2 Conditional Probability and Conditional Expectation

2.1 The Discrete Case

The conditional probability $\Pr\{A|B\}$ of the event A given the event B is defined by

$$\Pr\{A|B\} = \frac{\Pr\{A \text{ and } B\}}{\Pr\{B\}} \quad \text{if } \Pr\{B\} > 0, \tag{2.1}$$

and is not defined, or is assigned an arbitrary value, when $\Pr\{B\} = 0$. Let X and Y be random variables that can attain only countably many different values, say $0, 1, 2, \ldots$. The *conditional probability mass function* $p_{X|Y}(x|y)$ of X given $Y = y$ is defined by

$$p_{X|Y}(x|y) = \frac{\Pr\{X = x \text{ and } Y = y\}}{\Pr\{Y = y\}} \quad \text{if } \Pr\{Y = y\} > 0,$$

and is not defined, or is assigned an arbitrary value, whenever $\Pr\{Y = y\} = 0$. In terms of the joint and marginal probability mass functions $p_{XY}(x, y)$ and $p_Y(y) = \Sigma_x p_{XY}(x, y)$, respectively, the definition is

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} \quad \text{if } p_Y(y) > 0; \quad x, y = 0, 1, \ldots. \tag{2.2}$$

Observe that $p_{X|Y}(x|y)$ is a probability mass function in x for each fixed y, i.e., $p_{X|Y}(x|y) \geq 0$ and $\Sigma_\xi p_{X|Y}(\xi|y) = 1$, for all x, y.

The law of total probability takes the form

$$\Pr\{X = x\} = \sum_{y=0}^{\infty} p_{X|Y}(x|y) p_Y(y). \tag{2.3}$$

Notice in (2.3) that the points y where $p_{X|Y}(x|y)$ is not defined are exactly those values for which $p_Y(y) = 0$, and hence, do not affect the computation. The lack of a complete prescription for the conditional probability mass function, a nuisance in some instances, is always consistent with subsequent calculations.

Example Let X have a binomial distribution with parameters p and N, where N has a binomial distribution with parameters q and M. What is the marginal distribution of X?

An Introduction to Stochastic Modeling. DOI: 10.1016/B978-0-12-381416-6.00002-2

We are given the conditional probability mass function

$$p_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

and the marginal distribution

$$p_N(n) = \binom{M}{n} q^n (1-q)^{M-n}, \quad n = 0, 1, \ldots, M.$$

We apply the law of total probability in the form of (2.3) to obtain

$$\begin{aligned}
\Pr\{X = k\} &= \sum_{n=0}^{M} p_{X|N}(k|n) p_N(n) \\
&= \sum_{n=k}^{M} \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \frac{M!}{n!\,(M-n)!} q^n (1-q)^{M-n} \\
&= \frac{M!}{k!} p^k (1-q)^M \left(\frac{q}{1-q}\right)^k \sum_{n=k}^{M} \frac{1}{(n-k)!\,(M-n)!} (1-p)^{n-k} \\
&\quad \times \left(\frac{q}{1-q}\right)^{n-k} \\
&= \frac{M!}{k!\,(M-k)!} (pq)^k (1-q)^{M-k} \left[1 + \frac{q(1-p)}{1-q}\right]^{M-k} \\
&= \frac{M!}{k!\,(M-k)!} (pq)^k (1-pq)^{M-k}, \quad k = 0, 1, \ldots, M.
\end{aligned}$$

In words, X has a binomial distribution with parameters M and pq.

Example Suppose X has a binomial distribution with parameters p and N, where N has a Poisson distribution with mean λ. What is the marginal distribution for X?

Proceeding as in the previous example but now using

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

we obtain

$$\begin{aligned}
\Pr\{X = k\} &= \sum_{n=0}^{\infty} p_{X|N}(k|n) p_N(n) \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n e^{-\lambda}}{n!}
\end{aligned}$$

$$= \frac{\lambda^k e^{-\lambda} p^k}{k!} \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!}$$

$$= \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda(1-p)}$$

$$= \frac{(\lambda p)^k e^{-\lambda p}}{k!} \quad \text{for } k = 0, 1, \ldots.$$

In words, X has a Poisson distribution with mean λp.

Example Suppose X has a negative binomial distribution with parameters p and N, where N has the geometric distribution

$$p_N(n) = (1-\beta)\beta^{n-1} \quad \text{for } n = 1, 2, \ldots.$$

What is the marginal distribution for X?

We are given the conditional probability mass function

$$p_{X|N}(k|n) = \binom{n+k-1}{k} p^n (1-p)^k, \quad k = 0, 1, \ldots.$$

Using the law of total probability, we obtain

$$\Pr\{X = k\} = \sum_{n=0}^{\infty} p_{X|N}(k|n) p_N(n)$$

$$= \sum_{n=1}^{\infty} \frac{(n+k-1)!}{k!\,(n-1)!} p^n (1-p)^k (1-\beta)\beta^{n-1}$$

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

$$= (1-\beta)(1-p)^k p (1-\beta p)^{-k-1}$$

$$= \left(\frac{p-\beta p}{1-\beta p}\right)\left(\frac{1-p}{1-\beta p}\right)^k \quad \text{for } k = 0, 1, \ldots.$$

We recognize the marginal distribution of X as being of geometric form.

Let g be a function for which the expectation of $g(X)$ is finite. We define the *conditional* expected value of $g(X)$ given $Y = y$ by the formula

$$E[g(X)|Y = y] = \sum_x g(x) p_{X|Y}(x|y) \quad \text{if } p_Y(y) > 0, \tag{2.4}$$

and the conditional mean is not defined at values y for which $p_Y(y) = 0$. The law of total probability for conditional expectation reads

$$E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y). \tag{2.5}$$

The conditional expected value $E[g(X)|Y = y]$ is a function of the real variable y. If we evaluate this function at the random variable Y, we obtain a random variable that we denote by $E[g(X)|Y]$. The law of total probability in (2.5) now may be written in the form

$$E[g(X)] = E\{E[g(X)|Y]\}. \tag{2.6}$$

Since the conditional expectation of $g(X)$ given $Y = y$ is the expectation with respect to the conditional probability mass function $p_{X|Y}(x|y)$, conditional expectations behave in many ways like ordinary expectations. The following list summarizes some properties of conditional expectations. In this list, with or without affixes, X and Y are jointly distributed random variables; c is a real number; g is a function for which $E[|g(X)|] < \infty$; h is a bounded function; and v is a function of two variables for which $E[|v(X, Y)|] < \infty$. The properties are

1. $E[c_1 g_1(X_1) + c_2 g_2(X_2)|Y = y]$
$$= c_1 E[g_1(X_1)|Y = y] + c_2 E[g_2(X_2)|Y = y]. \tag{2.7}$$

2. if $g \geq 0$, then $E[g(X)|Y = y] \geq 0$. (2.8)

3. $E[v(X, Y)|Y = y] = E[v(X, y)|Y = y]$. (2.9)

4. $E[g(X)|Y = y] = E[g(X)]$ if X and Y are independent. (2.10)

5. $E[g(X)h(Y)|Y = y] = h(y)E[g(X)|Y = y]$. (2.11)

6. $E[g(X)h(Y)] = \sum_y h(y)E[g(X)|Y = y]p_Y(y)$
$$= E\{h(Y)E[g(X)|Y]\}. \tag{2.12}$$

As a consequence of (2.7), (2.11), and (2.12), with either $g \equiv 1$ or $h \equiv 1$, we obtain

$$E[c|Y = y] = c, \tag{2.13}$$

$$E[h(Y)|Y = y] = h(y), \tag{2.14}$$

$$E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y) = E\{E[g(X)|Y]\}. \tag{2.15}$$

Exercises

2.1.1 I roll a six-sided die and observe the number N on the uppermost face. I then toss a fair coin N times and observe X, the total number of heads to appear. What is the probability that $N = 3$ and $X = 2$? What is the probability that $X = 5$? What is $E[X]$, the expected number of heads to appear?

2.1.2 Four nickels and six dimes are tossed, and the total number N of heads is observed. If $N = 4$, what is the conditional probability that exactly two of the nickels were heads?

2.1.3 A poker hand of five cards is dealt from a normal deck of 52 cards. Let X be the number of aces in the hand. Determine $\Pr\{X > 1|X \geq 1\}$. This is the probability that the hand contains more than one ace, given that it has at least one ace. Compare this with the probability that the hand contains more than one ace, given that it contains the ace of spades.

