Chapter 5

Discrete Random Variables

Suppose that an experiment and a sample space are given. A *random variable* is a real-valued function of the outcome of the experiment. In other words, the random variable assigns a specific number to every possible outcome of the experiment. The numerical value of a particular outcome is simply called the *value* of the random variable. Because of the structure of real numbers, it is possible to define pertinent statistical properties on random variables that otherwise do not apply to probability spaces in general.
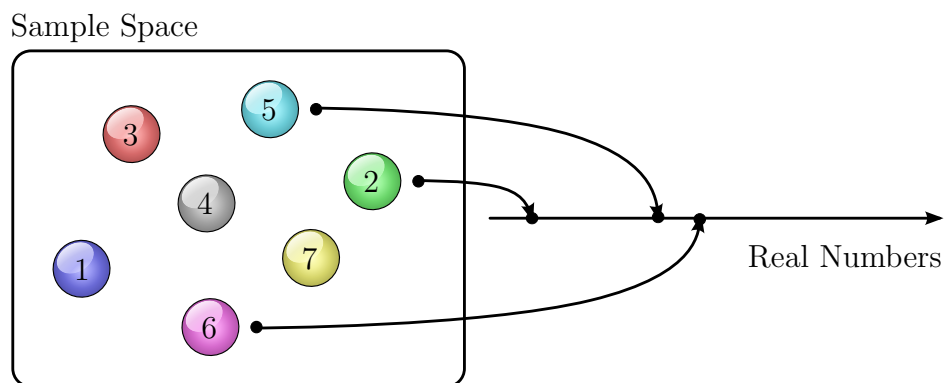
Figure 5.1: The sample space in this example has seven possible outcomes. A random variable maps each of these outcomes to a real number.

Example 33. *There are six possible outcomes to the rolling of a fair die, namely each of the six faces. These faces map naturally to the integers one*

through six.   The value of the random variable, in this case, is simply the number of dots that appear on the top face of the die.
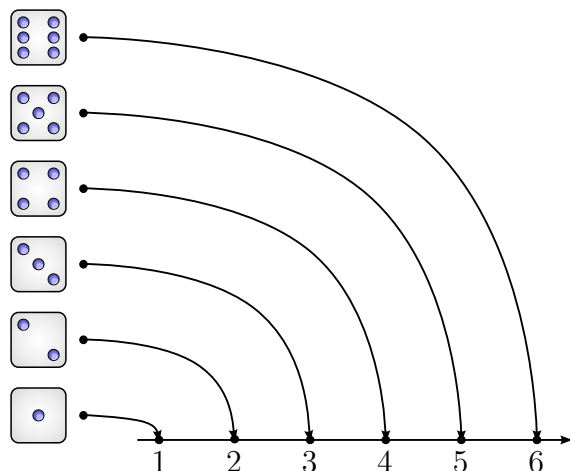

Figure 5.2: This random variable takes its input from the rolling of a die and assigns to each outcome a real number that corresponds to the number of dots that appear on the top face of the die.

A simple class of random variables is the collection of *discrete random variables*. A variable is called discrete if its range is finite or countably infinite; that is, it can only take a finite or countable number of values.

Example 34. *Consider the experiment where a coin is tossed repetitively until heads is observed.   The corresponding function, which maps the number of tosses to an integer, is a discrete random variable that takes a countable number of values.   The range of this random variable is given by the positive integers* $\{1, 2, \ldots\}$.

5.1   Probability Mass Functions

A discrete random variable X is characterized by the probability of each of the elements in its range. We identify the probabilities of individual elements in the range of X using the *probability mass function (PMF)* of X, which we denote by $p_X(\cdot)$. If x is a possible value of X then the *probability mass* of x,

written $p_X(x)$, is defined by

$$p_X(x) = \Pr(\{X = x\}) = \Pr(X = x). \tag{5.1}$$

Equivalently, we can think of $p_X(x)$ as the probability of the set of all outcomes in Ω for which X is equal to x,

