Manufacturing Systems Modeling and Analysis

Guy L. Curry · Richard M. Feldman

# Manufacturing Systems Modeling and Analysis

Springer

Prof. Guy L. Curry
Texas A & M University
Dept. Industrial & Systems
Engineering
3131 TAMU
College Station TX
77843-3131
USA
g-curry@tamu.edu

Prof. Richard M. Feldman
Texas A & M University
Dept. Industrial & Systems
Engineering
3131 TAMU
College Station TX
77843-3131
USA
richf@tamu.edu

ISBN: 978-3-540-88762-1 e-ISBN: 978-3-540-88763-8

DOI 10.1007/978-3-540-88763-8

Library of Congress Control Number: 2008939370



*Cover design:* eStudio Calamar S.L.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

*This book is dedicated to the two individuals who keep us going, tolerate our work ethic, and make life a wondrous journey, our wives: Jerrie Curry and Alice Feldman.*

# Preface

This textbook was developed to fill the need for an accessible but comprehensive presentation of the analytical approaches for modeling and analyzing models of manufacturing and production systems. It is an out growth of the efforts within the Industrial and Systems Engineering Department at Texas A&M to develop and teach an analytically based undergraduate course on probabilistic modeling of manufacturing type systems. The level of this textbook is directed at undergraduate and masters students in engineering and mathematical sciences. The only prerequisite for students using this textbook is a previous course covering calculus-based probability and statistics. The underlying methodology is queueing theory, and we shall develop the basic concepts in queueing theory in sufficient detail that the reader need not have previously covered it. Queueing theory is a well-established discipline dating back to the early 1900's work of A. K. Erlang, a Danish mathematician, on telephone traffic congestion. Although there are many textbooks on queueing theory, these texts are generally oriented to the methodological development of the field and exact results and not to the practical application of using approximations in realistic modeling situations. The application of queueing theory to manufacturing type systems started with the approximation based work of Ward Whitt in the 1980's. His paper on QNA (a queueing network analyzer) in 1983 is the base from which most applied modeling efforts have evolved.

There are several textbooks with titles similar to this book. Principle among these are: Modeling and Analysis of Manufacturing Systems by Askin and Standridge, Manufacturing Systems Engineering by Stanley Gershwin, Queueing Theory in Manufacturing Systems Analysis and Design by Papadopoulos, Heavey and Browne, Performance Analysis of Manufacturing Systems by Tayfur Altiok, Stochastic Modeling and Analysis of Manufacturing Systems, edited by David Yao, and Stochastic Models of Manufacturing Systems by Buzacott and Shanthikumar. Each of these texts, along with several others, contributes greatly to the field. The book that most closely aligns with the motivation, level, and intent of this book is Factory Physics by Hopp and Spearman. Their approach and analysis is highly recommended reading, however, their book's scope is on the larger field of produc-

tion and operations management. Thus, it does not provide the depth and breath of analytical modeling procedures that are presented in this text.

This text is about the development of analytical approximation models and their use in evaluating factory performance. The tools needed for the analytical approach are fully developed. One useful non-analytical tool that is not fully developed in this textbook is simulation modeling. In practice as well as in the development of the models in this text, simulation is extensively used as a verification tool. Even though the development of simulation models is only modestly addressed, we would encourage instructors who use this book in their curriculum after a simulation course to ask students to simulate some of the homework problems so that a comparison can be made of the analysis using the models presented here with simulation models. By developing simulation models students will have a better understanding of the modeling assumptions and the accuracy of the analytical approximations. In addition several chapters include an appendix that contains instructions in the use of Microsoft Excel as an aid in modeling or in building simple simulation models.

Two special sections are included to help the reader organize the many concepts contained in the text. Immediately after the Table of Contents, we have included a symbol table that contains most of the notation used throughout the text. Second, immediately after the final chapter a glossary of terms is included that summarizes the various definitions used. It is expected that these will prove valuable resources as the reader progresses through the text.

Many individuals have contributed to this book through our interactions in research efforts and discussions. Special thanks go to Professor Martin A. Wortman, Texas A&M University, who designed and taught the first presentation of the course for which this book was originally developed and Professor Bryan L. Deuermeyer, Texas A&M University, for his significant contributions to our joint research activities in this area and his continued interest and criticism. In addition several individuals have helped in improving the text by using a draft copy while teaching the material to undergraduates including Eylem Tekin at Texas A&M, Natarajan Gautam also at Texas A&M, and Kevin Gue at Auburn University. We also wish to acknowledge the contributions of Professors John A. Fowler, Arizona State University, and Mark L. Spearman, Factory Physics, Inc., for their continued interactions and discussions on modeling manufacturing systems. And we thank Ciriaco Valdez-Flores, a co-author of the first chapter covering basic probability for permission to include it as part of our book. Finally, we acknowledge our thanks through the words of the psalmist, "Give thanks to the Lord, for He is good; His love endures forever." (Psalms 107:1, NIV)

College Station, Texas
December 2008

*Guy L. Curry*
*Richard M. Feldman*

# Contents

# Symbols

| | |
|---|---|
| $\boldsymbol{\alpha}$ | Used in Chap. 9 as the row vector of initial probabilities associated with a phase type distribution. |
| $\alpha_k$ | In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of inter-arrival times into Subsystem $k$. |
| $\beta_k$ | In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of inter-arrival times into Subsystem $k$. |
| $\boldsymbol{\gamma}$ | Vector of mean arrival rates to the various workstations from an external source. |
| $\boldsymbol{\gamma}^i$ | Vector of mean arrival rates of Type $i$ jobs entering the various workstations from an external source. |
| $\gamma_{i,k}$ | Mean rate of Type $i$ jobs into Workstation $k$ from an external source. |
| $\widetilde{\gamma}_\ell^i$ | Mean rate of Type $i$ jobs to the $\ell^{th}$ step of the production plan from an external source (Property 6.5). |
| $\gamma_k$ | Mean rate of jobs arriving from an external source to Workstation $k$. In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of service times for Subsystem $k$. |
| $\lambda$ | Mean arrival rate. |
| $\boldsymbol{\lambda}$ | Vector of mean arrival rates into the various workstations. |
| $\lambda(B)$ | Mean arrival rate of batches of jobs. |
| $\lambda_e$ | The effective mean arrival rate (Def. 3.1). |
| $\boldsymbol{\lambda}^i$ | Vector of arrival rates of Type $i$ jobs entering the various workstations. |
| $\lambda(I)$ | Mean arrival rate of individual jobs. |
| $\lambda_{i,k}$ | Mean arrival rate of Type $i$ jobs entering Workstation $k$. |
| $\widetilde{\lambda}_{i,\ell}$ | Mean arrival rate of Type $i$ jobs to the $\ell^{th}$ step of the production plan (Property 6.5). |
| $\lambda_k$ | Mean arrival rate into Workstation $k$. |
| $\mu$ | Mean service rate (the reciprocal of the mean service time). |
| $\mu_k$ | Often used as the mean service rate for Workstation $k$. In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of service times for Subsystem $k$. |