2.1.4 A six-sided die is rolled, and the number N on the uppermost face is recorded. From a jar containing 10 tags numbered $1, 2, \ldots, 10$, we then select N tags at random without replacement. Let X be the smallest number on the drawn tags. Determine $\Pr\{X = 2\}$.

2.1.5 Let X be a Poisson random variable with parameter λ. Find the conditional mean of X given that X is odd.

2.1.6 Suppose U and V are independent and follow the geometric distribution

$$p(k) = \rho(1-\rho)^k \quad \text{for } k = 0, 1, \ldots.$$

Define the random variable $Z = U + V$.

(a) Determine the joint probability mass function $p_{U,Z}(u,z) = \Pr\{U = u, Z = z\}$.

(b) Determine the conditional probability mass function for U given that $Z = n$.

Problems

2.1.1 Let M have a binomial distribution with parameters N and p. Conditioned on M, the random variable X has a binomial distribution with parameters M and π.

(a) Determine the marginal distribution for X.

(b) Determine the covariance between X and $Y = M - X$.

2.1.2 A card is picked at random from N cards labeled $1, 2, \ldots, N$, and the number that appears is X. A second card is picked at random from cards numbered $1, 2, \ldots, X$ and its number is Y. Determine the conditional distribution of X given $Y = y$, for $y = 1, 2, \ldots$.

2.1.3 Let X and Y denote the respective outcomes when two fair dice are thrown. Let $U = \min\{X, Y\}$, $V = \max\{X, Y\}$, and $S = U + V$, $T = V - U$.

(a) Determine the conditional probability mass function for U given $V = v$.

(b) Determine the joint mass function for S and T.

2.1.4 Suppose that X has a binomial distribution with parameters $p = \frac{1}{2}$ and N, where N is also random and follows a binomial distribution with parameters $q = \frac{1}{4}$ and $M = 20$. What is the mean of X?

2.1.5 A nickel is tossed 20 times in succession. Every time that the nickel comes up heads, a dime is tossed. Let X count the number of heads appearing on tosses of the dime. Determine $\Pr\{X = 0\}$.

2.1.6 A dime is tossed repeatedly until a head appears. Let N be the trial number on which this first head occurs. Then, a nickel is tossed N times. Let X count the number of times that the nickel comes up tails. Determine $\Pr\{X=0\}$, $\Pr\{X=1\}$, and $E[X]$.

2.1.7 The probability that an airplane accident that is due to structural failure is correctly diagnosed is 0.85, and the probability that an airplane accident that is not due to structural failure is incorrectly diagnosed as being due to structural failure is 0.35. If 30% of all airplane accidents are due to structural failure, then find the probability that an airplane accident is due to structural failure given that it has been diagnosed as due to structural failure.

2.1.8 Initially an urn contains one red and one green ball. A ball is drawn at random from the urn, observed, and then replaced. If this ball is red, then an additional red ball is placed in the urn. If the ball is green, then a green ball is added. A second ball is drawn. Find the conditional probability that the first ball was red given that the second ball drawn was red.

2.1.9 Let N have a Poisson distribution with parameter $\lambda = 1$. Conditioned on $N = n$, let X have a uniform distribution over the integers $0, 1, \ldots, n+1$. What is the marginal distribution for X?

2.1.10 *Do men have more sisters than women have*? In a certain society, all married couples use the following strategy to determine the number of children that they will have: If the first child is a girl, they have no more children. If the first child is a boy, they have a second child. If the second child is a girl, they have no more children. If the second child is a boy, they have exactly one additional child. (We ignore twins, assume sexes are equally likely, and the sex of distinct children are independent random variables, etc.) (a) What is the probability distribution for the number of children in a family? (b) What is the probability distribution for the number of girl children in a family? (c) A male child is chosen at random from all of the male children in the population. What is the probability distribution for the number of sisters of this child? What is the probability distribution for the number of his brothers?

2.2 The Dice Game Craps

An analysis of the dice game known as craps provides an educational example of the use of conditional probability in stochastic modeling. In craps, two dice are rolled and the sum of their uppermost faces is observed. If the sum has value 2, 3, or 12, the player loses immediately. If the sum is 7 or 11, the player wins. If the sum is 4, 5, 6, 8, 9, or 10, then further rolls are required to resolve the game. In the case where the sum is 4, e.g., the dice are rolled repeatedly until either a sum of 4 reappears or a sum of 7 is observed. If the sum of 4 appears first, the roller wins; if the sum of 7 appears first, he or she loses.

Consider repeated rolls of the pair of dice and let Z_n for $n = 0, 1, \ldots$ be the sum observed on the nth roll. Then, $Z_0, Z_1, \ldots$ are independent identically distributed random variables. If the dice are fair, the probability mass function is

$$\begin{aligned}
p_Z(2) &= \frac{1}{36}, & p_Z(8) &= \frac{5}{36},\\
p_Z(3) &= \frac{2}{36}, & p_Z(9) &= \frac{4}{36},\\
p_Z(4) &= \frac{3}{36}, & p_Z(10) &= \frac{3}{36},\\
p_Z(5) &= \frac{4}{36}, & p_Z(11) &= \frac{2}{36},\\
p_Z(6) &= \frac{5}{36}, & p_Z(12) &= \frac{1}{36}.\\
p_Z(7) &= \frac{6}{36}, & &
\end{aligned} \tag{2.16}$$

Let A denote the event that the player wins the game. By the law of total probability,

$$\Pr\{A\} = \sum_{k=2}^{12} \Pr\{A|Z_0 = k\} p_Z(k). \tag{2.17}$$

Because $Z_0 = 2, 3,$ or 12 calls for an immediate loss, then $\Pr\{A|Z_0 = k\} = 0$ for $k = 2, 3,$ or 12. Similarly, $Z_0 = 7$ or 11 results in an immediate win, and thus $\Pr\{A|Z_0 = 7\} = \Pr\{A|Z_0 = 11\} = 1$. It remains to consider the values $Z_0 = 4, 5, 6, 8, 9,$ and 10, which call for additional rolls. Since the logic remains the same in each of these cases, we will argue only the case in which $Z_0 = 4$. Abbreviate with $\alpha = \Pr\{A|Z_0 = 4\}$. Then, α is the probability that in successive rolls $Z_1, Z_2, \ldots$ of a pair of dice, a sum of 4 appears before a sum of 7. Denote this event by B, and again bring in the law of total probability. Then,

$$\alpha = \Pr\{B\} = \sum_{k=2}^{12} \Pr\{B|Z_1 = k\} p_Z(k). \tag{2.18}$$

Now $\Pr\{B|Z_1 = 4\} = 1$, while $\Pr\{B|Z_1 = 7\} = 0$. If the first roll results in anything other than a 4 or a 7, the problem is repeated in a statistically identical setting. That is, $\Pr\{B|Z_1 = k\} = \alpha$ for $k \neq 4$ or 7. Substitution into (2.18) results in

$$\begin{aligned}
\alpha &= p_Z(4) \times 1 + p_Z(7) \times 0 + \sum_{k \neq 4,7} p_Z(k) \times \alpha\\
&= p_Z(4) + [1 - p_Z(4) - p_Z(7)]\alpha,
\end{aligned}$$

or

$$\alpha = \frac{p_Z(4)}{p_Z(4) + p_Z(7)}. \tag{2.19}$$

The same result may be secured by means of a longer, more computational, method. One may partition the event B into disjoint elemental events by writing

$$\begin{aligned} B = &\{Z_1 = 4\} \cup \{Z_1 \neq 4 \text{ or } 7, Z_2 = 4\} \\ &\cup \{Z_1 \neq 4 \text{ or } 7, Z_2 \neq 4 \text{ or } 7, Z_3 = 4\} \cup \cdots, \end{aligned}$$

and then

$$\begin{aligned} \Pr\{B\} = &\Pr\{Z_1 = 4\} + \Pr\{Z_1 \neq 4 \text{ or } 7, Z_2 = 4\} \\ &+ \Pr\{Z_1 \neq 4 \text{ or } 7, Z_2 \neq 4 \text{ or } 7, Z_3 = 4\} + \cdots. \end{aligned}$$

Now use the independence of $Z_1, Z_2, \ldots$ and sum a geometric series to secure

$$\begin{aligned} \Pr\{B\} &= p_Z(4) + [1 - p_Z(4) - p_Z(7)]p_Z(4) \\ &\quad + [1 - p_Z(4) - p_Z(7)]^2 p_Z(4) + \cdots \\ &= \frac{p_Z(4)}{p_Z(4) + p_Z(7)} \end{aligned}$$

in agreement with (2.19).