$$p_X(x) = \Pr(X^{-1}(x)) = \Pr(\{\omega \in \Omega | X(\omega) = x\}).$$

Here, $X^{-1}(x)$ denotes the *preimage* of x defined by $\{\omega \in \Omega | X(\omega) = x\}$. This is not to be confused with the inverse of a bijection.
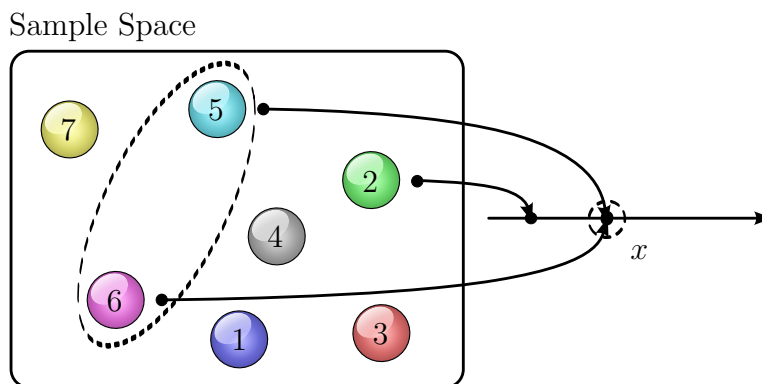

Figure 5.3: The probability mass of x is given by the probability of the set of all outcomes which X maps to x.

Let $X(\Omega)$ denote the collection of all the possible numerical values X can take; this set is known as the range of X. Using this notation, we can write

$$\sum_{x \in X(\Omega)} p_X(x) = 1. \tag{5.2}$$

We emphasize that the sets defined by $\{\omega \in \Omega | X(\omega) = x\}$ are disjoint and form a partition of the sample space Ω, as x ranges over all the possible values in $X(\Omega)$. Thus, (5.2) follows immediately from the countable additivity axiom and the normalization axiom of probability laws. In general, if X is a discrete random variable and S is a subset of $X(\Omega)$, we can write

$$\Pr(S) = \Pr\left(\{\omega \in \Omega | X(\omega) \in S\}\right) = \sum_{x \in S} p_X(x). \tag{5.3}$$

This equation offers an explicit formula to compute the probability of any subset of $X(\Omega)$, provided that X is discrete.

Example 35. *An urn contains three balls numbered one, two and three. Two balls are drawn from the urn without replacement. We wish to find the probability that the sum of the two selected numbers is odd.*

Let Ω be the set of ordered pairs corresponding to the possible outcomes of the experiment, $\Omega = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$. Note that these outcomes are equiprobable. We employ X to represent the sum of the two selected numbers. The PMF of random variable X is given by

$$p_X(3) = p_X(4) = p_X(5) = \frac{1}{3}.$$

If S denotes the event that the sum of the two numbers is odd, then the probability of the sum being odd can be computed as follows,

$$\Pr(S) = \Pr(\{3, 5\}) = p_X(3) + p_X(5) = \frac{2}{3}.$$

5.2   Important Discrete Random Variables

A number of discrete random variables appears frequently in problems related to probability. These random variables arise in many different contexts, and they are worth getting acquainted with. In general, discrete random variables occur primarily in situations where counting is involved.

5.2.1   Bernoulli Random Variables

The first and simplest random variable is the *Bernoulli random variable*. Let X be a random variable that takes on only two possible numerical values, $X(\Omega) = \{0, 1\}$. Then, X is a Bernoulli random variable and its PMF is given by

$$p_X(x) = \begin{cases} 1 - p, & \text{if } x = 0 \\ p, & \text{if } x = 1 \end{cases}$$

where $p \in [0, 1]$.

