mathbf{t}_2 \alpha_2$ captures the arrival of a platoon of a single packet type, with an end to the arrival of a platoon and the initiation of an inter-platoon inter-arrival time process, whilst
- the matrix $\delta \otimes \mathbf{t}_2 \alpha_1$ captures the arrival of a platoon consisting of at least two packets, with an end to the inter-platoon inter-arrival and the initiation of an intra-platoon inter-arrival time process.
- The matrix $F \otimes \mathbf{t}_1 \alpha_1$ captures the arrival of an intra-platoon packet, with an end to the intra-platoon inter-arrival and the initiation of a new intra-platoon inter-arrival time process, whilst
- the matrix $\mathbf{f} \otimes \mathbf{t}_1 \alpha_2$ captures the arrival of the last packet in a platoon, with an end to the intra-platoon inter-arrival and the initiation of an inter-platoon inter-arrival time process.

A simple example of this is where intra-platoon and inter-platoon inter-arrival times follow geometric distributions, with $p_1(j) = (1 - a_1)^{j-1} a_1$, $p_2(j) = (1 - a_2)^{j-1} a_2$, $j \geq 1$, and the distribution of platoon size is geometric with $p_0 = 1 - b$, $p_k = b(1 - c)^{k-1} c$, $k \geq 1$, where $0 < (a_1, a_2, b, c) < 1$. In this case we have

$$D_0 = \begin{bmatrix} 1 - a_2 & 0 \\ 0 & 1 - a_1 \end{bmatrix}, \quad D_1 = \begin{bmatrix} ba_2 (1 - b)a_2 \\ ba_1 (1 - b)a_1 \end{bmatrix}.$$

For a detail discussion of the discrete PAP see Alfa and Neuts (1995).

2.5.6.2 Batch Markovian Arrival Process (BMAP)

Define substochastic matrices D_k , $k \geq 0$, such that $D = \sum_{k=0}^{\infty} D_k$ is stochastic. The elements $(D_k)_{ij}$ refer to a transition from state i to state j with $k \geq 0$ arrivals.

If we define π such that

$$\pi D = \pi, \quad \pi \mathbf{1} = 1$$

then the arrival rate $\lambda' = E[X] = \pi \sum_{k=1}^{\infty} k D_k \mathbf{1}$. Let $\lambda'' = E[X^2] = \pi \sum_{k=1}^{\infty} k^2 D_k \mathbf{1}$ be its second moment about zero, then its variance σ_X^2 is given as

$$\sigma_X^2 = \lambda'' - (\lambda')^2. \quad (2.59)$$

Its j^{th} autocorrelation factor $ACF(j)$ is given as

$$ACF(j) = \frac{1}{\sigma_X^2} \left[\pi \left(\sum_{k=1}^{\infty} k D_k \right) D^{j-1} \left(\sum_{k=1}^{\infty} k D_k \right) \mathbf{1} - (\lambda')^2 \right], \quad j \geq 1. \quad (2.60)$$

2.5.7 Marked Markovian Arrival Process

Another class of arrival process of interest in telecommunication is the marked Markovian arrival process (MMAP[K]), with K classes. It is represented by $K + 1$ matrices D_0, D_1, \dots, D_K all of dimension $n_t \times n_t$. The elements $(D_k)_{ij}$, $k = 1, 2, \dots, K$, represent transitions from state i to state j with type k packet arrival and $(D_0)_{ij}$ represents no arrival. Let $D = \sum_{v=0}^K D_v$, then D is a stochastic matrix. For a detailed discussion of this class of arrival process see He and Neuts (1998).

An example application of this type of arrival process is in the telecommunication where we may have different paying classes of customers (users) who demand different types of service. The arrivals of these customers may be correlated and a MMAP can capture this process.

2.5.8 Semi Markov Processes

Service times with correlations can be described by a class of semi-Markov processes, which are the analogue of the MAP. For such examples, see Alfa and Chakravathy (1994) and Lucantoni and Neuts (1994).

2.5.9 Data Fitting for PH and MAP

Fitting of PH distribution is both a science and an art. This is because PH representation of a given probability distribution function is not unique. As a trivial example, the geometric distribution $r_k = q^{k-1}p$, $k = 1, 2, \dots$, can be represented as a PH distribution with parameters $(1, q)$ or even as (α, T) , where $\alpha = [0, 1, 0]$

and $T = \begin{bmatrix} b_1 & b_2 & b_3 \\ 0 & q & 0 \\ c_1 & c_2 & c_3 \end{bmatrix}$, where $0 \leq (b_i, c_i) \leq 1$ and $b_1 + b_2 + b_3 \leq 1$, $c_1 + c_2 + c_3 \leq 1$.