| | |
|---|---|
| $\nu_i$ | Number of steps within the production plan for a Type $i$ job (Def. 6.3). (Not to be confused with the letter $\nu$ used in Chap. 9.) |
| $a$ | Availability (Def. 4.2). |
| $c_k$ | The number of (identical) machines at Workstation $k$. |
| $C^2$ | Squared coefficient of variation which is the variance divided by the mean squared. |
| $C_a^2$ | Squared coefficient of variation of inter-arrival times. |
| $\mathbf{c}_a^2$ | A vector of the squared coefficients of variation of the inter-arrival times to the various workstations. |
| $C_a^2(B)$ | Squared coefficient of variation of the inter-arrival times of batches of jobs. |
| $C_a^2(I)$ | Squared coefficient of variation of the inter-arrival times of individual jobs. |
| $C_a^2(k)$ | Squared coefficient of variation of the stream of inter-arrival times entering Workstation $k$. |
| $C_a^2(k,j)$ | Squared coefficient of variation of the inter-arrival times into Workstation $j$ that come from Workstation $k$. If $k=0$, it refers to externally arriving jobs into Workstation $j$. |
| $C_d^2(k)$ | The squared coefficient of variation of the inter-departure times from Workstation $k$. |
| $C_s^2$ | Squared coefficient of variation of service times. |
| $C_s^2(B)$ | Squared coefficient of variation of the service times of batches of jobs. |
| $C_s^2(I)$ | Squared coefficient of variation of the service times of individual jobs. |
| $C_s^2(k)$ | Squared coefficient of variation of service times for an arbitrary job at Workstation $k$. |
| $C_s^2(i,k)$ | Squared coefficient of variation of service times for Type $i$ jobs at Workstation $k$. |
| $CT$ | Mean cycle time (Def. 2.1). |
| $CT_q(k)$ | Mean cycle time within the queue of Workstation $k$. |
| $CT_s$ | Mean cycle time for the system which includes all time spent within the factory. |
| $CT_s^i$ | Mean cycle time of a Type $i$ job for the system which includes all time spent within the factory. |
| $CT(i,k)$ | Mean cycle time within Workstation $k$ for a Type $i$ job including the time spent in the queue plus the time spent processing. |
| $CT(k)$ | Mean cycle time within Workstation $k$ including the time spent in the queue plus the time spent processing. |
| $CT_k(\cdot)$ | Mean cycle time at Workstation $k$ as a function of the CONWIP level. |
| $E$ | Expectation operator or the mean. |
| $F$ | Random variable denoting the time to failure. |
| $G$ | Used in Chap. 9 for a generator matrix usually associated with a $GE_2$ or an $MGE$ distribution. |
| $i$ | A general index. Starting with Chap. 6, it is most often used to indicate a job type. |
| $I(\cdot,\cdot)$ | An indicator function or identity matrix (Def. 6.4). |

| | |
|---|---|
| $k$ | A general index. Starting with Chap. 6, it is most often used to indicate a workstation, and in Chap. 7 it is also used for batch size. |
| $\ell$ | A general index. Most often used to denote the $\ell^{th}$ step of a production plan. In Chap. 8, it is sometimes used to indicate job type. |
| $m$ | Most often used for the total number of job types. |
| $n$ | Most often used for the total number of workstations. |
| $N$ | In Chap. 7, it is a random variable denoting batch size. |
| $P = (p_{j,k})$ | Routing matrix (Def. 5.2). |
| $P^i = (p^i_{j,k})$ | Routing matrix of Type $i$ jobs. |
| $\widetilde{P}^i = (\widetilde{p}^i_{\ell,j})$ | Step-wise routing matrix for Type $i$ jobs (Def. 6.3). |
| $p^F_{a,n}$ | In Chap. 9, the probability that an arrival to the $n^{th}$ (or final) subsystem, finds the subsystem full. |
| $p^{(i,F)}_{a,k}$ | In Chap. 9, the probability that an arrival to Subsystem $k$, for $k < n$, finds the subsystem full and the service-machine in Phase $i$. |
| $p^0_{d,1}$ | In Chap. 9, the probability that a departure from Subsystem 1 leaves the subsystem empty. |
| $p^{(i,0)}_{d,k}$ | In Chap. 9, the probability that a departure from Subsystem $k$, for $k > 1$, leaves the subsystem empty and the arrival-machine in Phase $i$. |
| $p_k$ | Often used for the steady-state probability of $k$ jobs being within a system. In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of inter-arrival times into Subsystem $k$. |
| $p_k(j,w)$ | The steady-state probability that there are $j$ jobs at Workstation $k$ when the CONWIP level for the factory is set to $w$. |
| $Q$ | Used in Chap. 9 for a generator matrix usually associated with finding the steady-state probabilities of two-node subsystems. |
| $q_k$ | In Chap. 9, it is used as a parameter for the $GE_2$ distribution that approximates the distribution of service times for Subsystem $k$. |
| $R$ | Random variable denoting repair time, except in Chap. 7 where it is the random variable denoting the setup time for a batch. |
| $r_k$ | The relative arrival rate into Workstation $k$. |
| $T_e$ | Random variable denoting the effective service time (Def. 4.1). |
| $T_a(B)$ | Random variable denoting inter-arrival times of batches of jobs. |
| $T_a(I)$ | Random variable denoting inter-arrival times of individual jobs. |
| $T_s(B)$ | Random variable denoting service times of batches of jobs. |
| $T_s(I)$ | Random variable denoting service times of individual jobs. |
| $T_s(i,k)$ | Random variable denoting service times for a Type $i$ job in Workstation $k$. |
| $T_s(k)$ | Random variable denoting service times for an arbitrary job in Workstation $k$. |
| $th$ | Mean throughput rate (Def. 2.3). |
| $th(k)$ | Mean throughput rate for Workstation $k$. |
| $u$ | Machine utilization. |
| $u_k$ | Utilization factor for Workstation $k$ (Eq. (6.2). |

