Chapter 5

Discrete Random Variables

5.1 Introduction

Having been introduced to the basic probabilistic concepts in Chapters 3 and 4, we now begin their application to solving problems of interest. To do so we define the random variable. It will be seen to be a function, also called a *mapping*, of the outcomes of a random experiment to the set of real numbers. With this association we are able to use the real number description to quantify items of interest. In this chapter we describe the *discrete* random variable, which is one that takes on a finite or countably infinite number of values. Later we will extend the definition to a random variable that takes on a continuum of values, the *continuous* random variable. The mathematics associated with a discrete random variable are inherently simpler and so conceptualization is facilitated by first concentrating on the discrete problem. The reader has already been introduced to the concept of a random variable in Chapter 2 in an informal way and hence may wish to review the computer simulation methodology described therein.

5.2 Summary

The random variable, which is a mapping from the sample space into the set of real numbers, is formally discussed and illustrated in Section 5.3. In Section 5.4 the probability of a random variable taking on its possible values is given by (5.2). Next the probability mass function is defined by (5.3). Some important probability mass functions are summarized in Section 5.5. They include the Bernoulli (5.5), the binomial (5.6), the geometric (5.7), and the Poisson (5.8). The binomial probability mass function can be approximated by the Poisson as shown in Figure 5.8 if $M \to \infty$ and $p \to 0$, with Mp remaining constant. This motivates the use of the Poisson probability mass function for traffic modeling. If a random variable is transformed to a new one via a mapping, then the new random variable has a probability mass function given by (5.9). Next the cumulative distribution function is introduced and is given by (5.10). It can be used as an equivalent description for the probability of a discrete random variable. Its properties are summarized in Section 5.8. The computer simulation of discrete random variables is revisited in Section 5.9 with the estimate of the probability mass function and the cumulative distribution function given by (5.14) and (5.15),(5.16), respectively. Finally, the application of the Poisson probability model to determining the resources required to service customers is described in Section 5.10.

5.3 Definition of Discrete Random Variable

We have previously used a coin toss and a die toss as examples of a random experiment. In the case of a die toss the outcomes comprised the sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. This was because each face of a die has a dot pattern consisting of 1, 2, 3, 4, 5, or 6 dots. A natural description of the outcome of a die toss is therefore the number of dots observed on the face that appears upward. In effect, we have mapped the *dot pattern* into the *number of dots* in describing the outcome. This type of experiment is called a *numerically valued* random phenomenon since the basic output is a real number. In the case of a coin toss the outcomes comprise the *nonnumerical* sample space $\mathcal{S} = \{\text{head, tail}\}$. We have, however, at times replaced the sample space by one consisting only of real numbers such as $\mathcal{S}_X = \{0, 1\}$, where a head is mapped into a 1 and a tail is mapped into a 0. This mapping is shown in Figure 5.1. For many applications this is a convenient mapping. For example, in

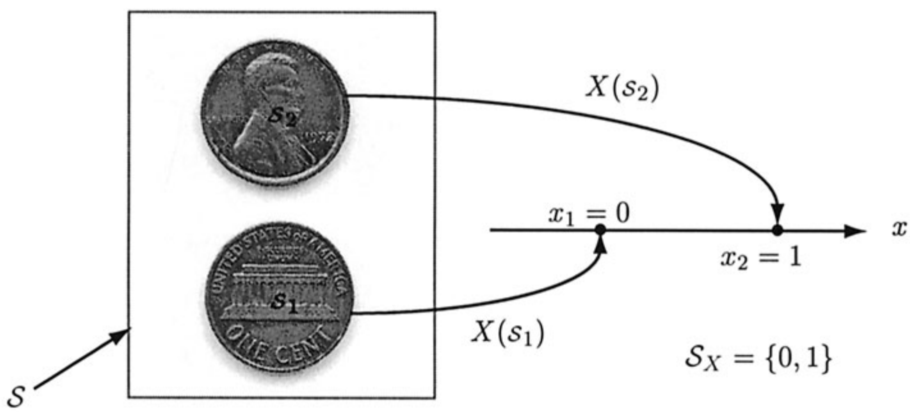


Figure 5.1: Mapping of the outcome of a coin toss into the set of real numbers.

a succession of M coin tosses, we might be interested in the total number of heads observed. With the defined mapping of

$$X(s_i) = \begin{cases} 0 & s_1 = \text{tail} \\ 1 & s_2 = \text{head} \end{cases}$$

we could represent the number of heads as $\sum_{i=1}^{M} X(s_i)$, where s_i is the outcome of the ith toss. The function that maps $\mathcal{S}$ into $\mathcal{S}_X$ and which is denoted by $X(\cdot)$ is called a *random variable*. It is a function that takes each outcome of $\mathcal{S}$ (which may not necessarily be a set of numbers) and maps it into the subset of the set of real numbers. Note that as previously mentioned in Chapter 2, a *capital* letter X will denote the random variable and a *lowercase* letter x its value. This convention for the coin toss example produces the assignment

$$X(s_i) = x_i \qquad i = 1, 2$$

where s_1 = tail and thus $x_1 = 0$, and s_2 = head and thus $x_2 = 1$. The name *random variable* is a poor one in that the function $X(\cdot)$ *is not random* but a known one and usually one of our own choosing. What is random is the input argument s_i and hence the output of the function is random. However, due to its long-standing usage in probability we will retain this terminology.

Sometimes it is more convenient to use a particular random variable for a given experiment. For example, in Chapter 2 we described a digital communication system called a PSK system. A bit is communicated using the transmitted signals

$$s(t) = \begin{cases} -A\cos 2\pi F_0 t & \text{for a } 0 \\ A\cos 2\pi F_0 t & \text{for a } 1. \end{cases}$$

Usually a 1 or a 0 occurs with equal probability so that the choice of a bit can be modeled as the outcome of a fair coin tossing experiment. If a head is observed, then a 1 is transmitted and a 0 otherwise. As a result, we could represent the transmitted signal with the model

$$s_i(t) = X(s_i) A\cos 2\pi F_0 t$$

where s_1 = tail and s_2 = head and hence we have the defined random variable

$$X(s_i) = \begin{cases} -1 & s_1 = \text{tail} \\ +1 & s_2 = \text{head}. \end{cases}$$

This random variable is a convenient one for this application.

In general, a *random variable* is a function that maps the sample space $\mathcal{S}$ into a subset of the real line. The real line will be denoted by R ($R = \{x : -\infty < x < \infty\}$). For a discrete random variable this subset is a finite or countably infinite set of points. The subset forms a new sample space which we will denote by $\mathcal{S}_X$, and which is illustrated in Figure 5.2. A discrete random variable may also map *multiple elements* of the sample space into the *same number* as illustrated in Figure 5.3. An example would be a die toss experiment in which we were only interested in whether the outcome is even or odd. To quantify this outcome we could define

$$X(s_i) = \begin{cases} 0 & \text{if } s_i = 1, 3, 5 \text{ dots} \\ 1 & \text{if } s_i = 2, 4, 6 \text{ dots}. \end{cases}$$

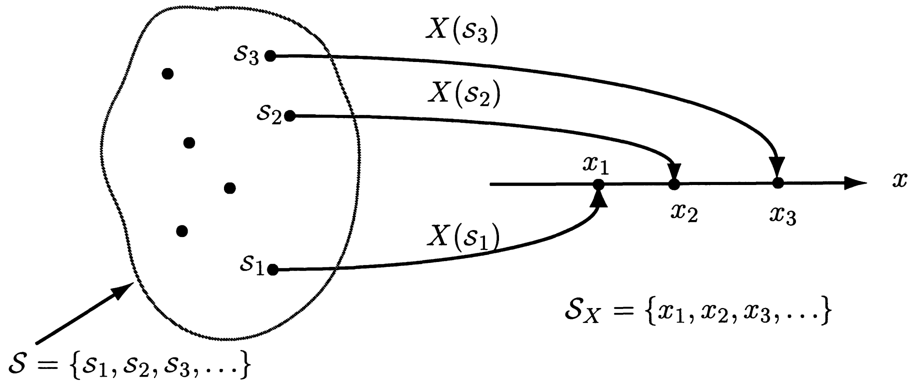


Figure 5.2: Discrete random variable as a *one-to-one* mapping of a countably infinite sample space into set of real numbers.