Extending the result just obtained to the other cases having more than one roll, we have

$$\Pr\{A|Z_0 = k\} = \frac{p_Z(k)}{p_Z(k) + p_Z(7)} \quad \text{for } k = 4, 5, 6, 8, 9, 10.$$

Finally, substitution into (2.17) yields the total win probability

$$\Pr\{A\} = p_Z(7) + p_Z(11) + \sum_{k=4,5,6,8,9,10} \frac{p_Z(k)^2}{p_Z(k) + p_Z(7)}. \tag{2.20}$$

The numerical values for $p_Z(k)$ given in (2.16), together with (2.20), determine the win probability

$$\Pr\{A\} = 0.49292929\cdots.$$

Having explained the computations, let us go on to a more interesting question. Suppose that the dice are not perfect cubes but are shaved so as to be slightly thinner in one dimension than in the other two. The numbers that appear on opposite faces on a single die always sum to 7. That is, 1 is opposite 6, 2 is opposite 5, and 3 is opposite 4. Suppose it is the 3-4 dimension that is smaller than the other two. See Figure 2.1. This

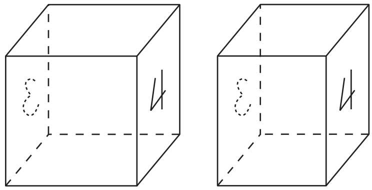

Figure 2.1 A cubic die versus a die that has been shaved down in one dimension.

will cause 3 and 4 to appear more frequently than the other faces, 1, 2, 5, and 6. To see this, think of the extreme case in which the 3-4 dimension is very thin, leading to a 3 or 4 on almost all tosses. Letting Y denote the result of tossing a single shaved die, we postulate that the probability mass function is given by

$$p_Y(3) = p_Y(4) = \frac{1}{6} + 2\varepsilon \equiv p_+,$$
$$p_Y(1) = p_Y(2) = p_Y(5) = p_Y(6) = \frac{1}{6} - \varepsilon \equiv p_-,$$

where $\varepsilon > 0$ is a small quantity depending on the amount by which the die has been biased.

If both dice are shaved in the same manner, the mass function for their sum can be determined in a straightforward manner from the following joint table:

	Die #1					
Die #2	**1** $p-$	**2** $p-$	**3** $p+$	**4** $p+$	**5** $p-$	**6** $p-$
1 $p-$	p_-^2	p_-^2	p_+p_-	p_+p_-	p_-^2	p_-^2
2 $p-$	p_-^2	p_-^2	p_+p_-	p_+p_-	p_-^2	p_-^2
3 $p+$	p_+p_-	p_+p_-	p_+^2	p_+^2	p_+p_-	p_+p_-
4 $p+$	p_+p_-	p_+p_-	p_+^2	p_+^2	p_+p_-	p_+p_-
5 $p-$	p_-^2	p_-^2	p_+p_-	p_+p_-	p_-^2	p_-^2
6 $p-$	p_-^2	p_-^2	p_+p_-	p_+p_-	p_-^2	p_-^2

It is easily seen that the probability mass function for the sum of the dice is

$$p(2) = p_-^2 = p(12),$$
$$p(3) = 2p_-^2 = p(11),$$
$$p(4) = p_-(p_- + 2p_+) = p(10),$$
$$p(5) = 4p_+p_- = p(9),$$
$$p(6) = p_-^2 + (p_+ + p_-)^2 = p(8),$$
$$p(7) = 4p_-^2 + 2p_+^2.$$

To obtain a numerical value to compare to the win probability $0.492929\cdots$ associated with fair dice, let us arbitrarily set $\varepsilon = 0.02$ so that $p_- = 0.146666\cdots$ and $p_+ = 0.206666\cdots$. Then, routine substitutions according to the table lead to

$$\begin{aligned} p(2) = p(12) &= 0.02151111, \quad p(5) = p(9) = 0.12124445, \\ p(3) = p(11) &= 0.04302222, \quad p(6) = p(8) = 0.14635556, \\ p(4) = p(10) &= 0.08213333, \quad p(7) = 0.17146667, \end{aligned} \tag{2.21}$$

and the win probability becomes $\Pr\{A\} = 0.5029237$.

The win probability of 0.4929293 with fair dice is unfavorable, i.e., is less than $\frac{1}{2}$. With shaved dice, the win probability is favorable, now being 0.5029237. What appears to be a slight change becomes, in fact, quite significant when a large number of games are played. See Chapter 3, Section 3.5.

Exercises

2.2.1 A red die is rolled a single time. A green die is rolled repeatedly. The game stops the first time that the sum of the two dice is either 4 or 7. What is the probability that the game stops with a sum of 4?

2.2.2 Verify the win probability of 0.5029237 by substituting from (2.21) into (2.20).

2.2.3 Determine the win probability when the dice are shaved on the 1–6 faces and $p_+ = 0.206666\cdots$ and $p_- = 0.146666\cdots$.

Problems

2.2.1 Let $X_1, X_2, \ldots$ be independent identically distributed positive random variables whose common distribution function is F. We interpret $X_1, X_2, \ldots$ as successive bids on an asset offered for sale. Suppose that the policy is followed of accepting the first bid that exceeds some prescribed number A. Formally, the accepted bid is X_N, where

$$N = \min\{k \geq 1; X_k > A\}.$$

Set $\alpha = \Pr\{X_1 > A\}$ and $M = E[X_N]$.

(a) Argue the equation

$$M = \int_A^\infty x\,\mathrm{d}F(x) + (1 - \alpha)M$$

by considering the possibilities, either the first bid is accepted or it is not.

(b) Solve for M, thereby obtaining

$$M = \alpha^{-1} \int_A^\infty x \, \mathrm{d}F(x).$$

(c) When X_1 has an exponential distribution with parameter λ, use the memoryless property to deduce $M = A + \lambda^{-1}$.

(d) Verify this result by calculation in (b).

2.2.2 Consider a pair of dice that are unbalanced by the addition of weights in the following manner: Die #1 has a small piece of lead placed near the four side, causing the appearance of the outcome 3 more often than usual, while die #2 is weighted near the three side, causing the outcome 4 to appear more often than usual. We assign the probabilities

Die #1

$$\begin{aligned} &p(1) = p(2) = p(5) = p(6) = 0.166667, \\ &p(3) = 0.186666, \\ &p(4) = 0.146666; \end{aligned}$$

Die #2

$$\begin{aligned} &p(1) = p(2) = p(5) = p(6) = 0.166667, \\ &p(3) = 0.146666, \\ &p(4) = 0.186666. \end{aligned}$$

Determine the win probability if the game of craps is played with these loaded dice.

2.3 Random Sums

Sums of the form $X = \xi_1 + \cdots + \xi_N$, where N is random, arise frequently and in varied contexts. Our study of random sums begins with a crisp definition and a precise statement of the assumptions effective in this section, followed by some quick examples.

We postulate a sequence $\xi_1, \xi_2, \ldots$ of independent and identically distributed random variables. Let N be a discrete random variable, independent of $\xi_1, \xi_2, \ldots$ and having the probability mass function $p_N(n) = \Pr\{N = n\}$ for $n = 0, 1, \ldots$. Define the random sum X by

$$X = \begin{cases} 0 & \text{if } N = 0, \\ \xi_1 + \cdots + \xi_N & \text{if } N > 0. \end{cases} \tag{2.22}$$

We save space by abbreviating (2.22) to simply $X = \xi_1 + \cdots + \xi_N$, understanding that $X = 0$ whenever $N = 0$.

Examples

(a) *Queueing* Let N be the number of customers arriving at a service facility in a specified period of time, and let ξ_i be the service time required by the ith customer. Then, $X = \xi_1 + \cdots + \xi_N$ is the total demand for service time.

(b) *Risk Theory* Suppose that a total of N claims arrives at an insurance company in a given week. Let ξ_i be the amount of the ith claim. Then, the total liability of the insurance company is $X = \xi_1 + \cdots + \xi_N$.

(c) *Population Models* Let N be the number of plants of a given species in a specified area, and let ξ_i be the number of seeds produced by the ith plant. Then, $X = \xi_1 + \cdots + \xi_N$ gives the total number of seeds produced in the area.