| | |
|---|---|
| $u_k(\cdot)$ | Utilization factor at Workstation $k$ as a function of the CONWIP level. |
| $V$ | The variance which also equals the second moment minus the mean squared. |
| $\mathbf{v}$ | In Chap. 9, a vector of steady-state probabilities derived for a generator matrix. (Not to be confused with the Greek letter $\nu$ used in Chap. 6.) |
| $v_i$ | In Chap. 9, the steady-state probability of being in State $i$. (Not to be confused with the Greek letter $\nu$ used in Chap. 6.) |
| $w$ | Used in Chaps. 8 and 9 as a variable for functions whose independent variable represents work-in-process. |
| $\mathbf{w}$ | A vector of dimension $m$, where $m$ is the number of job types, giving the CONWIP limits for each job type. |
| $w_{\max}$ | In Chaps. 8 and 9 constant indicated a maximum limit placed on work-in-process. |
| $\widetilde{w}^i(\cdot)$ | The workstation mapping function (Def. 6.2). |
| $WIP$ | Mean (time-averaged) work-in-process (Def. 2.2). |
| $WIP_q(k)$ | Mean (time-averaged) work-in-process for the queue of Workstation $k$. |
| $WIP_s$ | Mean (time-averaged) work-in-process within the system which includes all jobs within the factory. |
| $WIP(k)$ | Mean (time-averaged) work-in-process within Workstation $k$ including jobs in the queue and job(s) within the processor. |
| $WIP_k(\cdot)$ | Mean (time-averaged) work-in-process at Workstation $k$ as a function of the CONWIP level. |
| $WL_k$ | Workload at Workstation $k$ (Def. 6.1 and Eq. (6.1)). |

# Chapter 1
# Basic Probability Review

The background material for this textbook is a general understanding of probability and the properties of various distributions; thus, before discussing the modeling of the various manufacturing and production systems, it is important to review the fundamental concepts of basic probability. This material is not intended to teach probability theory, but it is used for review and to establish a common ground for the notation and definitions used throughout the book. Much of the material in this chapter is taken from [3], and for those already familiar with probability, this chapter can easily be skipped.

## 1.1 Basic Definitions

To understand probability , it is best to envision an experiment for which the outcome (result) is unknown. All possible outcomes must be defined and the collection of these outcomes is called the sample space. Probabilities are assigned to subsets of the sample space, called events. We shall give the rigorous definition for probability. However, the reader should not be discouraged if an intuitive understanding is not immediately acquired. This takes time and the best way to understand probability is by working problems.

**Definition 1.1.** An element of a *sample space* is an *outcome*. A set of outcomes, or equivalently a subset of the sample space, is called an *event*.

**Definition 1.2.** A *probability space* is a three-tuple $(\Omega, \mathscr{F}, \Pr)$ where $\Omega$ is a sample space, $\mathscr{F}$ is a collection of events from the sample space, and Pr is a probability measure that assigns a number to each event contained in $\mathscr{F}$. Furthermore, Pr must satisfy the following conditions, for each event $A, B$ within $\mathscr{F}$:

- $\Pr(\Omega) = 1$,
- $\Pr(A) \geq 0$,
- $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ if $A \cap B = \phi$, where $\phi$ denotes the empty set,

- $\Pr(A^c) = 1 - \Pr(A)$, where $A^c$ is the complement of $A$.

It should be noted that the collection of events, $\mathscr{F}$, in the definition of a probability space must satisfy some technical mathematical conditions that are not discussed in this text. If the sample space contains a finite number of elements, then $\mathscr{F}$ usually consists of all the possible subsets of the sample space. The four conditions on the probability measure Pr should appeal to one's intuitive concept of probability. The first condition indicates that something from the sample space must happen, the second condition indicates that negative probabilities are illegal, the third condition indicates that the probability of the union of two disjoint (or mutually exclusive) events is the sum of their individual probabilities and the fourth condition indicates that the probability of an event is equal to one minus the probability of its complement (all other events). The fourth condition is actually redundant but it is listed in the definitions because of its usefulness.

A probability space is the full description of an experiment; however, it is not always necessary to work with the entire space. One possible reason for working within a restricted space is because certain facts about the experiment are already known. For example, suppose a dispatcher at a refinery has just sent a barge containing jet fuel to a terminal 800 miles down river. Personnel at the terminal would like a prediction on when the fuel will arrive. The experiment consists of all possible weather, river, and barge conditions that would affect the travel time down river. However, when the dispatcher looks outside it is raining. Thus, the original probability space can be restricted to include only rainy conditions. Probabilities thus restricted are called conditional probabilities according to the following definition.

**Definition 1.3.** Let $(\Omega, \mathscr{F}, \Pr)$ be a probability space where $A$ and $B$ are events in $\mathscr{F}$ with $\Pr(B) \neq 0$. The *conditional probability* of $A$ given $B$, denoted $\Pr(A|B)$, is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} .$$

Venn diagrams are sometimes used to illustrate relationships among sets. In the diagram of Fig. 1.1, assume that the probability of a set is proportional to its area. Then the value of $\Pr(A|B)$ is the proportion of the area of set $B$ that is occupied by the set $A \cap B$.