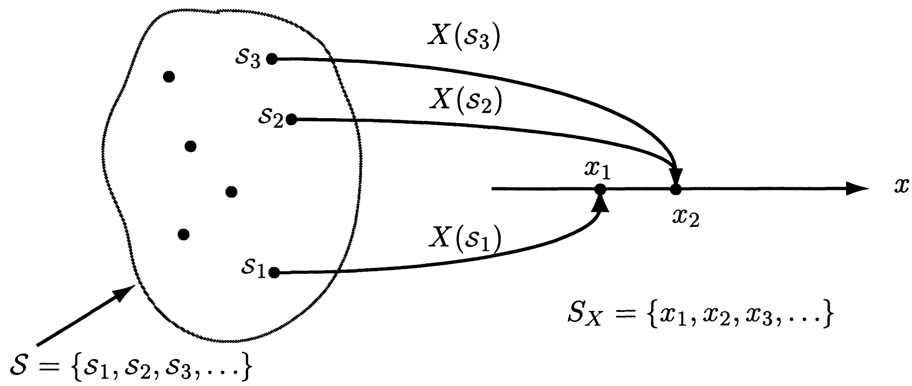


Figure 5.3: Discrete random variable as a *many-to-one* mapping of a countably infinite sample space into set of real numbers.

This type of mapping is usually called a *many-to-one* mapping while the previous one is called a *one-to-one* mapping. Note that for a many-to-one mapping we cannot recover the outcome of $\mathcal{S}$ if we know the value of $X(s)$. But as already explained, this is of little concern since we initially defined the random variable to output the item of interest. Lastly, for numerically valued random experiments in which s is contained in R, we can still use the random variable approach if we define $X(s) = s$ for all s. This allows the concept of a random variable to be used for all random experiments, with either numerical or nonnumerical outputs.

5.4 Probability of Discrete Random Variables

We would next like to determine the probabilities of the random variable taking on its possible values. In other words, what is the probability $P[X(s_i) = x_i]$ for each

$x_i \in \mathcal{S}_X$? Since the sample space $\mathcal{S}$ is discrete, the random variable can take on at most a countably infinite number of values or $X(s_i) = x_i$ for $i = 1, 2, \ldots$. It should be clear that if $X(\cdot)$ maps each s_i into a *different* x_i (or $X(\cdot)$ is one-to-one), then because s_i and x_i are just two different names for the same event

$$P[X(s) = x_i] = P[\{s_j : X(s_j) = x_i\}] = P[\{s_i\}] \tag{5.1}$$

or we assign a probability to the value of the random variable equal to that of the simple event in $\mathcal{S}$ that yields that value. If, however, there are multiple outcomes in $\mathcal{S}$ that map into the same value x_i (or $X(\cdot)$ is many-to-one) then

$$\begin{aligned} P[X(s) = x_i] &= P[\{s_j : X(s_j) = x_i\}] \\ &= \sum_{\{j : X(s_j) = x_i\}} P[\{s_j\}] \end{aligned} \tag{5.2}$$

since the s_j's are simple events in $\mathcal{S}$ and are therefore mutually exclusive. It is said that the events $\{X = x_i\}$, defined on $\mathcal{S}_X$, and $\{s_j : X(s_j) = x_i\}$, defined on $\mathcal{S}$, are *equivalent events*. As such they are assigned the same probability. Note that the probability assignment (5.2) subsumes that of (5.1) and that in either case we can summarize the probabilities that the random variable values take on by defining the *probability mass function* (PMF) as

$$p_X[x_i] = P[X(s) = x_i] \tag{5.3}$$

and use (5.2) to evaluate it from a knowledge of the mapping. It is important to observe that in the notation $p_X[x_i]$ the subscript X refers to the random variable and also the $[\cdot]$ notation is meant to remind the reader that the argument is a discrete one. Later, we will use $(\cdot)$ for continuous arguments. In summary, the probability mass function is the probability that the random variable X takes on the value x_i for each possible x_i. An example follows.

Example 5.1 – Coin toss – one-to-one mapping

The experiment consists of a single coin toss with a probability of heads equal to p. The sample space is $\mathcal{S} = \{\text{head, tail}\}$ and we define the random variable as

$$X(s_i) = \begin{cases} 0 & s_i = \text{tail} \\ 1 & s_i = \text{head.} \end{cases}$$

The PMF is therefore from from (5.3) and (5.1)

$$\begin{aligned} p_X[0] &= P[X(s) = 0] = 1 - p \\ p_X[1] &= P[X(s) = 1] = p. \end{aligned}$$

◇

Example 5.2 – Die toss – many-to-one mapping

The experiment consists of a single fair die toss. With a sample space of $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ and an interest only in whether the outcome is even or odd we define the random variable

$$X(s_i) = \begin{cases} 0 & \text{if } i = 1, 3, 5 \\ 1 & \text{if } i = 2, 4, 6. \end{cases}$$

Thus, using (5.3) and (5.2) we have the PMF

$$\begin{aligned} p_X[0] &= P[X(s) = 0] = \sum_{j=1,3,5} P[\{s_j\}] = \frac{3}{6} \\ p_X[1] &= P[X(s) = 1] = \sum_{j=2,4,6} P[\{s_j\}] = \frac{3}{6}. \end{aligned}$$

◇

The use of (5.2) may seem familiar and indeed it should. We have summed the probabilities of simple events in $\mathcal{S}$ to obtain the probability of an event in $\mathcal{S}$ using (3.10). Here, the event is just the subset of $\mathcal{S}$ for which $X(s) = x_i$ holds. The introduction of a random variable has quantified the events of interest!

Finally, because PMFs $p_X[x_i]$ are just new names for the probabilities $P[X(s) = x_i]$ they must satisfy the usual properties:

Property 5.1 – Range of values

$$0 \leq p_X[x_i] \leq 1$$

□

Property 5.2 – Sum of values

$$\sum_{i=1}^{M} p_X[x_i] = 1 \text{ if } \mathcal{S}_X \text{ consists of } M \text{ outcomes}$$

$$\sum_{i=1}^{\infty} p_X[x_i] = 1 \text{ if } \mathcal{S}_X \text{ is countably infinite.}$$

□

We will frequently omit the s argument of X to write $p_X[x_i] = P[X = x_i]$.

Once the PMF has been specified all subsequent probability calculations can be based on it, without referring back to the original sample space $\mathcal{S}$. Specifically, for an event A defined on $\mathcal{S}_X$ the probability is given by

$$P[X \in A] = \sum_{\{i: x_i \in A\}} p_X[x_i]. \tag{5.4}$$

An example follows.

Example 5.3 – Calculating probabilities based on the PMF

Consider a die whose sides have been labeled with two sides having 1 dot, two sides having 2 dots, and two sides having 3 dots. Hence, $\mathcal{S} = \{s_1, s_2, \ldots, s_6\} = \{\text{side 1, side 2, side 3, side 4, side 5, side 6}\}$. Then if we are interested in the probabilities of the outcomes displaying either 1, 2, or 3 dots, we would define a random variable as

$$X(s_i) = \begin{cases} 1 & i = 1,2 \\ 2 & i = 3,4 \\ 3 & i = 5,6. \end{cases}$$

It easily follows then that the PMF is from (5.2)

$$p_X[1] = p_X[2] = p_X[3] = \frac{1}{3}.$$

Now assume we are interested in the probability that a 2 or 3 occurs or $A = \{2,3\}$. Then from (5.4) we have

$$P[X \in \{2,3\}] = p_X[2] + p_X[3] = \frac{2}{3}.$$

There is no need to reconsider the original sample space $\mathcal{S}$ and all probability calculations of interest are obtainable from the PMF.

$\diamondsuit$

5.5 Important Probability Mass Functions

We have already encountered many of these in Chapter 4. We now summarize these in our new notation. Since the sample spaces $\mathcal{S}_X$ consist of integer values we will replace the notation x_i by k, which indicates an integer.

5.5.1 Bernoulli

$$p_X[k] = \begin{cases} 1-p & k = 0 \\ p & k = 1. \end{cases} \tag{5.5}$$

The PMF is shown in Figure 5.4 and is recognized as a *sequence* of numbers that is nonzero only for the indices $k = 0,1$. It is convenient to represent the Bernoulli PMF using the shorthand notation $\text{Ber}(p)$. With this notation we replace the description that "X is distributed according to a Bernoulli random variable with PMF $\text{Ber}(p)$" by the shorthand notation $X \sim \text{Ber}(p)$, where $\sim$ means "is distributed according to".