(d) *Biometrics* A wildlife sampling scheme traps a random number N of a given species. Let ξ_i be the weight of the ith specimen. Then, $X = \xi_1 + \cdots + \xi_N$ is the total weight captured.

The necessary background in conditional probability was covered in Section 2.1 for when $\xi_1, \xi_2, \ldots$ are discrete random variables. In order to study the random sum $X = \xi_1 + \cdots + \xi_N$ when $\xi_1, \xi_2, \ldots$ are continuous random variables, we need to extend our knowledge of conditional distributions.

2.3.1 Conditional Distributions: The Mixed Case

Let X and N be jointly distributed random variables and suppose that the possible values for N are the discrete set $n = 0, 1, 2, \ldots$. Then, the elementary definition of conditional probability (2.1) applies to define the *conditional distribution function* $F_{X|N}(x|n)$ of the random variable X, given that $N = n$, to be

$$F_{X|N}(x|n) = \frac{\Pr\{X \leq x \text{ and } N = n\}}{\Pr\{N = n\}} \quad \text{if } \Pr\{N = n\} > 0, \tag{2.23}$$

and the conditional distribution function is not defined at values of n for which $\Pr\{N = n\} = 0$. It is elementary to verify that $F_{X|N}(x|n)$ is a probability distribution function in x at each value of n for which it is defined.

The case in which X is a discrete random variable was covered in Section 2.1. Now let us suppose that X is continuous and that $F_{X|N}(x|n)$ is differentiable in x at each value of n for which $\Pr\{N = n\} > 0$. We define the *conditional probability density function* $f_{X|N}(x|n)$ for the random variable X given that $N = n$ by setting

$$f_{X|N}(x|n) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|N}(x|n) \quad \text{if } \Pr\{N = n\} > 0. \tag{2.24}$$

Again, $f_{X|N}(x|n)$ is a probability density function in x at each value of n for which it is defined. Moreover, the conditional density as defined in (2.24) has the appropriate properties, e.g.,

$$\Pr\{a \le X < b, N = n\} = \int_a^b f_{X|N}(x|n) p_N(n) \mathrm{d}x \tag{2.25}$$

for $a < b$ and where $p_N(n) = \Pr\{N = n\}$. The law of total probability leads to the marginal probability density function for X via

$$f_X(x) = \sum_{n=0}^{\infty} f_{X|N}(x|n) p_N(n). \tag{2.26}$$

Suppose that g is a function for which $E[|g(X)|] < \infty$. The conditional expectation of $g(X)$ given that $N = n$ is defined by

$$E[g(X)|N = n] = \int g(x) f_{X|N}(x|n)\, \mathrm{d}x. \tag{2.27}$$

Stipulated thus, $E[g(X)|N = n]$ satisfies the properties listed in (2.7) to (2.15) for the joint discrete case. For example, the law of total probability is

$$E[g(X)] = \sum_{n=0}^{\infty} E[g(X)|N = n] p_N(n) = E\{E[g(X)|N]\}. \tag{2.28}$$

2.3.2 The Moments of a Random Sum

Let us assume that ξ_k and N have the finite moments

$$\begin{aligned} E[\xi_k] &= \mu, \quad \mathrm{Var}[\xi_k] = \sigma^2, \\ E[N] &= \nu, \quad \mathrm{Var}[N] = \tau^2, \end{aligned} \tag{2.29}$$

and determine the mean and variance for $X = \xi_1 + \cdots + \xi_N$ as defined in (2.22). The derivation provides practice in manipulating conditional expectations, and the results,

$$E[X] = \mu\nu, \quad \mathrm{Var}[X] = \nu\sigma^2 + \mu^2\tau^2, \tag{2.30}$$

are useful and important. The properties of conditional expectation listed in (2.7) to (2.15) justify the steps in the determination.

If we begin with the mean $E[X]$, then

$$\begin{aligned} E[X] &= \sum_{n=0}^{\infty} E[X|N=n]p_N(n) && \text{[by (2.15)]} \\ &= \sum_{n=1}^{\infty} E[\xi_1+\cdots+\xi_N|N=n]p_N(n) && \text{(definition of X)} \\ &= \sum_{n=1}^{\infty} E[\xi_1+\cdots+\xi_n|N=n]p_N(n) && \text{[by (2.9)]} \\ &= \sum_{n=1}^{\infty} E[\xi_1+\cdots+\xi_n]p_N(n) && \text{[by (2.10)]} \\ &= \mu\sum_{n=1}^{\infty} np_N(n) = \mu\nu. \end{aligned}$$

To determine the variance, we begin with the elementary step

$$\begin{aligned} \operatorname{Var}[X] &= E\left[(X-\mu\nu)^2\right] = E\left[(X-N\mu+N\mu-\nu\mu)^2\right] \\ &= E\left[(X-N\mu)^2\right] + E\left[\mu^2(N-\nu)^2\right] \\ &\quad + 2E[\mu(X-N\mu)(N-\nu)]. \end{aligned} \tag{2.31}$$

Then,

$$\begin{aligned} E\left[(X-N\mu)^2\right] &= \sum_{n=0}^{\infty} E\left[(X-N\mu)^2|N=n\right]p_N(n) \\ &= \sum_{n=1}^{\infty} E\left[(\xi_1+\cdots+\xi_n-n\mu)^2|N=n\right]p_N(n) \\ &= \sigma^2+\sum_{n=1}^{\infty} np_N(n) = \nu\sigma^2, \end{aligned}$$

and

$$E\left[\mu^2(N-\nu)^2\right] = \mu^2E\left[(N-\nu)^2\right] = \mu^2\tau^2,$$

while

$$\begin{aligned} E[\mu(X-N\mu)(N-\nu)] &= \mu\sum_{n=0}^{\infty} E[(X-n\mu)(n-\nu)|N=n]p_N(n) \\ &= \mu\sum_{n=0}^{\infty}(n-\nu)E[(X-n\mu)|N=n]p_N(n) \\ &= 0 \end{aligned}$$

(because $E[(X-n\mu)|N=n]=E[\xi_1+\cdots+\xi_n-n\mu]=0$). Then, (2.31) with the subsequent three calculations validates the variance of X as stated in (2.30).

Example The number of offspring of a given species is a random variable having probability mass function $p(k)$ for $k=0,1,\ldots$. A population begins with a single parent who produces a random number N of progeny, each of which independently produces offspring according to $p(k)$ to form a second generation. Then, the total number of descendants in the second generation may be written $X=\xi_1+\cdots+\xi_N$, where ξ_k is the number of progeny of the kth offspring of the original parent. Let $E[N]=E[\xi_k]=\mu$ and $\text{Var}[N]=\text{Var}[\xi_k]=\sigma^2$. Then,

$$E[X]=\mu^2 \quad \text{and} \quad \text{Var}[X]=\mu\sigma^2(1+\mu).$$

2.3.3 The Distribution of a Random Sum

Suppose that the summands $\xi_1,\xi_2,\ldots$ are continuous random variables having a probability density function $f(z)$. For $n\geq 1$, the probability density function for the fixed sum $\xi_1+\cdots+\xi_n$ is the n-fold convolution of the density $f(z)$, denoted by $f^{(n)}(z)$ and recursively defined by

$$f^{(1)}(z)=f(z)$$

and

$$f^{(n)}(z)=\int f^{(n-1)}(z-u)f(u)\mathrm{d}u \quad \text{for } n>1. \tag{2.32}$$

(See Chapter 1, Section 1.2.5 for a discussion of convolutions.) Because N and $\xi_1,\xi_2,\ldots$ are independent, then $f^{(n)}(z)$ is also the conditional density function for $X=\xi_1+\cdots+\xi_N$ given that $N=n\geq 1$. Let us suppose that $\Pr\{N=0\}=0$. Then, by the law of total probability as expressed in (2.26), X is continuous and has the marginal density function

$$f_X(x)=\sum_{n=1}^{\infty}f^{(n)}(x)p_N(n). \tag{2.33}$$

Remark When $N=0$ can occur with positive probability, then $X=\xi_1+\cdots+\xi_N$ is a random variable having both continuous and discrete components to its distribution. Assuming that $\xi_1,\xi_2,\ldots$ are continuous with probability density function $f(z)$, then

$$\Pr\{X=0\}=\Pr\{N=0\}=p_N(0),$$

while for $0<a<b$ or $a<b<0$, then

$$\Pr\{a<X<b\}=\int_a^b\left\{\sum_{n=1}^{\infty}f^{(n)}(z)p_N(n)\right\}\mathrm{d}z. \tag{2.34}$$

Example *A Geometric Sum of Exponential Random Variables* In the following computational example, suppose

$$f(z) = \begin{cases} \lambda e^{-\lambda z} & \text{for } z \geq 0, \\ 0 & \text{for } z < 0, \end{cases}$$

and

$$p_N(n) = \beta(1-\beta)^{n-1} \quad \text{for } n = 1, 2, \ldots.$$

For $n \geq 1$, the n-fold convolution of $f(z)$ is the gamma density

$$f^{(n)}(z) = \begin{cases} \dfrac{\lambda^n}{(n-1)!} z^{n-1} e^{-\lambda z} & \text{for } z \geq 0, \\ 0 & \text{for } z < 0. \end{cases}$$

(See Chapter 1, Section 1.4.4 for discussion.)