*Example 1.1.* A telephone manufacturing company makes radio phones and plain phones and ships them in boxes of two (same type in a box). Periodically, a quality control technician randomly selects a shipping box, records the type of phone in the box (radio or plain), and then tests the phones and records the number that were defective. The sample space is

$$\Omega = \{(r,0),(r,1),(r,2),(p,0),(p,1),(p,2)\} ,$$

where each outcome is an ordered pair; the first component indicates whether the phones in the box are the radio type or plain type and the second component gives the number of defective phones. The set $\mathscr{F}$ is the set of all subsets, namely,

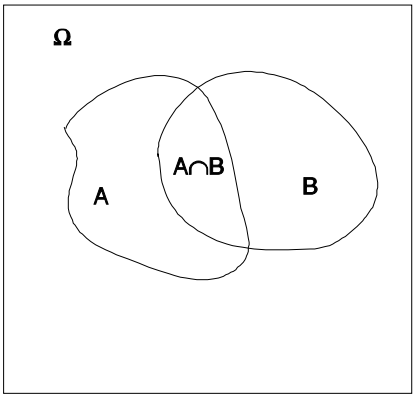


**Fig. 1.1** Venn diagram illustrating events $A$, $B$, and $A \cap B$

$$\mathscr{F} = \{\phi, \{(r,0)\}, \{(r,1)\}, \{(r,0),(r,1)\}, \cdots, \Omega\} .$$

There are many legitimate probability laws that could be associated with this space. One possibility is

$$\begin{aligned} &\Pr\{(r,0)\} = 0.45 , \;\; \Pr\{(p,0)\} = 0.37 , \\ &\Pr\{(r,1)\} = 0.07 , \;\; \Pr\{(p,1)\} = 0.08 , \\ &\Pr\{(r,2)\} = 0.01 , \;\; \Pr\{(p,2)\} = 0.02 . \end{aligned}$$

By using the last property in Definition 1.2, the probability measure can be extended to all events; for example, the probability that a box is selected that contains radio phones and at most one phone is defective is given by

$$\Pr\{(r,0),(r,1)\} = 0.52 .$$

Now let us assume that a box has been selected and opened. We observe that the two phones within the box are radio phones, but no test has yet been made on whether or not the phones are defective. To determine the probability that at most one phone is defective in the box containing radio phones, define the event $A$ to be the set $\{(r,0),(r,1),(p,0),(p,1)\}$ and the event $B$ to be $\{(r,0),(r,1),(r,2)\}$. In other words, $A$ is the event of having at most one defective phone, and $B$ is the event of having a box of radio phones. The probability statement can now be written as

$$\Pr\{A|B\} = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr\{(r,0),(r,1)\}}{\Pr\{(r,0),(r,1),(r,2)\}} = \frac{0.52}{0.53} = 0.991 .$$

□

- *Suggestion: Do Problems 1.1–1.2 and 1.20.*

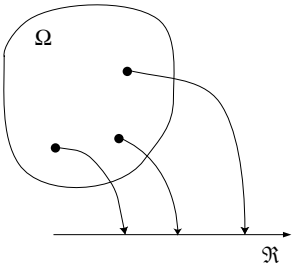


**Fig. 1.2** A random variable is a mapping from the sample space to the real numbers

## 1.2 Random Variables and Distribution Functions

It is often cumbersome to work with the outcomes directly in mathematical terms. Random variables are defined to facilitate the use of mathematical expressions and to focus only on the outcomes of interest.

**Definition 1.4.** A *random variable* is a function that assigns a real number to each outcome in the sample space.

Figure 1.2 presents a schematic illustrating a random variable. The name "random variable" is actually a misnomer, since it is not random and is not a variable. As illustrated in the figure, the random variable simply maps each point (outcome) in the sample space to a number on the real line[1].

Revisiting Example 1.1, let us assume that management is primarily interested in whether or not at least one defective phone is in a shipping box. In such a case a random variable $D$ might be defined such that it is equal to zero if all the phones within a box are good and equal to 1 otherwise; that is,

$$\begin{aligned} D(r,0) &= 0\,, \;\; D(p,0) = 0\,, \\ D(r,1) &= 1\,, \;\; D(p,1) = 1\,, \\ D(r,2) &= 1\,, \;\; D(p,2) = 1\,. \end{aligned}$$

The set $\{D = 0\}$ refers to the set of all outcomes for which $D = 0$ and a legitimate probability statement would be

$$\Pr\{D = 0\} = \Pr\{(r,0),(p,0)\} = 0.82\,.$$

To aid in the recognition of random variables, the notational convention of using only capital Roman letters (or possibly Greek letters) for random variables is followed. Thus, if you see a lower case Roman letter, you know immediately that it can not be a random variable.

[1] Technically, the space into which the random variable maps the sample space may be more general than the real number line, but for our purposes, the real numbers will be sufficient.

Random variables are either discrete or continuous depending on their possible values. If the possible values can be counted, the random variable is called discrete; otherwise, it is called continuous. The random variable $D$ defined in the previous example is discrete. To give an example of a continuous random variable, define $T$ to be a random variable that represents the length of time that it takes to test the phones within a shipping box. The range of possible values for $T$ is the set of all positive real numbers, and thus $T$ is a continuous random variable.

A cumulative distribution function (CDF) is often used to describe the probability measure underlying the random variable. The cumulative distribution function (usually denoted by a capital Roman letter or a Greek letter) gives the probability accumulated up to and including the point at which it is evaluated.

**Definition 1.5.** The function $F$ is the *cumulative distribution function* for the random variable $X$ if

$$F(a) = \Pr\{X \leq a\}$$

for all real numbers $a$.

The CDF for the random variable $D$ defined above is

$$F(a) = \begin{cases} 0 & \text{for } a < 0 \\ 0.82 & \text{for } 0 \leq a < 1 \\ 1.0 & \text{for } a \geq 1\,. \end{cases} \tag{1.1}$$

Figure 1.3 gives the graphical representation for $F$. The random variable $T$ defined to represent the testing time for phones within a randomly chosen box is continuous and there are many possibilities for its probability measure since we have not yet defined its probability space. As an example, the function $G$ (see Fig. 1.10) is the cumulative distribution function describing the randomness that might be associated with $T$:

$$G(a) = \begin{cases} 0 & \text{for } a < 0 \\ 1 - \mathrm{e}^{-2a} & \text{for } a \geq 0\,. \end{cases} \tag{1.2}$$

**Property 1.1.** *A cumulative distribution function $F$ has the following properties:*

- $\lim_{a \to -\infty} F(a) = 0$,
- $\lim_{a \to +\infty} F(a) = 1$,
- $F(a) \leq F(b)$ *if* $a < b$,
- $\lim_{a \to b^+} F(a) = F(b)$.