Example 36. *Consider the flipping of a biased coin, for which heads is obtained with probability p and tails is obtained with probability $1 - p$. A random variable that maps heads to one and tails to zero is a Bernoulli random variable with parameter p. In fact, every Bernoulli random variable is equivalent to the tossing of a coin.*
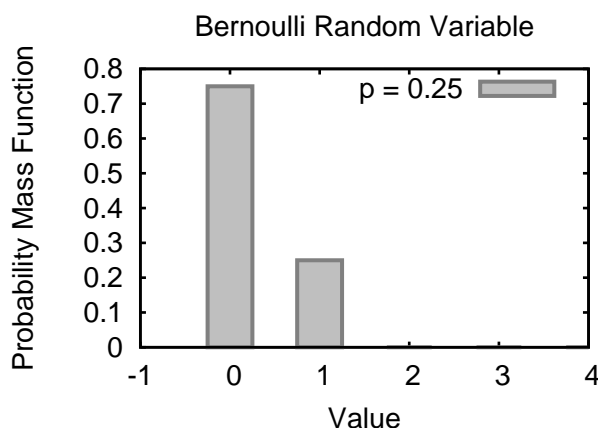
Figure 5.4: The PMF of a Bernoulli random variable appears above for parameter $p = 0.25$.

5.2.2 Binomial Random Variables

Multiple independent Bernoulli random variables can be combined to construct more sophisticated random variables. Suppose X is the sum of n independent and identically distributed Bernoulli random variables. Then X is called a *binomial random variable* with parameters n and p. The PMF of X is given by

$$p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $k = 0, 1, \ldots n$. We can easily verify that X fulfills the normalization axiom,

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Example 37. *The Brazos Soda Company creates an "Under the Caps" promotion whereby a customer can win an instant cash prize of \$1 by looking under a bottle cap. The likelihood to win is one in four, and it is independent from bottle to bottle. A customer buys eight bottles of soda from this company. We wish to find the PMF of the number of winning bottles, which we denote by X. Also, we want to compute the probability of winning more than \$4.*

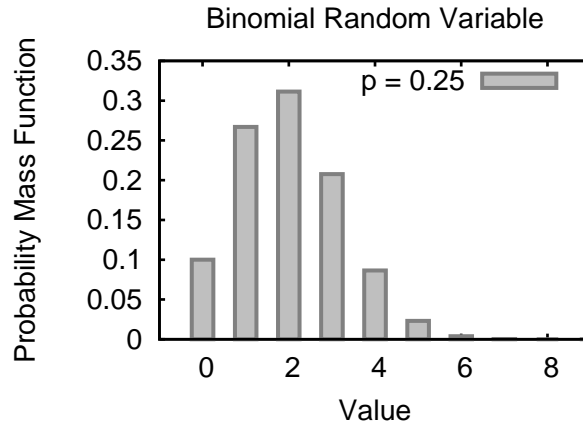*The random variable X is binomial with parameters $n = 8$ and $p = 1/4$.*

Figure 5.5: This figure shows a binomial random variable with parameters $n = 8$ and $p = 0.25$.

Its PMF is given by

$$p_X(k) = \binom{8}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{8-k} = \binom{8}{k} \frac{3^{8-k}}{4^8},$$

where $k = 0, 1, \ldots, 8$. The probability of winning more than \$4 is

$$\Pr(X > 4) = \sum_{k=5}^{8} \binom{8}{k} \frac{3^{8-k}}{4^8}.$$

5.2.3   Poisson Random Variables

The probability mass function of a *Poisson random variable* is given by

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

where λ is a positive number. Note that, using Taylor series expansion, we have

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1,$$

which shows that this PMF fulfills the normalization axiom of probability laws. The Poisson random variable is of fundamental importance when counting the number of occurrences of a phenomenon within a certain time period. It finds extensive use in networking, inventory management and queueing applications.
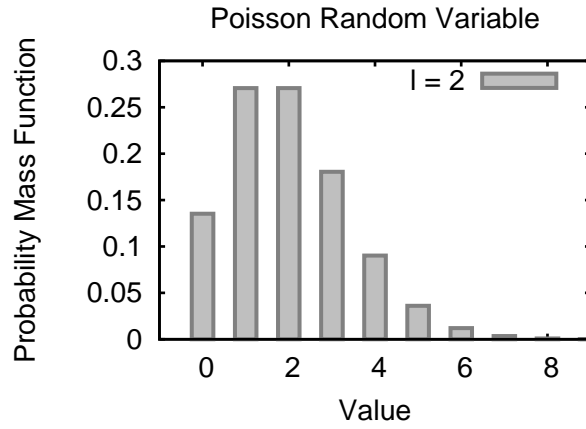
Figure 5.6: This figure shows the PMF of a Poisson random variable with parameter $\lambda = 2$. Note that the values of the PMF are only present for $k = 0, 1, \ldots, 8$.