The density for $X = \xi_1 + \cdots + \xi_N$ is given, according to (2.26), by

$$\begin{aligned} f_X(z) &= \sum_{n=1}^{\infty} f^{(n)}(z) p_N(n) \\ &= \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} z^{n-1} e^{-\lambda z} \beta(1-\beta)^{n-1} \\ &= \lambda \beta e^{-\lambda z} \sum_{n=1}^{\infty} \frac{[\lambda(1-\beta)z]^{n-1}}{(n-1)!} \\ &= \lambda \beta e^{-\lambda z} e^{\lambda(1-\beta)z} \\ &= \lambda \beta e^{-\lambda \beta z}, \quad z \geq 0. \end{aligned}$$

Surprise! X has an exponential distribution with parameter $\lambda\beta$.

Example *Stock Price Changes* Stochastic models for price fluctuations of publicly traded assets were developed as early as 1900.

Let Z denote the difference in price of a single share of a certain stock between the close of one trading day and the close of the next day. For an actively traded stock, a large number of transactions take place in a single day, and the total daily price change is the sum of the changes over these individual transactions. If we assume that price changes over successive transactions are independent random variables having a common finite variance,* then the central limit theorem applies. The price change over a large number of transactions should follow a normal, or Gaussian, distribution.

* Rather strong economic arguments in support of these assumptions can be given. The independence follows from concepts of a "perfect market" and the common variance from notions of time stationarity.

A variety of empirical studies have supported this conclusion. For the most part, these studies involved price changes over a fixed number of transactions. Other studies found discrepancies in that both very small and very large price changes occurred more frequently in the data than suggested by normal theory. At the same time, intermediate-size price changes were under represented in the data. For the most part, these studies examined price changes over fixed durations containing a random number of transactions.

A natural question arises: Does the random number of transactions in a given day provide a possible explanation for the departures from normality that are observed in data of daily price changes? Let us model the daily price change in the form

$$Z = \xi_0 + \xi_1 + \cdots + \xi_N = \xi_0 + X, \tag{2.35}$$

where $\xi_0, \xi_1, \ldots$ are independent normally distributed random variables with common mean zero and variance σ^2, and N has a Poisson distribution with mean ν.

We interpret N as the number of transactions during the day, ξ_i for $i \geq 1$ as the price change during the ith transaction, and ξ_0 as an initial price change arising between the close of the market on one day and the opening of the market on the next day. (An obvious generalization would allow the distribution of ξ_0 to differ from that of $\xi_1, \xi_2, \ldots$.)

Conditioned on $N = n$, the random variable $Z = \xi_0 + \xi_1 + \cdots + \xi_n$ is normally distributed with mean zero and variance $(n+1)\sigma^2$. The conditional density function is

$$\phi_n(z) = \frac{1}{\sqrt{2\pi(n+1)}\sigma} \exp\left\{-\frac{1}{2}\frac{1z^2}{(n+1)\sigma^2}\right\}.$$

Since the probability mass function for N is

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

using (2.33) we determine the probability density function for the daily price change to be

$$f_Z(z) = \sum_{n=0}^{\infty} \phi_n(z) \frac{\lambda^n e^{-\lambda}}{n!}.$$

The formula for the density $f_Z(z)$ does not simplify. Nevertheless, numerical calculations are possible. When $\lambda = 1$ and $\sigma^2 = \frac{1}{2}$, then (2.30) shows that the variance of the daily price change Z in the model (2.35) is $\mathrm{Var}[Z] = (1+\lambda)\sigma^2 = 1$. Thus, comparing the density $f_Z(z)$ when $\lambda = 1$ and $\sigma^2 = \frac{1}{2}$ to a normal density with mean zero and variance, one sheds some light on the question at hand.

The calculations were carried out and are shown in Figure 2.2.

The departure from normality that is exhibited by the random sum in Figure 2.2 is consistent with the departure from normality shown by stock price changes over fixed time intervals. Of course, our calculations do not *prove* that the observed departure

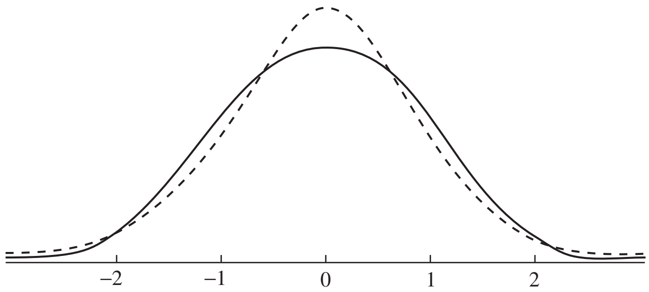

Figure 2.2 A standard normal density (solid line) as compared with a density for a random sum (dashed line). Both densities have zero mean and unit variance.

from normality is *caused* by the random number of transactions in a fixed time interval. Rather, the calculations show only that such an explanation is consistent with the data and is, therefore, a possible cause.

Exercises

2.3.1 A six-sided die is rolled, and the number N on the uppermost face is recorded. Then a fair coin is tossed N times, and the total number Z of heads to appear is observed. Determine the mean and variance of Z by viewing Z as a random sum of N Bernoulli random variables. Determine the probability mass function of Z, and use it to find the mean and variance of Z.

2.3.2 Six nickels are tossed, and the total number N of heads is observed. Then N dimes are tossed, and the total number Z of tails among the dimes is observed. Determine the mean and variance of Z. What is the probability that $Z = 2$?

2.3.3 Suppose that upon striking a plate a single electron is transformed into a number N of electrons, where N is a random variable with mean μ and standard deviation σ. Suppose that each of these electrons strikes a second plate and releases further electrons, independently of each other and each with the same probability distribution as N. Let Z be the total number of electrons emitted from the second plate. Determine the mean and variance of Z.

2.3.4 A six-sided die is rolled, and the number N on the uppermost face is recorded. From a jar containing 10 tags numbered $1, 2, \ldots, 10$ we then select N tags at random without replacement. Let X be the smallest number on the drawn tags. Determine $\Pr\{X = 2\}$ and $E[X]$.

2.3.5 The number of accidents occurring in a factory in a week is a Poisson random variable with mean 2. The number of individuals injured in different accidents is independently distributed, each with mean 3 and variance 4. Determine the mean and variance of the number of individuals injured in a week.

Problems

2.3.1 The following experiment is performed: An observation is made of a Poisson random variable N with parameter λ. Then N independent Bernoulli trials are performed, each with probability p of success. Let Z be the total number of successes observed in the N trials.

(a) Formulate Z as a random sum and thereby determine its mean and variance.

(b) What is the distribution of Z?

2.3.2 For each given p, let Z have a binomial distribution with parameters p and N. Suppose that N is itself binomially distributed with parameters q and M. Formulate Z as a random sum and show that Z has a binomial distribution with parameters pq and M.

2.3.3 Suppose that $\xi_1, \xi_2, \ldots$ are independent and identically distributed with $\Pr\{\xi_k = \pm 1\} = \frac{1}{2}$. Let N be independent of $\xi_1, \xi_2, \ldots$ and follow the geometric probability mass function

$$p_N(k) = \alpha(1-\alpha)^k \quad \text{for } k = 0, 1, \ldots,$$

where $0 < \alpha < 1$. Form the random sum $Z = \xi_1 + \cdots + \xi_N$.

(a) Determine the mean and variance of Z.