The first and second properties indicate that the graph of the cumulative distribution function always begins on the left at zero and limits to one on the right. The third property indicates that the function is nondecreasing. The fourth property indicates that the cumulative distribution function is right-continuous. Since the distribution function is monotone increasing, at each discontinuity the function value is defined

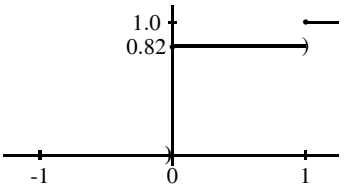


**Fig. 1.3** Cumulative distribution function for Eq. (1.1) for the discrete random variable $D$

by the larger of two limits: the limit value approaching the point from the left and the limit value approaching the point from the right.

It is possible to describe the random nature of a discrete random variable by indicating the size of jumps in its cumulative distribution function. Such a function is called a probability mass function (denoted by a lower case letter) and gives the probability of a particular value occurring.

**Definition 1.6.** The function $f$ is the *probability mass function* (pmf) of the discrete random variable $X$ if

$$f(k) = \Pr\{X = k\}$$

for every $k$ in the range of $X$.

If the pmf is known, then the cumulative distribution function is easily found by

$$\Pr\{X \leq a\} = F(a) = \sum_{k \leq a} f(k)\,. \tag{1.3}$$

The situation for a continuous random variable is not quite as easy because the probability that any single given point occurs must be zero. Thus, we talk about the probability of an interval occurring. With this in mind, it is clear that a mass function is inappropriate for continuous random variables; instead, a probability density function (denoted by a lower case letter) is used.

**Definition 1.7.** The function $g$ is called the *probability density function* (pdf) of the continuous random variable $Y$ if

$$\int_a^b g(u)\mathrm{d}u = \Pr\{a \leq Y \leq b\}$$

for all $a, b$ in the range of $Y$.

From Definition 1.7 it should be seen that the pdf is the derivative of the cumulative distribution function and

$$G(a) = \int_{-\infty}^{a} g(u)\mathrm{d}u\,. \tag{1.4}$$

The cumulative distribution functions for the example random variables $D$ and $T$ are defined in Eqs. (1.1 and 1.2). We complete that example by giving the pmf for $D$ and the pdf for $T$ as follows:

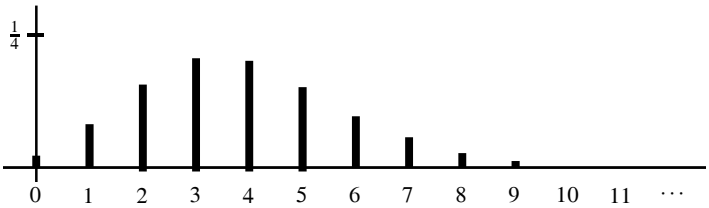


**Fig. 1.4** The Poisson probability mass function of Example 1.2

$$f(k) = \begin{cases} 0.82 & \text{for } k = 0 \\ 0.18 & \text{for } k = 1 \end{cases}. \tag{1.5}$$

and

$$g(a) = \begin{cases} 2\mathrm{e}^{-2a} & \text{for } a \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{1.6}$$

*Example 1.2.* Discrete random variables need not have finite ranges. A classical example of a discrete random variable with an infinite range is due to Rutherford, Chadwick, and Ellis from 1920 [7, pp. 209–210]. An experiment was performed to determine the number of $\alpha$-particles emitted by a radioactive substance in 7.5 seconds. The radioactive substance was chosen to have a long half-life so that the emission rate would be constant. After 2608 experiments, it was found that the number of emissions in 7.5 seconds was a random variable, $N$, whose pmf could be described by

$$\Pr\{N = k\} = \frac{(3.87)^k \mathrm{e}^{-3.87}}{k!} \quad \text{for } k = 0, 1, \cdots.$$

It is seen that the discrete random variable $N$ has a countably infinite range and the infinite sum of its pmf equals one. In fact, this distribution is fairly important and will be discussed later under the heading of the Poisson distribution. Figure 1.4 shows its pmf graphically. □

The notion of independence is very important when dealing with more than one random variable. Although we shall postpone the discussion on multivariate distribution functions until Sect. 1.5, we introduce the concept of independence at this point.

**Definition 1.8.** The random variables $X_1, \cdots, X_n$ are *independent* if

$$\Pr\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = \Pr\{X_1 \leq x_1\} \times \cdots \times \Pr\{X_n \leq x_n\}$$

for all possible values of $x_1, \cdots, x_n$.

Conceptually, random variables are independent if knowledge of one (or more) random variable does not "help" in making probability statements about the other random variables. Thus, an alternative definition of independence could be made using conditional probabilities (see Definition 1.3) where the random variables $X_1$

and $X_2$ are called independent if $\Pr\{X_1 \le x_1 | X_2 \le x_2\} = \Pr\{X_1 \le x_1\}$ for all values of $x_1$ and $x_2$.

For example, suppose that $T$ is a random variable denoting the length of time it takes for a barge to travel from a refinery to a terminal 800 miles down river, and $R$ is a random variable equal to 1 if the river condition is smooth when the barge leaves and 0 if the river condition is not smooth. After collecting data to estimate the probability laws governing $T$ and $R$, we would not expect the two random variables to be independent since knowledge of the river conditions would help in determining the length of travel time.

One advantage of independence is that it is easier to obtain the distribution for sums of random variables when they are independent than when they are not independent. When the random variables are continuous, the pdf of the sum involves an integral called a *convolution*.

**Property 1.2.** *Let $X_1$ and $X_2$ be independent continuous random variables with pdf's given by $f_1(\cdot)$ and $f_2(\cdot)$. Let $Y = X_1 + X_2$, and let $h(\cdot)$ be the pdf for $Y$. The pdf for $Y$ can be written, for all $y$, as*

$$h(y) = \int_{-\infty}^{\infty} f_1(y-x) f_2(x) \mathrm{d}x \, .$$

*Furthermore, if $X_1$ and $X_2$ are both nonnegative random variables, then*

$$h(y) = \int_{0}^{y} f_1(y-x) f_2(x) \mathrm{d}x \, .$$

*Example 1.3.* Our electronic equipment is highly sensitive to voltage fluctuations in the power supply so we have collected data to estimate when these fluctuations occur. After much study, it has been determined that the time between voltage spikes is a random variable with pdf given by (1.6), where the unit of time is hours. Furthermore, it has been determined that the random variables describing the time between two successive voltage spikes are independent. We have just turned the equipment on and would like to know the probability that within the next 30 minutes at least two spikes will occur.