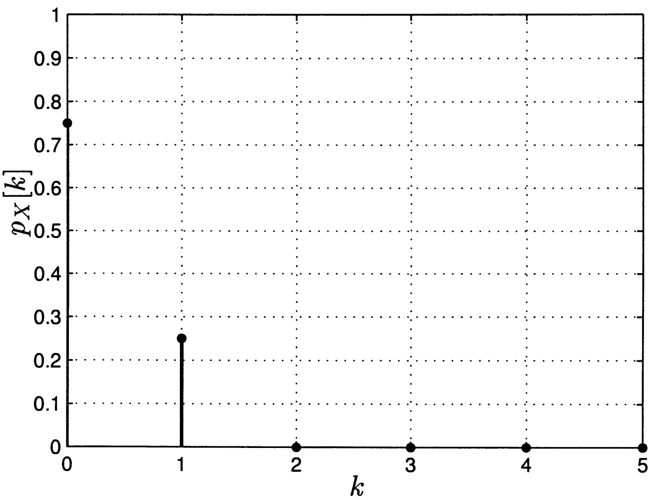


Figure 5.4: Bernoulli probability mass function for $p = 0.25$.

5.5.2 Binomial

$$p_X[k] = \binom{M}{k} p^k (1-p)^{M-k} \qquad k = 0, 1, \ldots, M. \tag{5.6}$$

The PMF is shown in Figure 5.5. The shorthand notation for the binomial PMF is

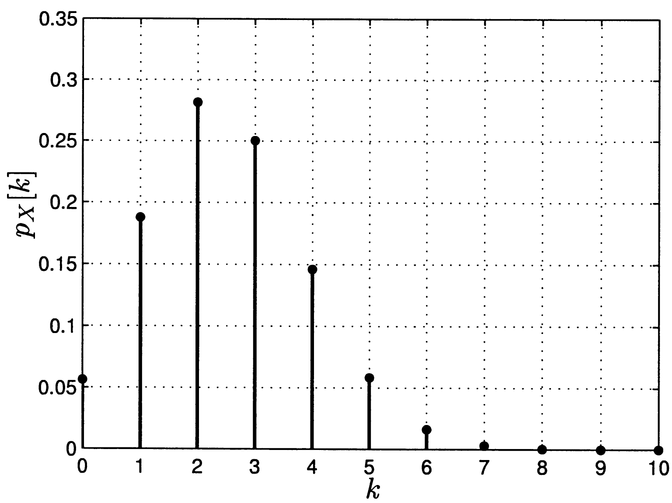


Figure 5.5: Binomial probability mass function for $M = 10, p = 0.25$.

$\text{bin}(M, p)$. The location of the maximum of the PMF can be shown to be given by $[(M+1)p]$, where $[x]$ denotes the largest integer less than or equal to x (see Problem 5.7).

5.5.3 Geometric

$$p_X[k] = (1-p)^{k-1}p \qquad k = 1, 2, \ldots. \tag{5.7}$$

The PMF is shown in Figure 5.6. The shorthand notation for the geometric PMF is geom(p).

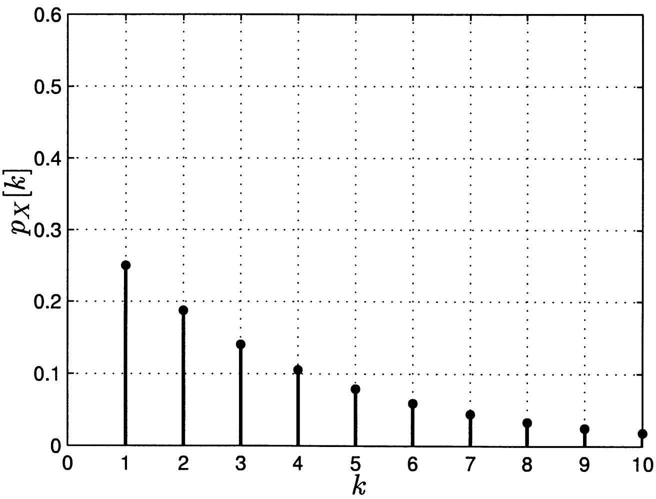


Figure 5.6: Geometric probability mass function for $M = 10, p = 0.25$.

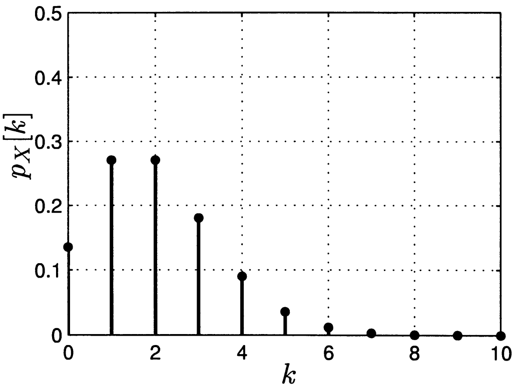
5.5.4 Poisson

$$p_X[k] = \exp(-\lambda)\frac{\lambda^k}{k!} \qquad k = 0, 1, 2, \ldots \tag{5.8}$$

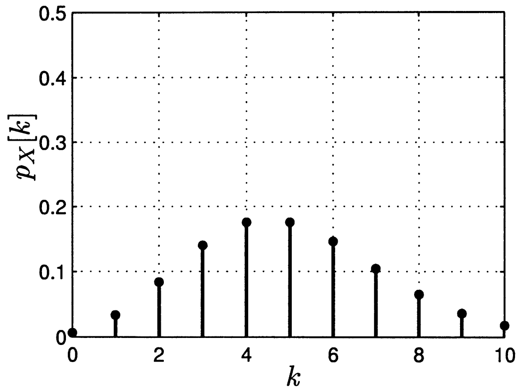
where $\lambda > 0$. The PMF is shown in Figure 5.7 for several values of λ. Note that the maximum occurs at $[\lambda]$ (see Problem 5.11). The shorthand notation is Pois(λ).

5.6 Approximation of Binomial PMF by Poisson PMF

The binomial and Poisson PMFs are related to each other under certain conditions. This relationship helps to explain why the Poisson PMF is used in various applications, primarily traffic modeling as described further in Section 5.10. The relationship is as follows. If in a binomial PMF, we let $M \to \infty$ as $p \to 0$ such that the product $\lambda = Mp$ remains constant, then bin(M, p) $\to$ Pois(λ). Note that $\lambda = Mp$ represents the expected or average number of successes in M Bernoulli trials (see Chapter 6 for definition of expectation). Hence, by keeping the average number of successes fixed but assuming more and more trials with smaller and smaller probabilities of success on each trial, we are led to a Poisson PMF. As an example, a comparison is shown in Figure 5.8 for $M = 10, p = 0.5$ and $M = 100, p = 0.05$. This

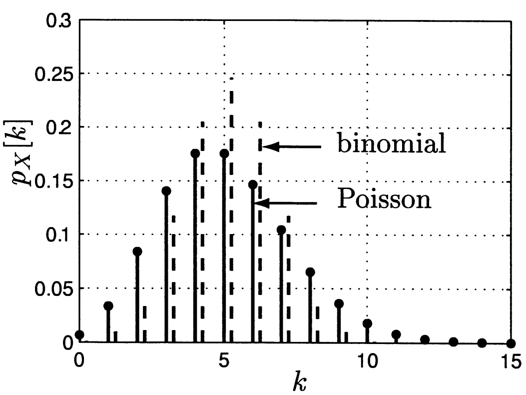


(a) $\lambda = 2$

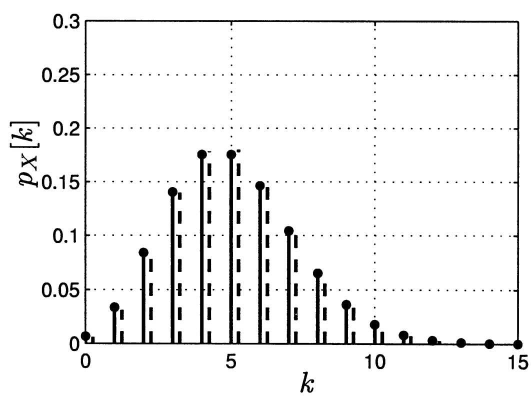


(b) $\lambda = 5$

Figure 5.7: The Poisson probability mass function for different values of λ.