(b) Evaluate the higher moments $m_3 = E\left[Z^3\right]$ and $m_4 = E\left[Z^4\right]$.

Hint: Express Z^4 in terms of the ξ_i's where $\xi_i^2 = 1$ and $E[\xi_i \xi_j] = 0$.

2.3.4 Suppose $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables having mean μ and variance σ^2. Form the random sum $S_N = \xi_1 + \cdots + \xi_N$.

(a) Derive the mean and variance of S_N when N has a Poisson distribution with parameter λ.

(b) Determine the mean and variance of S_N when N has a geometric distribution with mean $\lambda = (1-p)/p$.

(c) Compare the behaviors in (a) and (b) as $\lambda \to \infty$.

2.3.5 To form a slightly different random sum, let $\xi_0, \xi_1, \ldots$ be independent identically distributed random variables and let N be a nonnegative integer-valued random variable, independent of $\xi_0, \xi_1, \ldots$. The first two moments are

$$E[\xi_k] = \mu, \quad \mathrm{Var}[\xi_k] = \sigma^2,$$
$$E[N] = \nu, \quad \mathrm{Var}[N] = \tau^2.$$

Determine the mean and variance of the random sum $Z = \xi_0 + \cdots + \xi_N$.

2.4 Conditioning on a Continuous Random Variable*

Let X and Y be jointly distributed continuous random variables with joint probability density function $f_{X,Y}(x,y)$. We define the conditional probability density function $f_{X|Y}(x|y)$ for the random variable X given that $Y = y$ by the formula

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0, \tag{2.36}$$

* The reader may wish to defer reading this section until encountering Chapter 7, on renewal processes, where conditioning on a continuous random variable first appears.

and the conditional density is not defined at values y for which $f_Y(y) = 0$. The conditional distribution function for X given $Y = y$ is defined by

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(\xi|y)\mathrm{d}\xi \quad \text{if } f_Y(y) > 0. \tag{2.37}$$

Finally, given a function g for which $E[|g(X)|] < \infty$, the conditional expectation of $g(X)$ given that $Y = y$ is defined to be

$$E[g(X)|Y = y] = \int g(x)f_{X|Y}(x|y)\mathrm{d}x \quad \text{if } f_Y(y) > 0. \tag{2.38}$$

The definitions given in (2.36) to (2.38) are a significant extension of our elementary notions of conditional probability because they allow us to condition on certain events having zero probability. To understand the distinction, try to apply the elementary formula

$$\Pr\{A|B\} = \frac{\Pr\{A \text{ and } B\}}{\Pr\{B\}} \quad \text{if } \Pr\{B\} > 0 \tag{2.39}$$

to evaluate the conditional probability $\Pr\{a < X \le b|Y = y\}$. We set $A = \{a < X \le b\}$ and $B = \{Y = y\}$. But Y is a continuous random variable, and thus, $\Pr\{B\} = \Pr\{Y = y\} = 0$, and (2.39) cannot be applied. Equation (2.37) saves the day, yielding

$$\Pr\{a < X \le b|Y = y\} = F_{X|Y}(b|y) - F_{X|Y}(a|y) = \int_{a}^{b} f_{X|Y}(\xi|y)\mathrm{d}\xi, \tag{2.40}$$

provided only that the density $f_Y(y)$ is strictly positive at the point y.

To emphasize the important advance being made, we consider the following simple problem. A woman arrives at a bus stop at a time Y that is uniformly distributed between 0 (noon) and 1. Independently, the bus arrives at a time Z that is also uniformly distributed between 0 and 1. Given that the woman arrives at time $Y = 0.20$, what is the probability that she misses the bus?

On the one hand, the answer $\Pr\{Z < Y|Y = 0.20\} = 0.20$ is obvious. On the other hand, this elementary question cannot be answered by the elementary conditional probability formula (2.39) because the event $\{Y = 0.20\}$ has zero probability. To apply (2.36), start with the joint density function

$$f_{Z,Y}(z,y) = \begin{cases} 1 & \text{for } 0 \le z, y \le 1, \\ 0 & \text{elsewhere,} \end{cases}$$

and change variables according to $X = Y - Z$. Then,

$$f_{X,Y}(x,y) = 1 \quad \text{for } 0 \le y \le 1, y - 1 \le x \le y,$$

and, applying (2.36), we find that

$$f_{X|Y}(x|0.20) = \frac{f_{X,Y}(x, 0.20)}{f_Y(0.20)} = 1 \quad \text{for } -0.80 \le x \le 0.20.$$

Finally,

$$\Pr\{Z < Y|Y = 0.20\} = \Pr\{X > 0|Y = 0.20\} = \int_0^\infty f_{X|Y}(x|0.20)\mathrm{d}x = 0.20.$$

We see that the definition in (2.36) leads to the intuitively correct answer.

The conditional density function that is prescribed by (2.36) possesses all of the properties that are called for by our intuition and the basic concept of conditional probability. In particular, one can calculate the probability of joint events by the formula

$$\Pr\{a < X < b, c < Y < d\} = \int_c^d \left\{ \int_a^b f_{X|Y}(x|y)\mathrm{d}x \right\} f_Y(y)\mathrm{d}y, \tag{2.41}$$

which becomes the law of total probability by setting $c = -\infty$ and $d = +\infty$;

$$\Pr\{a < X < b\} = \int_{-\infty}^{+\infty} \left\{ \int_a^b f_{X|Y}(x|y)\mathrm{d}x \right\} f_Y(y)\mathrm{d}y. \tag{2.42}$$

For the same reasons, the conditional expectation as defined in (2.38) satisfies the requirements listed in (2.7) to (2.11). The property (2.12), adapted to a continuous random variable Y, is written

$$\begin{aligned} E[g(X)h(Y)] &= E\{h(Y)E[g(X)|Y]\} \\ &= \int h(y)E[g(X)|Y = y]f_Y(y)\mathrm{d}y, \end{aligned} \tag{2.43}$$

valid for any bounded function h, and assuming $E[|g(X)|] < \infty$. When $h(y) \equiv 1$, we recover the law of total probability in the form

$$E[g(X)] = E\{E[g(X)|Y]\} = \int E[g(X)|Y = y]f_Y(y)\mathrm{d}y. \tag{2.44}$$

Both the discrete and continuous cases of (2.43) and (2.44) are contained in the expressions

$$E[g(X)h(Y)] = E\{h(Y)E[g(X)|Y]\} = \int h(y)E[g(X)|Y = y]\,\mathrm{d}F_Y(y), \tag{2.45}$$

and

$$E[g(X)] = E\{E[g(X)|Y]\} = \int E[g(X)|Y=y]\,\mathrm{d}F_Y(y). \tag{2.46}$$

[See the discussion following (1.9) in Chapter 1 for an explanation of the symbolism in (2.45) and (2.46).]

The following exercises provide practice in deriving conditional probability density functions and in manipulating the law of total probability.

Example Suppose X and Y are jointly distributed random variables having the density function

$$f_{XY}(x,y) = \frac{1}{y}\mathrm{e}^{-(x/y)-y} \quad \text{for } x, y > 0.$$

We first determine the marginal density for y, obtaining

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_{XY}(x,y)\,\mathrm{d}x \\ &= \mathrm{e}^{-y}\int_0^\infty y^{-1}\mathrm{e}^{-(x/y)}\,\mathrm{d}x = \mathrm{e}^{-y} \quad \text{for } y > 0. \end{aligned}$$

Then,

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = y^{-1}\mathrm{e}^{-(x/y)} \quad \text{for } x, y > 0.$$

That is, conditional on $Y = y$, the random variable X has an exponential distribution with parameter $1/y$. It is easily seen that $E[X|Y=y] = y$.

Example For each given p, let X have a binomial distribution with parameters p and N. Suppose that p is uniformly distributed on the interval [0, 1]. What is the resulting distribution of X?

We are given the marginal distribution for p and the conditional distribution for X. Applying the law of total probability and the beta integral in Chapter 1 (1.66), we obtain

$$\begin{aligned} \Pr\{X=k\} &= \int_0^1 \Pr\{X=k|p=\xi\}f_p(\xi)\,\mathrm{d}\xi \\ &= \int_0^1 \frac{N!}{k!\,(N-k)!}\xi^k(1-\xi)^{N-k}\,\mathrm{d}\xi \\ &= \frac{N!}{k!\,(N-k)!}\,\frac{k!\,(N-k)!}{(N+1)!} \\ &= \frac{1}{N+1} \quad \text{for } k = 0, 1, \ldots, N. \end{aligned}$$

That is, X is uniformly distributed on the integers $0, 1, \dots, N$.