Let $X_1$ denote the time interval from when the equipment is turned on until the first voltage spike occurs, and let $X_2$ denote the time interval from when the first spike occurs until the second occurs. The question of interest is to find $\Pr\{Y \le 0.5\}$, where $Y = X_1 + X_2$. Let the pdf for $Y$ be denoted by $h(\cdot)$. Property 1.2 yields

$$\begin{aligned} h(y) &= \int_0^y 4\mathrm{e}^{-2(y-x)}\mathrm{e}^{-2x}\mathrm{d}x \\ &= 4\mathrm{e}^{-2y}\int_0^y \mathrm{d}x = 4y\mathrm{e}^{-2y} \, , \end{aligned}$$

for $y \ge 0$. The pdf of $Y$ is now used to answer our question, namely,

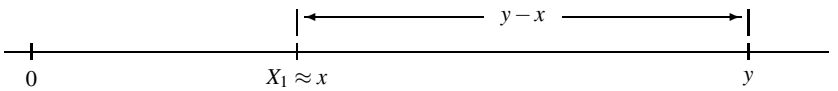


**Fig. 1.5** Time line illustrating the convolution

$$\Pr\{Y \leq 0.5\} = \int_0^{0.5} h(y)\mathrm{d}y = \int_0^{0.5} 4y\mathrm{e}^{-2y}\mathrm{d}y = 0.264\,.$$

□

It is also interesting to note that the convolution can be used to give the cumulative distribution function if the first pdf in the above property is replaced by the CDF; in other words, for *nonnegative* random variables we have

$$H(y) = \int_0^y F_1(y-x)f_2(x)\mathrm{d}x\,. \tag{1.7}$$

Applying (1.7) to our voltage fluctuation question yields

$$\Pr\{Y \leq 0.5\} \equiv H(0.5) = \int_0^{0.5} (1-\mathrm{e}^{-2(0.5-x)})2\mathrm{e}^{-2x}\mathrm{d}x = 0.264\,.$$

We rewrite the convolution of Eq. (1.7) slightly to help in obtaining an intuitive understanding of why the convolution is used for sums. Again, assume that $X_1$ and $X_2$ are independent, nonnegative random variables with pdf's $f_1$ and $f_2$, then

$$\Pr\{X_1 + X_2 \leq y\} = \int_0^y F_2(y-x)f_1(x)\mathrm{d}x\,.$$

The interpretation of $f_1(x)dx$ is that it represents the probability that the random variable $X_1$ falls in the interval $(x, x+dx)$ or, equivalently, that $X_1$ is approximately $x$. Now consider the time line in Fig. 1.5. For the sum to be less than $y$, two events must occur: first, $X_1$ must be some value (call it $x$) that is less than $y$; second, $X_2$ must be less than the remaining time that is $y-x$. The probability of the first event is approximately $f_1(x)dx$, and the probability of the second event is $F_2(y-x)$. Since the two events are independent, they are multiplied together; and since the value of $x$ can be any number between 0 and $y$, the integral is from 0 to $y$.

- *Suggestion: Do Problems 1.3–1.6.*

## 1.3 Mean and Variance

Many random variables have complicated distribution functions and it is therefore difficult to obtain an intuitive understanding of the behavior of the random variable by simply knowing the distribution function. Two measures, the mean and variance, are defined to aid in describing the randomness of a random variable. The mean equals the arithmetic average of infinitely many observations of the random variable and the variance is an indication of the variability of the random variable. To illustrate this concept we use the square root of the variance which is called the *standard deviation*. In the 19*th* century, the Russian mathematician P. L. Chebyshev showed that for any given distribution, *at least* 75% of the time the observed value of a random variable will be within two standard deviations of its mean and *at least* 93.75% of the time the observed value will be within four standard deviations of the mean. These are general statements, and specific distributions will give much tighter bounds. (For example, a commonly used distribution is the normal "bell shaped" distribution. With the normal distribution, there is a 95.44% probability of being within two standard deviations of the mean.) Both the mean and variance are defined in terms of the expected value operator, that we now define.

**Definition 1.9.** Let $h$ be a function defined on the real numbers and let $X$ be a random variable. The *expected value* of $h(X)$ is given, for $X$ discrete, by

$$E[h(X)] = \sum_k h(k) f(k)$$

where $f$ is its pmf, and for $X$ continuous, by

$$E[h(X)] = \int_{-\infty}^{\infty} h(s) f(s) \mathrm{d}s$$

where $f$ is its pdf.

*Example 1.4.* A supplier sells eggs by the carton containing 144 eggs. There is a small probability that some eggs will be broken and he refunds money based on broken eggs. We let $B$ be a random variable indicating the number of broken eggs per carton with a pmf given by

| $k$ | $f(k)$ |
|---|---|
| 0 | 0.779 |
| 1 | 0.195 |
| 2 | 0.024 |
| 3 | 0.002 |

.

A carton sells for $4.00, but a refund of 5 cents is made for each broken egg. To determine the expected income per carton, we define the function $h$ as follows

| $k$ | $h(k)$ |
|---|---|
| 0 | 4.00 |
| 1 | 3.95 |
| 2 | 3.90 |
| 3 | 3.85 |

.

Thus, $h(k)$ is the net revenue obtained when a carton is sold containing $k$ broken eggs. Since it is not known ahead of time how many eggs are broken, we are interested in determining the *expected* net revenue for a carton of eggs. Definition 1.9 yields

$$\begin{aligned} E[h(B)] &= 4.00 \times 0.779 + 3.95 \times 0.195 \\ &\quad + 3.90 \times 0.024 + 3.85 \times 0.002 = 3.98755\,. \end{aligned}$$

□

The expected value operator is a linear operator, and it is not difficult to show the following property.