Several examples of this form can be presented. It is also known that PH representation is non-minimal, as demonstrated by the last example. So, fitting a PH distribution usually involves selecting the number of phases in advance and then finding the best fit using standard statistical methods such as the maximum likelihood method of moments. Alternatively one may select the structure of the PH and then find the best fitting order and parameters. So the art is in the pre-selection process which is often guided by what the PH distribution is going to be used for in the end. For example if the PH is for representing service times in a queueing problem we want to have a small dimension so as to reduce the computational load associated with the queueing model. In some other instances the structure of the PH may be more important if a specific structure will make computation easier.

In general assume we are given a set of N observations of inter-event times y_1, y_2, \dots, y_N , with $\mathbf{y} = [y_1, y_2, \dots, y_N]$. If we want to fit a PH distribution (α, T) of order n and/or known structure to this set of observations we can proceed by of the following two methods:

- **Method of moments:** We need to estimate a maximum of $n^2 + n - 1$ parameters. This is because we need a maximum of n^2 parameters for the matrix T and $n - 1$ parameters for the vector α . If we have a specific structure in mind then the number of unknowns that need to be determined can be reduced. For example if we want to fit a negative binomial distribution of order n then all we need is one parameter. But if it is a general negative binomial then we need n parameters. Let the number of parameters needed to be determined by $m \leq n^2 + n - 1$. Then we need to compute m moments of the dataset \mathbf{y} , which we write as $\tilde{\mu}_k$, $k = 1, 2, \dots, m$. With this knowledge and also knowing that the factorial moments of the PH distribution are given as $\mu'_k = k! \alpha (I - T)^{-k} \mathbf{1}$ we can then obtain the moments of the PH μ_k from the factorial, equate them to the moments from the observed data. This will lead to a set of m non-linear equations, which need to be solved to obtain the best fitting parameters.
- **Maximum likelihood (ML) method:** This second approach is more popularly used. It is based on the likelihoodness of observing the dataset. Let $f_i(y_i, \alpha, T) = \alpha T^{y_i-1} \mathbf{1}$, $i = 1, 2, \dots, N$. Then the likelihood function is

$$\mathcal{L}(\mathbf{y}, \alpha, T) = \prod_{i=1}^N f_i(y_i, \alpha, T).$$

This function is usually converted to its log form and then different methods can be used to find the best fitting parameters.

Data fitting of the PH distribution is outside the scope of this book. However, several methods can be found in Telek (2000) and Bobbio et al (2004). Bobbio et al (2004) specifically presented methods for fitting acyclic PH. Acyclic PH is a special PH (α, T) for which $T_{i,j} = 0, \forall i \geq j$ and $\alpha_1 = 1$.

There has not been much work with regards to the fitting of discrete MAP. However, Breuer and Alfa (2005) did present the ME algorithm based on ML for estimating parameters for the PAP.

2.6 Service Times: What does this really mean?

We conclude this chapter with a point that has not been addressed much in queueing. What does service distribution really imply? Is the distribution because a server cannot guarantee to process a job that requires k units of work time in k units of time due to its own lack of consistency? Or is it that the server can complete k units of work in k units of time but the arriving customers come with different service time requirements? This has a major effect when considering telecommunication problems.

2.7 Problems

1. Question 1: Packets arrive at a router for processing. Each of the packets requires 2 units of time for set up before processing; this is a set up time which we call S_s . After the set up time of 2 units a packet requires additional units of service and this additional units of service has a geometric distribution with parameter a ; i.e. if the additional service is S_a then $Pr\{S_a = j\} = (1 - a)^{j-1}a, j \geq 1$. Therefore the total service time of a packet is $S = S_s + S_a$. Show how this S can be represented by a discrete phase type distribution, and give the representation. What is the mean of S , i.e. $E[S]$? Find the z -transform of this processing time. What is the probability that the processing time of two packets is 8 units of time?
2. Question 2: Consider a service system with each item's service time following a geometric distribution with parameter p . Suppose the server can take a break after serving n items. Let X be the length of time the server works before taking a break and $x_k = Pr\{X = k\}$. Write down the expression for $x_k, k \geq n$. If the service time of an individual item follows a phase type distribution (α, T) , write down the expression for x_k .
3. Question 3: Consider three boxes with infinite capacities. Packets are generated according to the phase type distribution (α, T) . Suppose these items are placed in the boxes according to the following rule; items 1, 4, 7, 10, \dots , go into first box; items 2, 5, 8, 11, \dots , go into the second box and items 3, 6, 9, 12, \dots , go into the third box. Develop an expression that captures the arrival process of items into any of the boxes. Extend the idea to the case of $n \geq 1$ boxes.