Example 38. *Requests at an Internet server arrive at a rate of λ connections per second. The number of service requests per second is modeled as a random variable with a Poisson distribution. We wish to find the probability that no service requests arrive during a time interval of one second.*

Let N be a random variable that represents the number of requests that arrives within a span of one second. By assumption, N is a Poisson random variable with PMF

$$p_N(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

The probability that no service requests arrive in one second is simply given by $p_N(0) = e^{-\lambda}$.

It is possible to obtain a Poisson random variable as the limit of a sequence of binomial random variables. Fix λ and let $p_n = \lambda/n$. For $k = 1, 2, \ldots n$, we define the PMF of the random variable X_n as

$$p_{X_n}(k) = \Pr(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n!}{n^k(n-k)!} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

In the limit, as n approaches infinity, we get

$$\lim_{n\to\infty} p_{X_n}(k) = \lim_{n\to\infty} \frac{n!}{n^k(n-k)!}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Thus, the sequence of binomial random variables $\{X_n\}$ converges in distribution to the Poisson random variable X.
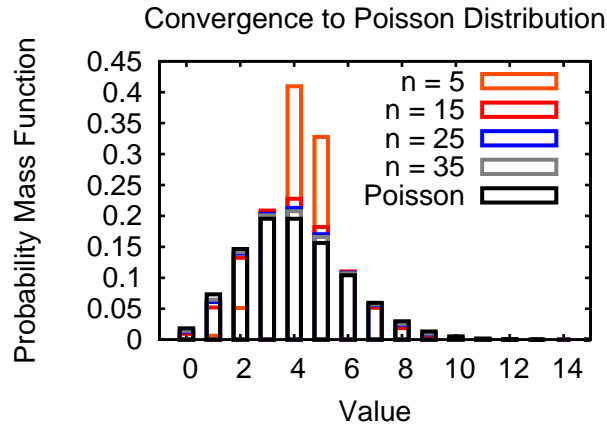

Figure 5.7: The levels of a binomial PMF with parameter $p = \lambda/n$ converge to the probabilities of a Poisson PMF with parameter λ as n increases to infinity.

This discussion suggests that the Poisson PMF can be employed to approximate the PMF of a binomial random variable in certain circumstances. Suppose that Y is a binomial random variable with parameters n and p. If n is large and p is small then the probability that Y equals k can be approximated by

$$p_Y(k) = \frac{n!}{n^k(n-k)!}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda},$$

where $\lambda = np$. The latter expression can be computed numerically in a straightforward manner.

Example 39. *The probability of a bit error on a communication channel is equal to 10^{-2}. We wish to approximate the probability that a block of 1000 bits has four or more errors.*

Assume that the probability of individual errors is independent from bit to bit. The transmission of each bit can be modeled as a Bernoulli trial, with a

zero indicating a correct transmission and a one representing a bit error. The total number of errors in 1000 *transmissions then corresponds to a binomial random variable with parameters* $n = 1000$ *and* $p = 10^{-2}$. *The probability of making four or more errors can be approximated using a Poisson random variable with constant* $\lambda = np = 10$. *Thus, we can approximate the probability that a block of* 1000 *bits has four or more errors by*

$$\Pr(N \geq 4) = 1 - \Pr(N < 4) \approx 1 - \sum_{k=0}^{3} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= 1 - e^{-10} \left(1 + 10 + 50 + \frac{500}{3} \right)$$

$$\approx 0.9897.$$

This is in contrast to the exact answer, which can be written as 0.9899 *when truncated to four decimal places.*

5.2.4 Geometric Random Variables

Consider a random experiment where a Bernoulli trial is repeated multiple times until a one is observed. At each time step, the probability of getting a one is equal to p and the probability of getting a zero is $1 - p$. The number of trials carried out before completion, which we denote by X, is recorded as the outcome of this experiment. The random variable X is a *geometric random variable*, and its PMF is given by

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots$$

We stress that $(1 - p)^{k-1} p$ simply represents the probability of obtaining a sequence of $k - 1$ zero immediately followed by a one.