(a) $M = 10, p = 0.5$



(b) $M = 100, p = 0.05$

Figure 5.8: The Poisson approximation to the binomial probability mass function.

result is primarily useful since Poisson PMFs are easier to manipulate and also arise in the modeling of point processes as described in Chapter 21.

To make this connection we have for the binomial PMF with $p = \lambda/M \to 0$ as

$M \to \infty$ (and λ fixed)

$$\begin{aligned}
p_X[k] &= \binom{M}{k} p^k (1-p)^{M-k} \\
&= \frac{M!}{(M-k)!k!} \left(\frac{\lambda}{M}\right)^k \left(1 - \frac{\lambda}{M}\right)^{M-k} \\
&= \frac{(M)_k}{k!} \frac{\lambda^k}{M^k} \frac{(1-\lambda/M)^M}{(1-\lambda/M)^k} \\
&= \frac{\lambda^k}{k!} \frac{(M)_k}{M^k} \frac{(1-\lambda/M)^M}{(1-\lambda/M)^k}.
\end{aligned}$$

But for a fixed k, as $M \to \infty$, we have that $(M)_k/M^k \to 1$. Also, for a fixed k, $(1-\lambda/M)^k \to 1$ so that we need only find the limit of $g(M) = (1-\lambda/M)^M$ as $M \to \infty$. This is shown in Problem 5.15 to be $\exp(-\lambda)$ and therefore

$$p_X[k] \to \frac{\lambda^k}{k!} \exp(-\lambda).$$

Also, since the binomial PMF is defined for $k = 0, 1, \ldots, M$, as $M \to \infty$ the limiting PMF is defined for $k = 0, 1, \ldots$. This result can also be found using characteristic functions as shown in Chapter 6.

5.7 Transformation of Discrete Random Variables

It is frequently of interest to be able to determine the PMF of a *transformed* random variable. Mathematically, we desire the PMF of the new random variable $Y = g(X)$, where X is a discrete random variable. For example, consider a die whose faces are labeled with the numbers $0, 0, 1, 1, 2, 2$. We wish to find the PMF of the number observed when the die is tossed, assuming all sides are equally likely to occur. If the original sample space is composed of the possible cube sides that can occur, so that $\mathcal{S}_X = \{1, 2, 3, 4, 5, 6\}$, then the transformation appears as shown in Figure 5.9. Specifically, we have that

$$Y = \begin{cases} y_1 = 0 & \text{if } x = x_1 = 1 \text{ or } x = x_2 = 2 \\ y_2 = 1 & \text{if } x = x_3 = 3 \text{ or } x = x_4 = 4 \\ y_3 = 2 & \text{if } x = x_5 = 5 \text{ or } x = x_6 = 6. \end{cases}$$

Note that the transformation is many-to-one. Since events such as $\{y : y = y_1 = 0\}$ and $\{x : x = x_1 = 1, x = x_2 = 2\}$, for example, are equivalent, they should be assigned the same probability. Thus, using the property that the events $\{X = x_i\}$ are simple events defined on $\mathcal{S}_X$, we have that

$$p_Y[y_i] = \begin{cases} p_X[1] + p_X[2] = \frac{1}{3} & i = 1 \\ p_X[3] + p_X[4] = \frac{1}{3} & i = 2 \\ p_X[5] + p_X[6] = \frac{1}{3} & i = 3. \end{cases}$$

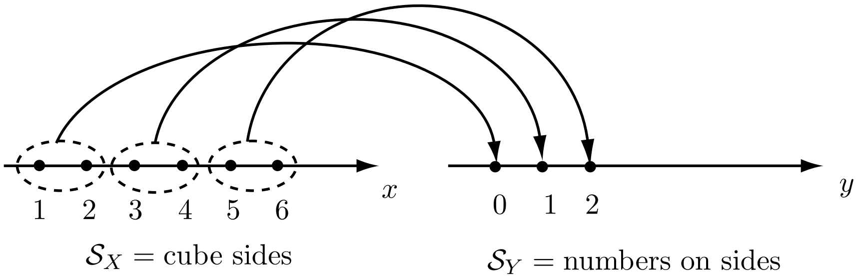


Figure 5.9: Transformation of discrete random variable.

In general, we have that

$$p_Y[y_i] = \sum_{\{j: g(x_j)=y_i\}} p_X[x_j]. \tag{5.9}$$

We just sum up the probabilities for all the values of $X = x_j$ that are mapped into $Y = y_i$. This is reminiscent of (5.2) in which the transformation was from the objects s_j defined on $\mathcal{S}$ to the numbers x_i defined on $\mathcal{S}_X$. In fact, it is nearly identical except that we have replaced the objects that are to be transformed by numbers, i.e., the x_j's. Some examples of this procedure follow.

Example 5.4 – One-to-one transformation of Bernoulli random variable
If $X \sim \text{Ber}(p)$ and $Y = 2X - 1$, determine the PMF of Y. The sample space for X is $\mathcal{S}_X = \{0, 1\}$ and consequently that for Y is $\mathcal{S}_Y = \{-1, 1\}$. It follows that $x_1 = 0$ maps into $y_1 = -1$ and $x_2 = 1$ maps into $y_2 = 1$. As a result, we have from (5.9)

$$\begin{aligned} p_Y[-1] &= p_X[0] = 1 - p \\ p_Y[1] &= p_X[1] = p. \end{aligned}$$

Note that this mapping is particularly simple since it is one-to-one. A slightly more complicated example is next.

◊

Example 5.5 – Many-to-one transformation
Let the transformation be $Y = g(X) = X^2$ which is defined on the sample space $\mathcal{S}_X = \{-1, 0, 1\}$ so that $\mathcal{S}_Y = \{0, 1\}$. Clearly, $g(x_j) = x_j^2 = 0$ only for $x_j = 0$. Hence,

$$p_Y[0] = p_X[0].$$

However, $g(x_j) = x_j^2 = 1$ for $x_j = -1$ and $x_j = 1$. Thus, using (5.9) we have

$$\begin{aligned} p_Y[1] &= \sum_{\{x_j : x_j^2 = 1\}} p_X[x_j] \\ &= p_X[-1] + p_X[1]. \end{aligned}$$

Note that we have determined $p_Y[y_i]$ by summing the probabilities of all the x_j's that map into y_i via the transformation $y = g(x)$. This is in essence the meaning of (5.9).

◊

Example 5.6 – Many-to-one transformation of Poisson random variable
Now consider $X \sim \text{Pois}(\lambda)$ and define the transformation $Y = g(X)$ as

$$Y = \begin{cases} 1 & \text{if } X = k \text{ is even} \\ -1 & \text{if } X = k \text{ is odd.} \end{cases}$$

To find the PMF for Y we use

$$p_Y[k] = P[Y = k] = \begin{cases} P[X \text{ is even}] & k = 1 \\ P[X \text{ is odd}] & k = -1. \end{cases}$$

We need only determine $p_Y[1]$ since $p_Y[-1] = 1 - p_Y[1]$. Thus, from (5.9)

$$\begin{aligned} p_Y[1] &= \sum_{j=0 \text{ and even}}^{\infty} p_X[j] \\ &= \sum_{j=0 \text{ and even}}^{\infty} \exp(-\lambda)\frac{\lambda^j}{j!}. \end{aligned}$$

To evaluate the infinite sum in closed form we use the following "trick"

$$\begin{aligned} \sum_{j=0 \text{ and even}}^{\infty} \frac{\lambda^j}{j!} &= \frac{1}{2}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} + \frac{1}{2}\sum_{j=0}^{\infty}\frac{(-\lambda)^j}{j!} \\ &= \frac{1}{2}\exp(\lambda) + \frac{1}{2}\exp(-\lambda) \end{aligned}$$

since the Taylor expansion of $\exp(x)$ is known to be $\sum_{j=0}^{\infty} x^j/j!$ (see Problem 5.22). Finally, we have that

$$\begin{aligned} p_Y[1] &= \exp(-\lambda)\left[\frac{1}{2}\exp(\lambda) + \frac{1}{2}\exp(-\lambda)\right] = \frac{1}{2}(1 + \exp(-2\lambda)) \\ p_Y[-1] &= 1 - p_Y[1] = \frac{1}{2}(1 - \exp(-2\lambda)). \end{aligned}$$