**Property 1.3.** *Let $X$ and $Y$ be two random variables with $c$ being a constant, then*

- $E[c] = c$,
- $E[cX] = cE[X]$,
- $E[X+Y] = E[X] + E[Y]$.

In the egg example since the cost per broken egg is a constant $(c = 0.05)$, the expected revenue per carton could be computed as

$$\begin{aligned} E[4.0 - 0.05B] &= 4.0 - 0.05E[B] \\ &= 4.0 - 0.05\,(\,0 \times 0.779 + 1 \times 0.195 + 2 \times 0.024 + 3 \times 0.002\,) \\ &= 3.98755\,. \end{aligned}$$

The expected value operator provides us with the procedure to determine the mean and variance.

**Definition 1.10.** The *mean*, $\mu$ or $E[X]$, and *variance*,$\sigma^2$ or $V[X]$, of a random variable $X$ are defined as

$$\mu = E[X], \quad \sigma^2 = E[(X-\mu)^2]\,,$$

respectively. The *standard deviation* is the square root of the variance.

**Property 1.4.** *The following are often helpful as computational aids:*

- $V[X] = \sigma^2 = E[X^2] - \mu^2$
- $V[cX] = c^2 V[X]$
- *If* $X \geq 0$, $E[X] = \int_0^\infty [1 - F(s)]\mathrm{d}s$ *where* $F(x) = \Pr\{X \leq x\}$
- *If* $X \geq 0$, *then* $E[X^2] = 2\int_0^\infty s[1 - F(s)]\mathrm{d}s$ *where* $F(x) = \Pr\{X \leq x\}$.

*Example 1.5.* The mean and variance calculations for a discrete random variable can be easily illustrated by defining the random variable $N$ to be the number of defective phones within a randomly chosen box from Example 1.1. In other words, $N$ has the pmf given by

$$\Pr\{N = k\} = \begin{cases} 0.82 & \text{for } k = 0 \\ 0.15 & \text{for } k = 1 \\ 0.03 & \text{for } k = 2\,. \end{cases}$$

The mean and variance is, therefore, given by

$$\begin{aligned} E[N] &= 0 \times 0.82 + 1 \times 0.15 + 2 \times 0.03 \\ &= 0.21, \end{aligned}$$

$$\begin{aligned} V[N] &= (0 - 0.21)^2 \times 0.82 + (1 - 0.21)^2 \times 0.15 + (2 - 0.21)^2 \times 0.03 \\ &= 0.2259\,. \end{aligned}$$

Or, an easier calculation for the variance (Property 1.4) is

$$\begin{aligned} E[N^2] &= 0^2 \times 0.82 + 1^2 \times 0.15 + 2^2 \times 0.03 \\ &= 0.27 \end{aligned}$$

$$\begin{aligned} V[N] &= 0.27 - 0.21^2 \\ &= 0.2259\,. \end{aligned}$$

□

*Example 1.6.* The mean and variance calculations for a continuous random variable can be illustrated with the random variable $T$ whose pdf was given by Eq. 1.6. The mean and variance is therefore given by

$$E[T] = \int_0^\infty 2s\mathrm{e}^{-2s}\mathrm{d}s = 0.5\,,$$

$$V[T] = \int_0^\infty 2(s - 0.5)^2\mathrm{e}^{-2s}\mathrm{d}s = 0.25\,.$$

Or, an easier calculation for the variance (Property 1.4) is

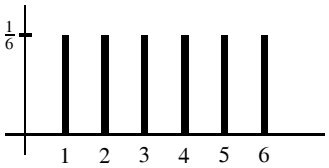


**Fig. 1.6** A discrete uniform probability mass function

$$E[T^2] = \int_0^\infty 2s^2 e^{-2s} ds = 0.5 \,,$$

$$V[T] = 0.5 - 0.5^2 = 0.25 \,.$$

□

The final definition in this section is used often as a descriptive statistic to give an intuitive feel for the variability of processes.

**Definition 1.11.** The *squared coefficient of variation*, $C^2$, of a nonnegative random variable $T$ is the ratio of the the variance to the mean squared; that is,

$$C^2[T] = \frac{V[T]}{E[T]^2} \,.$$

- *Suggestion: Do Problems 1.7–1.14.*

## 1.4 Important Distributions

There are many distribution functions that are used so frequently that they have become known by special names. In this section, some of the major distribution functions are given. The student will find it helpful in years to come if these distributions are committed to memory. There are several textbooks (my favorite is [6, chap. 6]) that give more complete descriptions of distributions, and we recommend gaining a familiarity with a variety of distribution functions before any serious modeling is attempted.

**Uniform-Discrete:** The random variable $N$ has a discrete uniform distribution if there are two integers $a$ and $b$ such that the pmf of $N$ can be written as

$$f(k) = \frac{1}{b-a+1} \quad \text{for } k = a, a+1, \cdots, b \,. \tag{1.8}$$

Then,

$$E[N] = \frac{a+b}{2}; \quad V[N] = \frac{(b-a+1)^2 - 1}{12} \,.$$

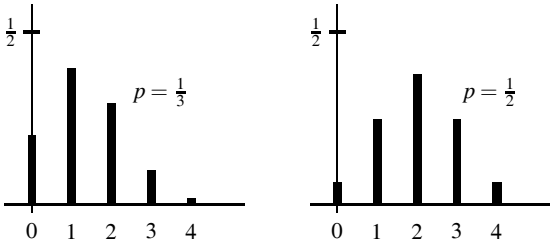


**Fig. 1.7** Two binomial probability mass functions

*Example 1.7.* Consider rolling a fair die. Figure 1.6 shows the uniform pmf for the "number of dots" random variable. Notice in the figure that, as the name "uniform" implies, all the probabilities are the same. □

**Bernoulli:** The random variable $N$ has a Bernoulli distribution if there is a number $0 < p < 1$ such that the pmf of $N$ can be written as

$$f(k) = \begin{cases} 1-p & \text{for } k = 0 \\ p & \text{for } k = 1 \,. \end{cases} \tag{1.9}$$

Then,

$$E[N] = p; \quad V[N] = p(1-p); \quad c^2[N] = \frac{1-p}{p} \,.$$

**Binomial:** (By James Bernoulli, 1654-1705; published posthumously in 1713.) The random variable $N$ has a binomial distribution if there is a number $0 < p < 1$ and a positive integer $n$ such that the pmf of $N$ can be written as

$$f(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \cdots, n \,. \tag{1.10}$$

Then,

$$E[N] = np; \quad V[N] = np(1-p); \quad C^2[N] = \frac{1-p}{np} \,.$$

The number $p$ is often though of as the probability of a success. The binomial pmf evaluated at $k$ thus gives the probability of $k$ successes occurring out of $n$ trials. The binomial random variable with parameters $p$ and $n$ is the sum of $n$ (independent) Bernoulli random variables each with parameter $p$.