Example 40. *The Brazos Soda Company introduces another "Under the Caps" promotion. This time, a customer can win an additional bottle of soda by looking under the cap of her bottle. The probability to win is* 1/5, *and it is independent from bottle to bottle. A customer purchases one bottle of soda from the Brazos Soda Company and thereby enters the contest. For every extra bottle of soda won by looking under the cap, the customer gets an additional chance*
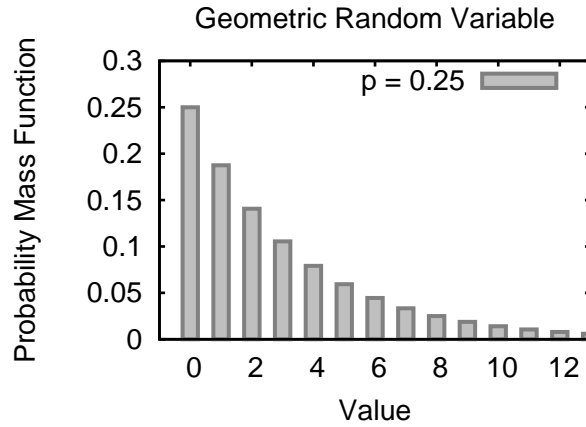
Figure 5.8: The PMF of a geometric random variable is a decreasing function of k. It is plotted above for $p = 0.25$. The values of the PMF are only present for $k = 1, 2, \ldots, 12$.

to play. We wish to find the PMF of the number of bottles obtained by this customer.

Let X denote the total number of bottles obtained by the customer. The random variable X is geometric and its PMF is

$$p_X(k) = \left(\frac{1}{5}\right)^{k-1} \frac{4}{5},$$

where $k = 1, 2, \ldots$

Memoryless Property: A remarkable aspect of the geometric random variable is that it satisfies the *memoryless property*,

$$\Pr(X = k + j | X > k) = \Pr(X = j).$$

This can be established using the definition of conditional probability. Let X be a geometric random variable with parameter p, and assume that k and j are positive integers. We can write the conditional probability of X as

$$\Pr(X = k + j | X > k) = \frac{\Pr(\{X = k + j\} \cap \{X > k\})}{\Pr(X > k)}$$

$$= \frac{\Pr(X = k + j)}{\Pr(X > k)} = \frac{(1-p)^{k+j-1}p}{(1-p)^k}$$

$$= (1-p)^{j-1}p = \Pr(X = j).$$

In words, the probability that the number of trials carried out before completion is $k + j$, given k unsuccessful trials, is equal to the unconditioned probability that the total number of trials is j. It can be shown that the geometric random variable is the only discrete random variable that possesses the memoryless property.

5.2.5 Discrete Uniform Random Variables

A finite random variable where all the possible outcomes are equally likely is called a *discrete uniform random variable.* Let X be a uniform random variable taking value over $X(\Omega) = \{1, 2, \ldots, n\}$. Its PMF is therefore given by

$$p_X(k) = \begin{cases} 1/n, & \text{if } k = 1, 2, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

We encountered specific cases of this random variable before. The tossing of a fair coin and the rolling of a die can both be used to construct uniform random variables.
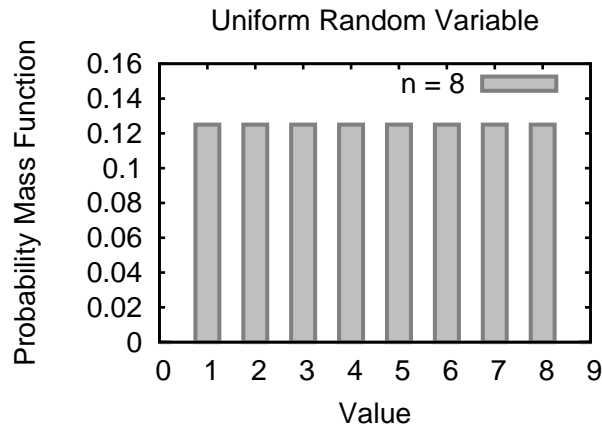

Figure 5.9: A uniform random variable corresponds to the situation where all the possible values of the random variable are equally likely. It is shown above for the case where $n = 8$.