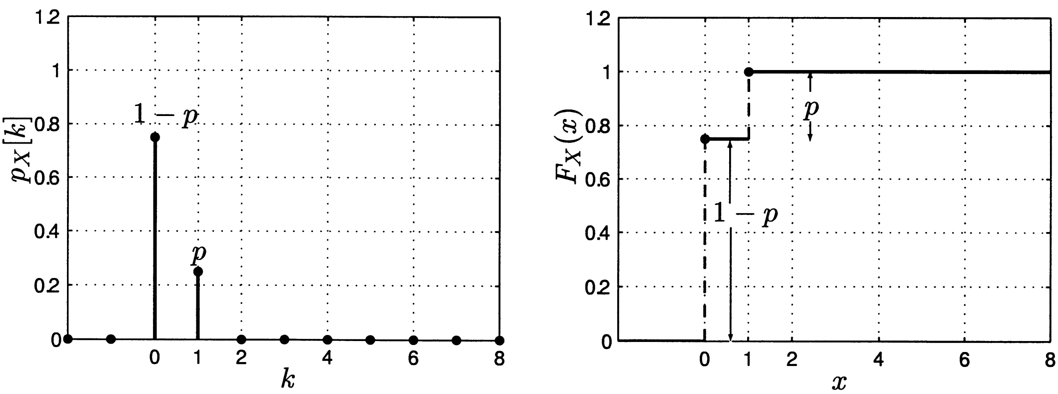
◊

5.8 Cumulative Distribution Function

An alternative means of summarizing the probabilities of a discrete random variable is the *cumulative distribution function* (CDF). It is sometimes referred to more succinctly as the *distribution function.* The CDF for a random variable X and evaluated at x is given by $P[\{\text{real numbers } x' : x' \leq x\}]$, which is the probability that X lies in the semi-infinite interval $(-\infty, x]$. It is therefore defined as

$$F_X(x) = P[X \leq x] \quad -\infty < x < \infty. \tag{5.10}$$

It is important to observe that the value $X = x$ is included in the interval. As an example, if $X \sim \text{Ber}(p)$, then the PMF and the corresponding CDF are shown in Figure 5.10. Because the random variable takes on only the values 0 and 1, the CDF



(a) PMF (b) CDF

Figure 5.10: The Bernoulli probability mass function and cumulative distribution function for $p = 0.25$.

changes its value only at these points, where it jumps. The CDF can be thought of as a "running sum" which adds up the probabilities of the PMF starting at $-\infty$ and ending at $+\infty$. When the value x of $F_X(x)$ encounters a nonzero value of the PMF, the additional mass causes the CDF to jump, with the size of the jump equal to the value of the PMF at that point. For example, referring to Figure 5.10b, at $x = 0$ we have $F_X(0) = p_X[0] = 1 - p = 3/4$ and at $x = 1$ we have $F_X(1) = p_X[0] + p_X[1] = 1$, with the jump having size $p_X[1] = p = 1/4$. Another example follows.

Example 5.7 – CDF for geometric random variable

Since $p_X[k] = (1-p)^{k-1}p$ for $k = 1, 2, \ldots$, we have the CDF

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \displaystyle\sum_{i=1}^{[x]} (1-p)^{i-1}p & x \geq 1 \end{cases}$$

where $[x]$ denotes the largest integer less than or equal to x. This evaluates to

$$F_X(x) = \begin{cases} 0 & x < 1 \\ p & 1 \leq x < 2 \\ p + (1-p)p & 2 \leq x < 3 \\ \text{etc.} & \end{cases}$$

The PMF and CDF are plotted in Figure 5.11 for $p = 0.5$. Since the CDF jumps at

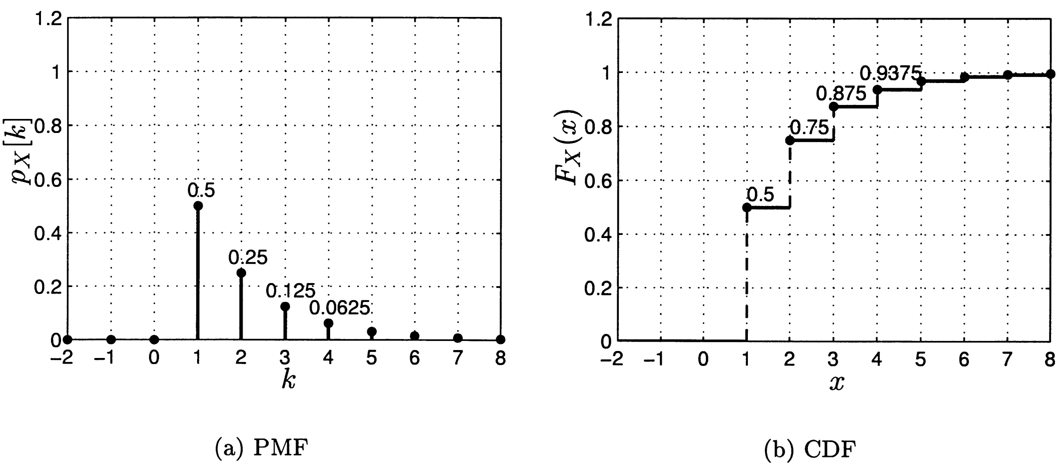


Figure 5.11: The geometric probability mass function and cumulative distribution function for $p = 0.5$.

each nonzero value of the PMF and the jump size is that value of the PMF, we can recover the PMF from the CDF. In particular, we have that

$$p_X[x] = F_X(x^+) - F_X(x^-)$$

where x^+ denotes a value just slightly larger than x and x^- denotes a value just slightly smaller than x. Thus, if $F_X(x)$ does not have a discontinuity at x the value of the PMF is zero. At a discontinuity the value of the PMF is just the jump size as previously asserted. Also, because of the definition of the CDF, i.e., that $F_X(x) = P[X \leq x] = P[X < x \text{ or } X = x]$, the value of $F_X(x)$ is the value *after* the jump. The CDF is said to be *right-continuous* which is sometimes stated mathematically as $\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$ at the point $x = x_0$.

$\Diamond$

From the previous example we see that the PMF and CDF are equivalent descriptions of the probability assignment for X. Either one can be used to find the probability of X being in an interval (even an interval of length zero). For example, to determine $P[3/2 < X \leq 7/2]$ for the geometric random variable

$$\begin{aligned} P\left[\frac{3}{2} < X \leq \frac{7}{2}\right] &= p_X[2] + p_X[3] \\ &= F_X\left(\frac{7}{2}\right) - F_X\left(\frac{3}{2}\right) \end{aligned}$$

as is evident by referring to Figure 5.11b. We need to be careful, however, to note whether the endpoints of the interval are included or not. This is due to the discontinuities of the CDF. Because of the definition of the CDF as the probability of X being within the interval $(-\infty, x]$, which *includes* the right-most point, we have for the interval $(a, b]$

$$P[a < X \leq b] = F_X(b^+) - F_X(a^+). \tag{5.11}$$

Also, the other intervals (a, b), $[a, b)$, and $[a, b]$ will in general have different probabilities than that given by (5.11). From Figure 5.11b and (5.11) we have as an example that

$$P[2 < X \leq 3] = F_X(3^+) - F_X(2^+) = p_X[3] = (1-p)^2 p = 0.125$$

but

$$P[2 \leq X \leq 3] = F_X(3^+) - F_X(2^-) = (1-p)p + (1-p)^2 p = 0.375.$$

From the definition of the CDF and as further illustrated in Figures 5.10 and 5.11 the CDF has several important properties. They are now listed and proven.

Property 5.3 – CDF is between 0 and 1.

$$0 \leq F_X(x) \leq 1 \qquad -\infty < x < \infty$$

Proof: Since by definition $F_X(x) = P[X \leq x]$ is a probability for all x, it must lie between 0 and 1.

□

Property 5.4 – Limits of CDF as $x \to -\infty$ and as $x \to \infty$

$$\begin{aligned} \lim_{x \to -\infty} F_X(x) &= 0 \\ \lim_{x \to +\infty} F_X(x) &= 1. \end{aligned}$$

Proof:

$$\lim_{x \to -\infty} F_X(x) = P[\{s : X(s) < -\infty\}] = P[\emptyset] = 0$$

since the values that $X(s)$ can take on do not include $-\infty$. Also,

$$\lim_{x\to+\infty} F_X(x) = P[\{s : X(s) < +\infty\}] = P[S] = 1$$

since the values that $X(s)$ can take on are all included on the real line.