*Example 1.8.* We are monitoring calls at a switchboard in a large manufacturing firm and have determined that one third of the calls are long distance and two thirds of the calls are local. We have decided to pick four calls at random and would like to know how many calls in the group of four are long distance. In other words, let $N$ be a random variable indicating the number of long distance calls in the group of four. Thus, $N$ is binomial with $n = 4$ and $p = 1/3$. It also happens that in this company, half of the individuals placing calls are women and half are men. We would also like to know how many of the group of four were calls placed by men. Let $M$ denote

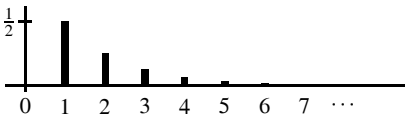


**Fig. 1.8** A geometric probability mass function

the number of men placing calls; thus, $M$ is binomial with $n = 4$ and $p = 1/2$. The pmf's for these two random variables are shown in Fig. 1.7. Notice that for $p = 0.5$, the pmf is symmetric, and as $p$ varies from 0.5, the graph becomes skewed. □

**Geometric:** The random variable $N$ has a geometric distribution if there is a number $0 < p < 1$ such that the pmf of $N$ can be written as

$$f(k) = p(1-p)^{k-1} \text{ for } k = 1, 2, \cdots . \tag{1.11}$$

Then,

$$E[N] = \frac{1}{p}; \quad V[N] = \frac{1-p}{p^2}; \quad C^2[N] = 1 - p.$$

The idea behind the geometric random variable is that it represents the number of trials until the first success occurs. In other words, $p$ is thought of as the probability of success for a single trial, and we continually perform the trials until a success occurs. The random variable $N$ is then set equal to the number of trial that we had to perform. Note that although the geometric random variable is discrete, its range is infinite.

*Example 1.9.* A car saleswoman has made a statistical analysis of her previous sales history and determined that each day there is a 50% chance that she will sell a luxury car. After careful further analysis, it is also clear that a luxury car sale on one day is independent of the sale (or lack of it) on another day. On New Year's Day (a holiday in which the dealership was closed) the saleswoman is contemplating when she will sell her first luxury car of the year. If $N$ is the random variable indicating the day of the first luxury car sale ($N = 1$ implies the sale was on January 2), then $N$ is distributed according to the geometric distribution with $p = 0.5$, and its pmf is shown in Fig. 1.8. Notice that theoretically the random variable has an infinite range, but for all practical purposes the probability of the random variable being larger than seven is negligible. □

**Poisson:** (By Simeon Denis Poisson, 1781-1840; published in 1837.) The random variable $N$ has a Poisson distribution if there is a number $\lambda > 0$ such that the pmf of $N$ can be written as

$$f(k) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \text{ for } k = 0, 1, \cdots . \tag{1.12}$$

Then,

$$E[N] = \lambda; \quad V[N] = \lambda; \quad C^2[N] = 1/\lambda \ .$$

The Poisson distribution is the most important discrete distribution in stochastic modeling. It arises in many different circumstances. One use is as an approximation to the binomial distribution. For $n$ large and $p$ small, the binomial is approximated by the Poisson by setting $\lambda = np$. For example, suppose we have a box of 144 eggs and there is a 1% probability that any one egg will break. Assuming that the breakage of eggs is independent of other eggs breaking, the probability that exactly 3 eggs will be broken out of the 144 can be determined using the binomial distribution with $n = 144$, $p = 0.01$, and $k = 3$; thus

$$\frac{144!}{141!3!}(0.01)^3(0.99)^{141} = 0.1181\,,$$

or by the Poisson approximation with $\lambda = 1.44$ that yields

$$\frac{(1.44)^3 e^{-1.44}}{3!} = 0.1179\,.$$

In 1898, L. V. Bortkiewicz [7, p. 206] reported that the number of deaths due to horse-kicks in the Prussian army was a Poisson random variable. Although this seems like a silly example, it is very instructive. The reason that the Poisson distribution holds in this case is due to the binomial approximation feature of the Poisson. Consider the situation: there would be a small chance of death by horse-kick for any one person (i.e., $p$ small) but a large number of individuals in the army (i.e., $n$ large). There are many analogous situations in modeling that deal with large populations and a small chance of occurrence for any one individual within the population. In particular, arrival processes (like arrivals to a bus station in a large city) can often be viewed in this fashion and thus described by a Poisson distribution. Another common use of the Poisson distribution is in population studies. The population size of a randomly growing organism often can be described by a Poisson random variable. W. S. Gosset, using the pseudonym of Student, showed in 1907 that the number of yeast cells in 400 squares of haemocytometer followed a Poisson distribution. Radioactive emissions are also Poisson as indicated in Example 1.2. (Fig. 1.4 also shows the Poisson pmf.)

Many arrival processes are well approximated using the Poisson probabilities. For example, the number of arriving telephone calls to a switchboard during a specified period of time, or the number of arrivals to a teller at a bank during a fixed period of time are often modeled as a Poisson random variable. Specifically, we say that an arrival process is a Poisson process with mean rate $\lambda$ if arrivals occur one-at-a-time and the number of arrivals during an interval of length $t$ is given by the random variable $N_t$ where

$$\Pr\{N_t = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad \text{for } k = 0, 1, \cdots . \tag{1.13}$$

**Uniform-Continuous:** The random variable $X$ has a continuous uniform distribution if there are two numbers $a$ and $b$ with $a < b$ such that the pdf of $X$ can be written as