4. Question 4: Consider two classes of items A and B . Items A arrive according to the phase type distribution with parameter (α, T) and item B according to phase type distribution (β, S) . Write down an expression for the probability that item A arrives before item B . What is the probability that the first arrival of an item is at time n ?

References

- Alfa AS (2004) Markov chain representations of discrete distributions applied to queueing models. *Computers and Operations Research* 31:2365–2385
- Alfa AS, Chakravorthy S (1994) A discrete queue with Markovian arrival processes and phase type primary and secondary services. *Stochastic Models* 10:437–451
- Alfa AS, Neuts MF (1995) Modelling vehicular traffic using the discrete time Markovian arrival process. *Transportation Science* 29:109–117
- Alfa AS, Dolhun KL, Chakravorthy S (1995) A discrete single server queue with Markovian arrivals and phase type group service. *Journal of Applied Mathematics and Stochastic Analysis* 8:151–176
- Asmussen S, Nerman O (1991) Fitting phase-type distributions via the EM algorithm. In: *Symposium i Anvendt Statistik*, K. Vest Nielsen (ed.), Copenhagen, pp 335–346
- Blondia C (1992) A discrete-time batch Markovian arrival process as B-ISDN traffic model. *Belgian Journal of Operations Research, Statistics and Computer Science* 32:3–23
- Bobbio A, Cumani A (1992) ML estimation of the parameters of a PH distribution in triangular canonical form. *Computer Performance Evaluation*, G Balbo and G Serazzi (ed), Elsevier Science Pub BV pp 33–46
- Bobbio A, Telek M (1992) Parameter estimation of phase type distributions. In: *20th European Meeting of Statisticians*, Bath, UK, pp 335–346
- Bobbio A, Hovarth A, Scarpa M, Telek M (2004) Acyclic discrete phase type distributions: properties and a parameter estimation algorithm. *Performance Evaluation* 54:1–32
- Breuer L, Alfa AS (2005) An EM algorithm for platoon arrival processes in discrete time. *Operations Res Lett* 33:535–543
- He QM, Neuts MF (1998) Markov arrival process with marked transitions. *Stochastic Processes and Applications* 74:37–52
- Johnson M, Taaffe M (1989) Matching moments to phase distributions: mixtures of erlang distributions of common order. *Stochastic Models* 5
- Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic modeling. In: *Pennsylvania*, Philadelphia, Pennsylvania
- Lucantoni DM (1991) New results on single server queue with a batch Markovian arrival process. *Stochastic Models* 7:1–46
- Lucantoni DM, Neuts MF (1994) Some steady-state distributions of the MAP/SM/1 queue. *Stochastic Models* 10
- Neuts MF (1979) A versatile Markovian point process. *Journ Appl Prob* 16:764–779
- Neuts MF (1981) Matrix-geometric solutions in stochastic models - an algorithmic approach. In: *The Johns Hopkins University Press*, Baltimore
- Neuts MF (1992) Models based on the Markovian arrival process. *IEICE Trans Commun* E75-B:1255–1265
- Park D, H P, Yamashita H (1994) Approximate analysis of discrete-time tandem queueing networks with bursty and correlated input traffic and customer loss. *Oper Res Letters* 15:95–104
- Shi D, Liu D (1998) Markovian models for non-negative random variables. In: Alfa AS, Chakravorthy S (eds) *Advances in Matrix-analytic Methods in Stochastic Models*, Notable Publications, pp 403–427

- Shi D, Guo J, Liu L (1996) SPH-distributions and the rectangle-iterative algorithm. In: Chakravorthy S, Alfa AS (eds) *Matrix-analytic Methods in Stochastic Models*, Marcel Dekker, pp 207–224
- Telek M (2000) The minimal coefficient of variation of discrete phase type distributions. 3rd Conf on MAM pp 391–400
- Wolff RW (1989) *Stochastic Modelling and the Theory of Queues*. Prentice hall, Engelwood Cliffs, New Jersey



<http://www.springer.com/978-1-4939-3418-8>

Applied Discrete-Time Queues

Alfa, A.S.

2016, XXI, 383 p. 26 illus., Hardcover

ISBN: 978-1-4939-3418-8