□

Property 5.5 – CDF is monotonically increasing.
A monotonically increasing function $g(\cdot)$ is one in which for every x_1 and x_2 with $x_1 \leq x_2$, it follows that $g(x_1) \leq g(x_2)$ or the function increases or stays the same as the argument increases (see also Problem 5.29).
Proof:

$$\begin{aligned} F_X(x_2) &= P[X \leq x_2] && \text{(definition)} \\ &= P[(X \leq x_1) \cup (x_1 < X \leq x_2)] \\ &= P[X \leq x_1] + P[x_1 < X \leq x_2] && \text{(Axiom 3)} \\ &= F_X(x_1) + P[x_1 < X \leq x_2] \geq F_X(x_1). && \text{(definition and Axiom 1)} \end{aligned}$$

Alternatively, if $A = \{-\infty < X \leq x_1\}$ and $B = \{-\infty < X \leq x_2\}$ with $x_1 \leq x_2$, then $A \subset B$. From Property 3.5 (montonicity) $F_X(x_2) = P[B] \geq P[A] = F_X(x_1)$.

□

Property 5.6 – CDF is right-continuous.
By right-continuous it is meant that as we approach the point x_0 from the right, the limiting value of the CDF should be the value of the CDF at that point. Mathematically, it is expressed as

$$\lim_{x\to x_0^+} F_X(x) = F_X(x_0).$$

Proof:
The proof relies on the continuity property of the probability function. It can be found in [Ross 2002].

□

Property 5.7 – Probability of interval found using the CDF

$$P[a < X \leq b] = F_X(b) - F_X(a) \tag{5.12}$$

or more explicitly to remind us of possible discontinuities

$$P[a < X \leq b] = F_X(b^+) - F_X(a^+). \tag{5.13}$$

Proof:
Since for $a < b$

$$\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}$$

and the intervals on the right-hand-side are disjoint (mutually exclusive events), by Axiom 3

$$P[-\infty < X \leq b] = P[-\infty < X \leq a] + P[a < X \leq b]$$

or rearranging terms we have that

$$P[a < X \leq b] = P[-\infty < X \leq b] - P[-\infty < X \leq a] = F_X(b) - F_X(a).$$

□

5.9 Computer Simulation

In Chapter 2 we discussed how to simulate a discrete random variable on a digital computer. In particular, Section 2.4 presented some MATLAB code. We now continue that discussion to show how to simulate a discrete random variable and estimate its PMF and CDF. Assume that X can take on values in $\mathcal{S}_X = \{1, 2, 3\}$ with a PMF

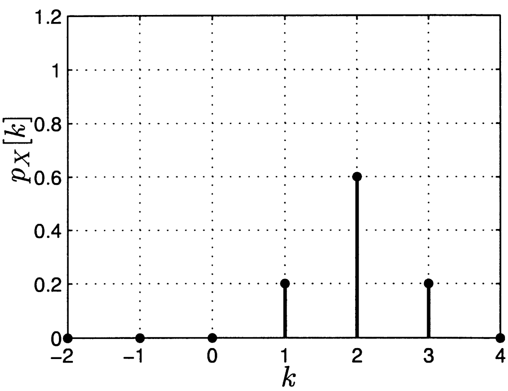
$$p_X[x] = \begin{cases} p_1 = 0.2 & \text{if } x = x_1 = 1 \\ p_2 = 0.6 & \text{if } x = x_2 = 2 \\ p_3 = 0.2 & \text{if } x = x_3 = 3. \end{cases}$$

The PMF and CDF are shown in Figure 5.12. The code from Section 2.4 for generating M realizations of X is

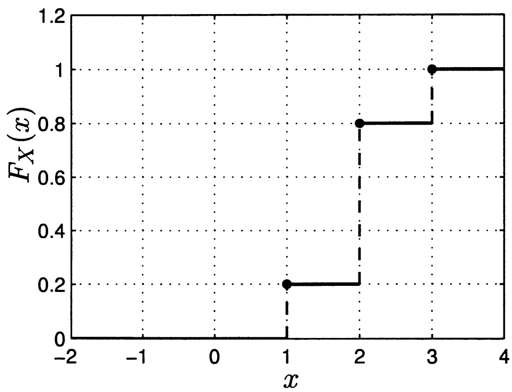
```
for i=1:M
  u=rand(1,1);
  if u<=0.2
    x(i,1)=1;
  elseif u>0.2 & u<=0.8
    x(i,1)=2;
  elseif u>0.8
    x(i,1)=3;
  end
end
```

Recall that U is a random variable whose values are equally likely to fall within the interval $(0, 1)$. It is called the *uniform random variable* and is described further in Chapter 10. Now to estimate the PMF $p_X[k] = P[X = k]$ for $k = 1, 2, 3$ we use the relative frequency interpretation of probability to yield

$$\hat{p}_X[k] = \frac{\text{Number of outcomes equal to } k}{M} \qquad k = 1, 2, 3. \tag{5.14}$$



(a) PMF



(b) CDF

Figure 5.12: The probability mass function and cumulative distribution function for computer simulation example.

For $M = 100$ this is shown in Figure 5.13a. Also, the CDF is estimated for all x via

$$\hat{F}_X(x) = \frac{\text{Number of outcomes } \leq x}{M} \tag{5.15}$$

or equivalently by

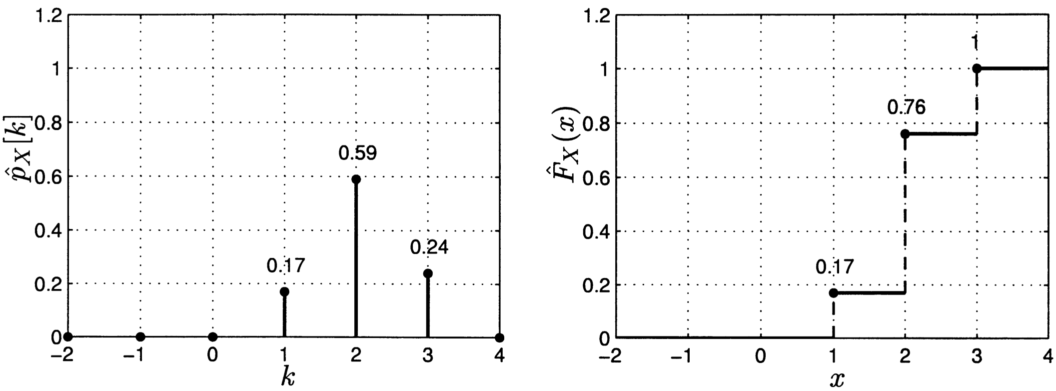
$$\hat{F}_X(x) = \sum_{\{k:k\leq x\}} \hat{p}_X[k] \tag{5.16}$$

and is shown in Figure 5.13b. For finite sample spaces this approach to simulate a discrete random variable is adequate. But for infinite sample spaces such as for the geometric and Poisson random variables a different approach is needed. See Problem 5.30 for a further discussion.

Before concluding our discussion we wish to point out a useful property of CDFs that simplifies the computer generation of random variable outcomes. Note from Figure 5.12b with $u = F_X(x)$ that we can define an inverse CDF as $x = F_X^{-1}(u)$ where

$$x = F_X^{-1}(u) = \begin{cases} 1 & \text{if } 0 < u \leq 0.2 \\ 2 & \text{if } 0.2 < u \leq 0.8 \\ 3 & \text{if } 0.8 < u < 1 \end{cases}$$

or we choose the value of x as shown in Figure 5.14. But if u is the outcome of a uniform random variable U on $(0, 1)$, then this procedure is identical to that implemented in the previous MATLAB program used to generate realizations of X. A more general program is given in Appendix 6B as `PMFdata.m`. This is not merely a coincidence but can be shown to follow from the definition of the CDF.



(a) PMF

(b) CDF

Figure 5.13: The estimated probability mass function and corresponding estimated cumulative distribution function.

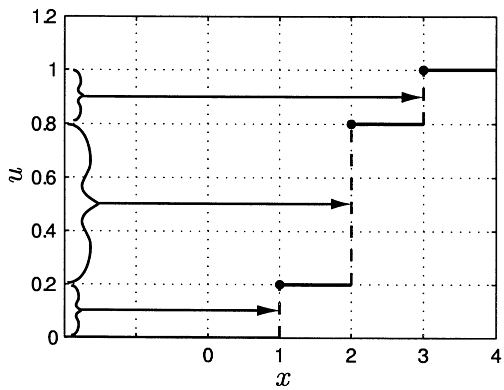


Figure 5.14: Relationship of inverse CDF to generation of discrete random variable. Value of u is mapped into value of x.