When p has the beta distribution with parameters r and s, then similar calculations give

$$\Pr\{X=k\} = \frac{N!}{k!\,(N-k)!}\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}\int_0^1 \xi^{r-1}(1-\xi)^{s-1}\xi^k(1-\xi)^{N-k}\mathrm{d}\xi$$

$$= \binom{N}{k}\frac{\Gamma(r+s)\Gamma(r+k)\Gamma(s+N-k)}{\Gamma(r)\Gamma(s)\Gamma(N+r+s)} \quad \text{for } k=0,1,\dots,N.$$

Example A random variable Y follows the exponential distribution with parameter θ. Given that $Y=y$, the random variable X has a Poisson distribution with mean y. Applying the law of total probability then yields

$$\begin{aligned}\Pr\{X=k\} &= \int_0^\infty \frac{y^k\mathrm{e}^{-y}}{k!}\theta\mathrm{e}^{-\theta y}\mathrm{d}y\\ &= \frac{\theta}{k!}\int_0^\infty y^k\mathrm{e}^{-(1+\theta)y}\mathrm{d}y\\ &= \frac{\theta}{k!\,(1+\theta)^{k+1}}\int_0^\infty u^k\mathrm{e}^{-u}\mathrm{d}u\\ &= \frac{\theta}{(1+\theta)^{k+1}} \quad \text{for } k=0,1,\dots.\end{aligned}$$

Suppose that Y has the gamma density

$$f_Y(y) = \frac{\theta}{\Gamma(\alpha)}(\theta y)^{\alpha-1}\mathrm{e}^{-\theta y}, \quad y \ge 0.$$

Then, similar calculations yield

$$\begin{aligned}\Pr\{X=k\} &= \int_0^\infty \frac{y^k\mathrm{e}^{-y}}{k!}\frac{\theta}{\Gamma(\alpha)}(\theta y)^{\alpha-1}\mathrm{e}^{-\theta y}\mathrm{d}y\\ &= \frac{\theta^\alpha}{k!\,\Gamma(\alpha)(1+\theta)^{k+\alpha}}\int_0^\infty u^{k+\alpha-1}\mathrm{e}^{-u}\mathrm{d}u\\ &= \frac{\Gamma(k+\alpha)}{k!\,\Gamma(\alpha)}\left(\frac{\theta}{1+\theta}\right)^\alpha\left(\frac{1}{1+\theta}\right)^k, \quad k=0,1,\dots.\end{aligned}$$

This is the negative binomial distribution.

Exercises

2.4.1 Suppose that three contestants on a quiz show are each given the same question and that each answers it correctly, independently of the others, with probability p. But the difficulty of the question is itself a random variable, so let us suppose, for the sake of illustration, that p is uniformly distributed over the interval $(0, 1]$. What is the probability that exactly two of the contestants answer the question correctly?

2.4.2 Suppose that three components in a certain system each function with probability p and fail with probability $1-p$, each component operating or failing independently of the others. But the system is in a random environment so that p is itself a random variable. Suppose that p is uniformly distributed over the interval $(0, 1]$. The system operates if at least two of the components operate. What is the probability that the system operates?

2.4.3 A random variable T is selected that is uniformly distributed over the interval $(0, 1]$. Then, a second random variable U is chosen, uniformly distributed on the interval $(0, T]$. What is the probability that U exceeds $\frac{1}{2}$?

2.4.4 Suppose X and Y are independent random variables, each exponentially distributed with parameter λ. Determine the probability density function for $Z = X/Y$.

2.4.5 Let U be uniformly distributed over the interval $[0, L]$ where L follows the gamma density $f_L(x) = xe^{-x}$ for $x \geq 0$. What is the joint density function of U and $V = L - U$?

Problems

2.4.1 Suppose that the outcome X of a certain chance mechanism depends on a parameter p according to $\Pr\{X = 1\} = p$ and $\Pr\{X = 0\} = 1 - p$, where $0 \leq p \leq 1$. Suppose that p is chosen at random, uniformly distributed over the unit interval $[0, 1]$, and then, that two independent outcomes X_1 and X_2 are observed. What is the unconditional correlation coefficient between X_1 and X_2?

Note: Conditionally independent random variables may become dependent if they share a common parameter.

2.4.2 Let N have a Poisson distribution with parameter $\lambda > 0$. Suppose that, conditioned on $N = n$, the random variable X is binomially distributed with parameters $N = n$ and p. Set $Y = N - X$. Show that X and Y have Poisson distributions with respective parameters λp and $\lambda(1 - p)$ and that X and Y are independent.

Note: Conditionally dependent random variables may become independent through randomization.

2.4.3 Let X have a Poisson distribution with parameter $\lambda > 0$. Suppose λ itself is random, following an exponential density with parameter θ.

(a) What is the marginal distribution of X?

(b) Determine the conditional density for λ given $X = k$.

2.4.4 Suppose X and Y are independent random variables having the same Poisson distribution with parameter λ, but where λ is also random, being exponentially distributed with parameter θ. What is the conditional distribution for X given that $X+Y=n$?

2.4.5 Let X and Y be jointly distributed random variables whose joint probability mass function is given in the following table:

			x	
		-1	0	1
y	-1	$\frac{1}{9}$	$\frac{2}{9}$	0
	0	0	$\frac{1}{9}$	$\frac{2}{9}$
	1	$\frac{2}{9}$	0	$\frac{1}{9}$

$p(x,y) = \Pr\{X=x, Y=y\}$

Show that the covariance between X and Y is zero even though X and Y are not independent.

2.4.6 Let $X_0, X_1, X_2, \ldots$ be independent identically distributed nonnegative random variables having a continuous distribution. Let N be the first index k for which $X_k > X_0$. That is, $N = 1$ if $X_1 > X_0, N = 2$ if $X_1 \leq X_0$ and $X_2 > X_0$, etc. Determine the probability mass function for N and the mean $E[N]$. (Interpretation: $X_0, X_1, \ldots$ are successive offers or bids on a car that you are trying to sell. Then, N is the index of the first bid that is better than the initial bid.)

2.4.7 Suppose that X and Y are independent random variables, each having the same exponential distribution with parameter α. What is the conditional probability density function for X, given that $Z = X+Y = z$?

2.4.8 Let X and Y have the normal density given in Chapter 1, in (1.47). Show that the conditional density function for X, given that $Y = y$, is normal with moments

$$\mu_{X|Y} = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y)$$

and

$$\sigma_{X|Y} = \sigma_X\sqrt{1-\rho^2}.$$

2.5 Martingales*

Stochastic processes are characterized by the dependence relationships that exist among their variables. The martingale property is one such relationship that captures a notion of a game being fair. The martingale property is a restriction solely on the

* Some problems scattered throughout the text call for the student to identify certain stochastic processes as martingales. Otherwise, the material of this section is not used in the sequel.

conditional means of some of the variables, given values of others, and does not otherwise depend on the actual distribution of the random variables in the stochastic process. Despite the apparent weakness of the martingale assumption, the consequences are striking, as we hope to suggest.

2.5.1 The Definition

We begin the presentation with the simplest definition.

Definition A stochastic process $\{X_n; n = 0, 1, \ldots\}$ is a martingale if for $n = 0, 1, \ldots,$

(a) $E[|X_n|] < \infty$,
and
(b) $E[X_{n+1}|X_0, \ldots, X_n] = X_n$.

Taking expectations on both sides of (b),

$$E\{E[X_{n+1}|X_0, \ldots, X_n]\} = E\{X_n\},$$

and using the law of total probability in the form

$$E\{E[X_{n+1}|X_1, \ldots, X_n]\} = E[X_{n+1}]$$

shows that

$$E[X_{n+1}] = E[X_n],$$

and consequently, a martingale has constant mean:

$$E[X_0] = E[X_k] = E[X_n], \quad 0 \le k \le n. \tag{2.47}$$

A similar conditioning (see Problem 2.5.1) verifies that the martingale equality (b) extends to future times in the form

$$E[X_m|X_0, \ldots, X_n] = X_n \quad \text{for } m \ge n. \tag{2.48}$$

To relate the martingale property to concepts of fairness in gambling, consider X_n to be a certain player's fortune after the nth play of a game. The game is "fair" if on average, the player's fortune neither increases nor decreases at each play. The martingale property (b) requires the player's fortune after the next play to equal, on average, his current fortune and not be otherwise affected by previous history. Some early work in martingale theory was motivated in part by problems in gambling. For example, *martingale systems theorems* consider whether an astute choice of betting strategy can turn a fair game into a favorable one, and the name "martingale" derives from a French term for the particular strategy of doubling one's bets until a win is secured. While it remains popular to illustrate martingale concepts with gambling examples, today, martingale theory has such broad scope and diverse applications that to think of it purely in terms of gambling would be unduly restrictive and misleading.