5.3 Functions of Random Variables

Recall that a random variable is a function of the outcome of an experiment. Given a random variable X, it is possible to create a new random variable Y by applying a real-valued function $g(\cdot)$ to X. If X is a random variable then $Y = g(X)$ is also a random variable since it associates a numerical value to every outcome of the experiment. In particular, if $\omega \in \Omega$ is the outcome of the experiment, then X takes on value $x = X(\omega)$ and Y takes on value $Y(\omega) = g(x)$.
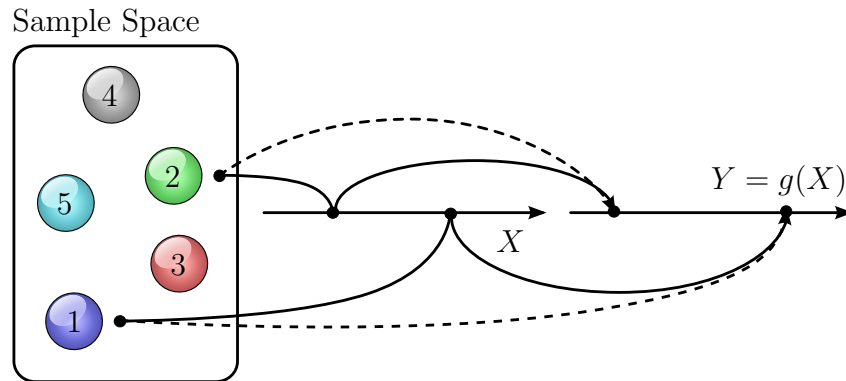

Figure 5.10: A real-valued function of a random variable is a random variable itself. In this figure, Y is obtained by passing random variable X through a function $g(\cdot)$.

Furthermore, if X is a discrete random variable, then so is Y. The set of possible values Y can take is denoted by $g(X(\Omega))$, and the number of elements in $g(X(\Omega))$ is no greater than the number of elements in $X(\Omega)$. The PMF of Y, which we represent by $p_Y(\cdot)$, is obtained as follows. If $y \in g(X(\Omega))$ then

$$p_Y(y) = \sum_{\{x \in X(\Omega) | g(x) = y\}} p_X(x); \tag{5.4}$$

otherwise, $p_Y(y) = 0$. In particular, $p_Y(y)$ is non-zero only if there exists an $x \in X(\Omega)$ such that $g(x) = y$ and $p_X(x) > 0$.

Example 41. *Let X be a random variable and let $Y = g(X) = aX + b$, where a and b are constant. That is, Y is an affine function of X. Suppose that*

$a \neq 0$, then the probability of Y being equal to value y is given by

$$p_Y(y) = p_X\left(\frac{y-b}{a}\right).$$

Linear and affine functions of random variables are commonplace in applied probability and engineering.

Example 42. *A taxi driver works in New York City. The distance of a ride with a typical client is a discrete uniform random variable taking value over $X(\Omega) = \{1, 2, \ldots, 10\}$. The metered rate of fare, according to city rules, is $2.50 upon entry and $0.40 for each additional unit (one-fifth of a mile). We wish to find the PMF of Y, the value of a typical fare.*

Traveling one mile is five units and costs an extra $2.00. The smallest possible fare is therefore $4.50, with probability 1/10. Similarly, a ride of X miles will generate a revenue of $Y = 2.5 + 2X$ dollars for the cab driver. The PMF of Y is thus given by

$$p_Y(2.5 + 2k) = \frac{1}{10}$$

for $k = 1, 2, \ldots, 10$; and it is necessarily zero for any other argument.

Further Reading

1. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Sections 4.1–4.2, 4.6–4.8.

2. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Sections 2.1–2.3.

3. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Section 2.7.

4. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Sections 2.1–1.2.