Although little more than a curiousity now, it will become important when we simulate continuous random variables in Chapter 10.

5.10 Real-World Example – Servicing Customers

A standard problem in many disciplines is the allocation of resources to service customers. It occurs in determining the number of cashiers needed at a store, the computer capacity needed to service download requests, and the amount of equipment necessary to service phone customers, as examples. In order to service these customers in a timely manner, it is necessary to know the distribution of arrival times of their requests. Since this will vary depending on many factors such as time of day, popularity of a file request, etc., the best we can hope for is a determination of the probabilities of these arrivals. As we will see shortly, the Poisson probability PMF is particularly suitable as a model. We now focus on the problem of determining the number of cashiers needed in a supermarket.

A supermarket has one express lane open from 5 to 6 PM on weekdays (Monday through Friday). This time of the day is usually the busiest since people tend to stop on their way home from work to buy groceries. The number of items allowed in the express lane is limited to 10 so that the average time to process an order is fairly constant at about 1 minute. The manager of the supermarket notices that there is frequently a long line of people waiting and hears customers grumbling about the wait. To improve the situation he decides to open additional express lanes during this time period. If he does, however, he will have to "pull" workers from other jobs around the store to serve as cashiers. Hence, he is reluctant to open more lanes than necessary. He hires Professor Poisson to study the problem and tell him how many lanes should be opened. The manager tells Professor Poisson that there should be no more than one person waiting in line 95% of the time. Since the processing time is 1 minute, there can be at most two arrivals in each time slot of 1 minute length. He reasons that one will be immediately serviced and the other will only have to wait a maximum of 1 minute. After a week of careful study, Professor Poisson tells the manager to open two lanes from 5 to 6 PM. Here is his reasoning.

First Professor Poisson observes the arrivals of customers in the express lane on a Monday from 5 to 6 PM. The observed arrivals are shown in Figure 5.15, where the arrival times are measured in seconds. On Monday there are a total of

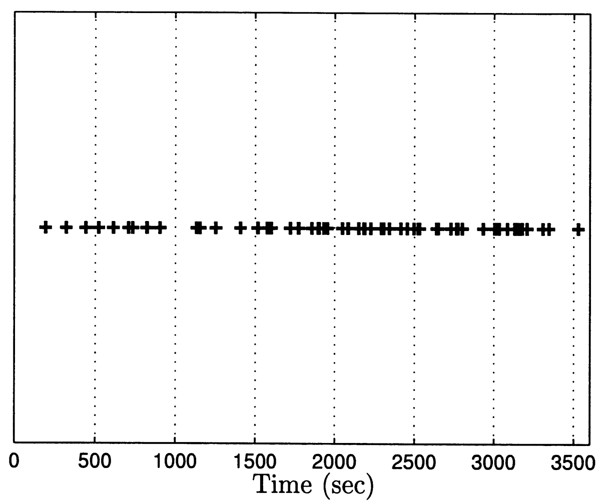


Figure 5.15: Arrival times at one express lane on Monday (a '+' indicates an arrival).

80 arrivals. He repeats his experiment on the following 4 days (Tuesday through Friday) and notes total arrivals of 68, 70, 59, and 66 customers, respectively. On the average there are 68.6 arrivals, which he rounds up to 70. Thus, the arrival rate is 1.167 customers per minute. He then likens the arrival process to one in which the 5 to 6 PM time interval is broken up into 3600 time slots of 1 second each. He reasons that there is at most 1 arrival in a given time slot and there may be no arrivals in that time slot. (This of course would not be valid if for instance, two friends did their shopping together and arrived at the same time.) Hence, Professor Poisson reasons that a good arrival model is a sequence of independent Bernoulli trials, where 0 indicates no arrival and 1 indicates an arrival in each 1-second time slot. The probability p of a 1 is estimated from his observed data as the number of arrivals from 5 to 6 PM divided by the total number of time slots in seconds. This yields $\hat{p} = 70/3600 = 0.0194$ for each 1-second time slot. Instead of using the binomial PMF to describe the number of arrivals in each 1-minute time slot (for which $p = 0.0194$ and $M = 60$), he decides to approximate it using his favorite PMF, the Poisson model. Therefore, the probability of k arrivals (or successes) in a time interval of 60 seconds would be

$$p_{X_1}[k] = \exp(-\lambda_1)\frac{\lambda_1^k}{k!} \qquad k = 0, 1, \ldots \tag{5.17}$$

where the subscripts on X and λ are meant to remind us that we will initially consider the arrivals at *one* express lane. The value of λ_1 to be used is $\lambda_1 = Mp$, which is estimated as $\hat{\lambda}_1 = M\hat{p} = 60(70/3600) = 7/6$. This represents the expected number of customers arriving in the 1-minute interval. According to the manager's requirements, within this time interval there should be at most 2 customers arriving 95% of the time. Hence, we require that

$$P[X_1 \leq 2] = \sum_{k=0}^{2} p_{X_1}[k] \geq 0.95.$$

But from (5.17) this becomes

$$P[X_1 \leq 2] = \exp(-\lambda_1)\left(1 + \lambda_1 + \frac{1}{2}\lambda_1^2\right) = 0.88$$

using $\lambda_1 = 7/6$. Hence, the probability of 2 or fewer customers arriving at the express lane is not greater than 0.95. If a second express lane is opened, then the average number of arrivals at each lane during the 1-minute time interval will be halved to 35. Therefore, the Poisson PMF for the number of arrivals at each lane will be characterized by $\lambda_2 = 7/12$. Now, however, there are two lanes and two sets of arrivals. Since the arrivals are modeled as *independent* Bernoulli trials, we can assert that

$$\begin{aligned} P[\text{2 or fewer arrivals at both lanes}] &= P[\text{2 or fewer arrivals at lane 1}] \\ &\quad \cdot P[\text{2 or fewer arrivals at lane 2}] \\ &= P[\text{2 or fewer arrivals at lane 1}]^2 \\ &= P[X_1 \leq 2]^2 \end{aligned}$$

so that

$$\begin{aligned} P[\text{2 or fewer arrivals at both lanes}] &= \left(\sum_{k=0}^{2} p_{X_1}[k]\right)^2 \\ &= \left[\exp(-\lambda_2)\left(1 + \lambda_2 + \frac{1}{2}\lambda_2^2\right)\right]^2 = 0.957 \end{aligned}$$

which meets the requirement. An example is shown for one of the two express lanes with an average number of customer arrivals per minute of 7/12 in Figures 5.16 and 5.17, with the latter an expanded version of the former. The dashed vertical lines

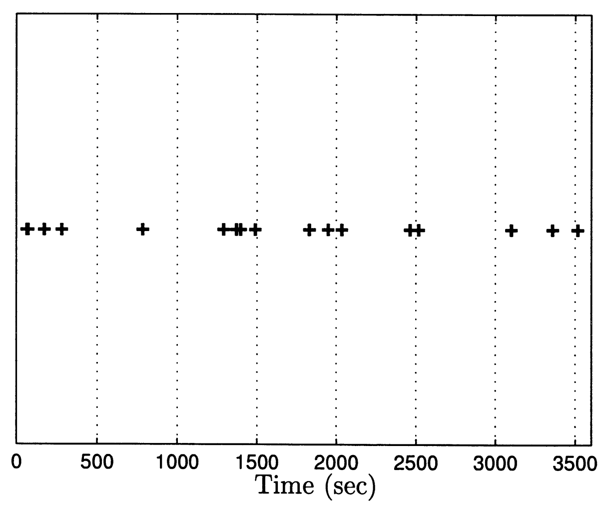


Figure 5.16: Arrival times at one of the two express lanes (a '+' indicates an arrival).

in Figure 5.17 indicate 1-minute intervals. There are no 1-minute intervals with more than 2 arrivals, as we expect.

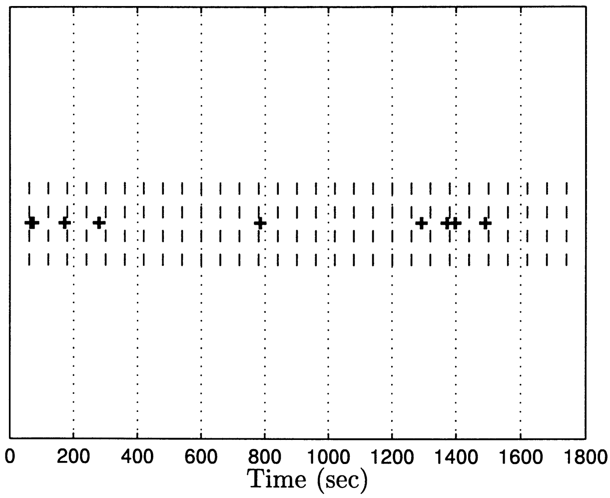


Figure 5.17: Expanded version of Figure 5.16 (a '+' indicates an arrival). Time slots of 60 seconds are shown by dashed lines.

References

Ross, S., *A First Course in Probability*, Prentice-Hall, Upper Saddle River, NJ, 2002.

Problems

5.1 (w) Draw a picture depicting a mapping of the outcome of a die toss, i.e., the pattern of dots that appear, to the numbers $1,2,3,4,5,6$.

5.2 (w) Repeat Problem 5.1 for a mapping of the sides that display 1, 2, or 3 dots to the number 0 and the remaining sides to the number 1.

5.3 (w) Consider a random experiment for which $\mathcal{S} = \{s_i : s_i = i,\ i = 1,2,\ldots,10\}$ and the outcomes are equally likely. If a random variable is defined as $X(s_i) = s_i^2$, find $\mathcal{S}_X$ and the PMF.

5.4 (⌣̈) **(w)** Consider a random experiment for which $\mathcal{S} = \{s_i : s_i = -3,-2,-1,0,1,2,3\}$ and the outcomes are equally likely. If a random variable is defined as $X(s_i) = s_i^2$, find $\mathcal{S}_X$ and the PMF.

5.5 (w) A man is late for his job by $s_i = i$ minutes, where $i = 1,2,\ldots$. If $P[s_i] = (1/2)^i$ and he is fined \$0.50 per minute, find the PMF of his fine. Next find the probability that he will be fined more than \$10.

5.6 (⌣) **(w)** If $p_X[k] = \alpha p^k$ for $k = 2, 3, \ldots$ is to be a valid PMF, what are the possible values for α and p?

5.7 (t) The maximum value of the binomial PMF occurs for the unique value $k = [(M+1)p]$, where $[x]$ denotes the largest integer less than or equal to x, if $(M+1)p$ is not an integer. If, however, $(M+1)p$ is an integer, then the PMF will have the same maximum value at $k = (M+1)p$ and $k = (M+1)p - 1$. For the latter case when $(M+1)p$ is an integer you are asked to prove this result. To do so first show that

$$p_X[k]/p_X[k-1] = 1 + \frac{(M+1)p - k}{k(1-p)}.$$

5.8 (⌣) **(w)** At a party a large barrel is filled with 99 gag gifts and 1 diamond ring, all enclosed in identical boxes. Each person at the party is given a chance to pick a box from the barrel, open the box to see if the diamond is inside, and if not, to close the box and return it to the barrel. What is the probability that at least 19 persons will choose gag gifts before the diamond ring is selected?

5.9 (f,c) If X is a geometric random variable with $p = 0.25$, what is the probability that $X \geq 4$? Verify your result by performing a computer simulation.

5.10 (c) Using a computer simulation to generate a geom(0.25) random variable, determine the average value for a large number of realizations. Relate this to the value of p and explain the results.

5.11 (t) Prove that the maximum value of a Poisson PMF occurs at $k = [\lambda]$. Hint: See Problem 5.7 for the approach.

5.12 (w,c) If $X \sim \text{Pois}(\lambda)$, plot $P[X \geq 2]$ versus λ and explain your results.

5.13 (⌣) **(c)** Use a computer simulation to generate realizations of a Pois(λ) random variable with $\lambda = 5$ by approximating it with a bin(100,0.05) random variable. What is the average value of X?

5.14 (⌣) **(w)** If $X \sim \text{bin}(100, 0.01)$, determine $p_X[5]$. Next compare this to the value obtained using a Poisson approximation.

5.15 (t) Prove the following limit:

$$\lim_{M\to\infty} g(M) = \lim_{M\to\infty} \left(1 + \frac{x}{M}\right)^M = \exp(x).$$

To do so note that the same limit is obtained if M is replaced by a continuous variable, say u, and that one can consider $\ln g(u)$ since the logarithm is a continuous function. Hint: Use L'Hospital's rule.

5.16 (f,c) Compare the PMFs for Pois(1) and bin(100,0.01) random variables.

5.17 (c) Generate realizations of a Pois(1) random variable by using a binomial approximation.

5.18 (⌣̈) (c) Compare the theoretical value of $P[X = 3]$ for the Poisson random variable to the estimated value obtained from the simulation of Problem 5.17.

5.19 (f) If $X \sim \text{Ber}(p)$, find the PMF for $Y = -X$.

5.20 (⌣̈) (f) If $X \sim \text{Pois}(\lambda)$, find the PMF for $Y = 2X$.

5.21 (f) A discrete random variable X has the PMF

$$p_X[x_i] = \begin{cases} \frac{1}{2} & x_1 = -1 \\ \frac{1}{4} & x_2 = -\frac{1}{2} \\ \frac{1}{8} & x_3 = 0 \\ \frac{1}{16} & x_4 = \frac{1}{2} \\ \frac{1}{16} & x_5 = 1. \end{cases}$$

If $Y = \sin \pi X$, find the PMF for Y.

5.22 (t) In this problem we derive the Taylor expansion for the function $g(x) = \exp(x)$. To do so note that the expansion about the point $x = 0$ is given by

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} x^n$$

where $g^{(0)}(0) = g(0)$ and $g^{(n)}(0)$ is the nth derivative of $g(x)$ evaluated at $x = 0$. Prove that it is given by

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

5.23 (f) Plot the CDF for

$$p_X[k] = \begin{cases} \frac{1}{4} & k = 1 \\ \frac{1}{2} & k = 2 \\ \frac{1}{4} & k = 3. \end{cases}$$

5.24 (w) A horizontal bar of negligible weight is loaded with three weights as shown in Figure 5.18. Assuming that the weights are concentrated at their center locations, plot the total mass of the bar starting at the left end (where $x = 0$ meters) to any point on the bar. How does this relate to a PMF and a CDF?

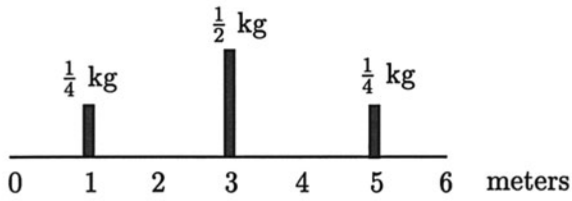


Figure 5.18: Weightless bar supporting three weights.

5.25 (f) Find and plot the CDF of $Y = -X$ if $X \sim \text{Ber}(\frac{1}{4})$.

5.26 ($\ddot\smile$) (w) Find the PMF if X is a discrete random variable with the CDF

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{\lfloor x \rfloor}{5} & 0 \le x \le 5 \\ 1 & x > 5. \end{cases}$$

5.27 (w) Is the following a valid CDF? If not, why not, and how could you modify it to become a valid one?

$$F_X(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{2} & 2 \le x \le 3 \\ \frac{3}{4} & 3 < x \le 4 \\ 1 & x \ge 4. \end{cases}$$

5.28 ($\ddot\smile$) (f) If X has the CDF shown in Figure 5.11b, determine $P[2 \le X \le 4]$ from the CDF.

5.29 (t) Prove that the function $g(x) = \exp(x)$ is a monotonically increasing function by showing that $g(x_2) \ge g(x_1)$ if $x_2 \ge x_1$.

5.30 (c) Estimate the PMF for a geom(0.25) random variable for $k = 1, 2, \ldots, 20$ using a computer simulation and compare it to the true PMF. Also, estimate the CDF from your computer simulation.

5.31 ($\ddot\smile$) (f,c) The arrival rate of calls at a mobile switching station is 1 per second. The probability of k calls in a T second interval is given by a Poisson PMF with λ = arrival rate $\times$ T. What is the probability that there will be more than 100 calls placed in a 1-minute interval?