Example *Stock Prices in a Perfect Market* Let X_n be the closing price at the end of day n of a certain publicly traded security such as a share of stock. While daily prices may fluctuate, many scholars believe that, in a perfect market, these price sequences should be martingales. In a perfect market freely open to all, they argue, it should not be possible to predict with any degree of accuracy whether a future price X_{n+1} will be higher or lower than the current price X_n. For example, if a future price could be expected to be higher, then a number of buyers would enter the market, and their demand would raise the current price X_n. Similarly, if a future price could be predicted as lower, a number of sellers would appear and tend to depress the current price. Equilibrium obtains where the future price cannot be predicted, on average, as higher or lower, i.e., where price sequences are martingales.

2.5.2 The Markov Inequality

What does the mean of a random variable tell us about its distribution? For a nonnegative random variable X, Markov's inequality is $\lambda \Pr\{X \geq \lambda\} \leq E[X]$, for any positive constant λ. For example, if $E[X] = 1$, then $\Pr\{X \geq 4\} \leq \frac{1}{4}$, no matter what the actual distribution of X is. The proof uses two properties: (i) $X \geq 0$ (X is a nonnegative random variable), and (ii) $E[X\mathbf{1}\{X \geq \lambda\}] \geq \lambda \Pr\{X \geq \lambda\}$. (Recall that $\mathbf{1}(A)$ is the *indicator* of an event A and is one if A occurs and zero otherwise. See Chapter 1, Section 1.3.1.) Then, by the law of total probability,

$$\begin{aligned} E[X] &= E\left[X1_{[\lambda,\infty)}(X)\right] + E\left[X1_{(-\infty,\lambda)}(X)\right] \\ &\geq E\left[X1_{[\lambda,\infty)}(X)\right] \\ &\geq \lambda \Pr[X \geq \lambda] \end{aligned}$$

and Markov's inequality results.

2.5.3 The Maximal Inequality for Nonnegative Martingales

Because a martingale has constant mean, Markov's inequality applied to a nonnegative martingale immediately yields

$$\Pr\{X_n \geq \lambda\} \leq \frac{E[X_0]}{\lambda}, \quad \lambda > 0.$$

We will extend the reasoning behind Markov's inequality to achieve an inequality of far greater power:

$$\Pr\left\{\max_{0 \leq n \leq m} X_n \geq \lambda\right\} \leq \frac{E[X_0]}{\lambda}. \tag{2.49}$$

Instead of limiting the probability of a large value for a single observation X_n, the *maximal inequality* (2.49) limits the probability of observing a large value anywhere in the time interval $0, \ldots, m$, and since the right side of (2.49) does not depend on the length of the interval, the maximal inequality limits the probability of observing a large value at any time in the infinite future of the martingale!

In order to prove the maximal inequality for nonnegative martingales, we need but a single additional fact: If X and Y are jointly distributed random variables and B is an arbitrary set, then

$$E[X\,1_B(Y)] = E[E(X|Y)\,1_B(Y)] \tag{2.50}$$

But (2.50) follows from the conditional expectation property (2.12), $E[g(X)h(Y)] = E\{h(Y)E[g(X)|Y]\}$, with $g(x) = x$ and $h(y) = \mathbf{1}(y \text{ in } B\}$. We will have need of (2.50) with $X = X_m$ and $Y = (X_0, \ldots, X_n)$, whereupon (2.50) followed by (2.48) then justifies

$$\begin{aligned} &E[X_m \mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}] \\ &\quad = E[E\{X_m|X_0, \ldots, X_n\}\mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}] \\ &\quad = E[X_n \mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}]. \end{aligned} \tag{2.51}$$

Theorem 2.1. *Let $X_0, X_1, \ldots$ be a martingale with nonnegative values; i.e.,* $\Pr\{X_n \geq 0\} = 1$ *for* $n = 0, 1, \ldots$. *For any* $\lambda > 0$*,*

$$\Pr\left\{\max_{0\leq n\leq m} X_n \geq \lambda\right\} \leq \frac{E[X_0]}{\lambda}, \quad \text{for } 0 \leq n \leq m \tag{2.52}$$

and

$$\Pr\left\{\max_{n\geq 0} X_n > \lambda\right\} \leq \frac{E[X_0]}{\lambda}, \quad \text{for all } n. \tag{2.53}$$

Proof. Inequality (2.53) follows from (2.52) because the right side of (2.52) does not depend on m. We begin with the law of total probability, as in Chapter 1, Section 1.2.1. Either the $\{X_0, \ldots, X_m\}$ sequence rises above λ for the first time at some index n or else it remains always below λ. As these possibilities are mutually exclusive and exhaustive, we apply the law of total probability to obtain

$$\begin{aligned} E[X_m] &= \sum_{n=0}^{m} E[X_m \mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\} \\ &\quad + E[X_m \mathbf{1}\{X_0 < \lambda, \ldots, X_m < \lambda\}] \\ &\geq \sum_{n=0}^{m} E[X_m \mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}] \quad (X_m \geq 0) \end{aligned}$$

$$= \sum_{n=0}^{m} E[X_n \mathbf{1}\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}] \quad \text{[using (2.51)]}$$
$$\geq \lambda \sum_{n=0}^{m} \Pr\{X_0 < \lambda, \ldots, X_{n-1} < \lambda, X_n \geq \lambda\}$$
$$= \lambda \Pr\left\{\max_{0 \leq n \leq m} X_n \geq \lambda\right\}.$$

■

Example A gambler begins with a unit amount of money and faces a series of independent fair games. Beginning with $X_0 = 1$, the gambler bets the amount p, $0 < p < 1$. If the first game is a win, which occurs with probability $\frac{1}{2}$, the gambler's fortune is $X_1 = 1 + pX_0 = 1 + p$. If the first game is a loss, then $X_1 = 1 - pX_0 = 1 - p$. After the nth play and with a current fortune of X_n, the gambler wagers pX_n, and

$$X_{n+1} = \begin{cases} (1+p)X_n & \text{with probability } \dfrac{1}{2}, \\ (1-p)X_n & \text{with probability } \dfrac{1}{2}. \end{cases}$$

Then, $\{X_n\}$ is a nonnegative martingale, and the maximal inequality (2.52) with $\lambda = 2$, e.g., asserts that *the probability that the gambler ever doubles his money is less than or equal to* $\frac{1}{2}$, *and this holds no matter what the game is, as long as it is fair*, *and no matter what fraction p of his fortune is wagered at each play*. Indeed, the fraction wagered may vary from play to play, as long as it is chosen without knowledge of the next outcome.

As amply demonstrated by this example, the maximal inequality is a very strong statement. Indeed, more elaborate arguments based on the maximal and other related martingale inequalities are used to show that a nonnegative martingale converges: If $\{X_n\}$ is a nonnegative martingale, then there exists a random variable, let us call it X_∞, for which $\lim_{n\to\infty} X_n = X_\infty$. We cannot guarantee the equality of the expectations in the limit, but the inequality $E[X_0] \geq E[X_\infty] \geq 0$ can be established.

Example In Chapter 3, Section 3.8, we will introduce the branching process model for population growth. In this model, X_n is the number of individuals in the population in the nth generation, and $\mu > 0$ is the mean family size or expected number of offspring of any single individual. The mean population size in the nth generation is $X_0\mu^n$. In this branching process model, X_n/μ^n is a nonnegative martingale (see Chapter 3, Problem 3.8.4), and the maximal inequality implies that the probability of the actual population ever exceeding 10 times the mean size is less than or equal to 1/10. The nonnegative martingale convergence theorem asserts that the evolution of such a population after many generations may be described by a single random variable X_∞ in the form

$$X_n \approx X_\infty \mu^n, \quad \text{for large } n.$$

Example *How NOT to generate a uniformly distributed random variable* An urn initially contains one red and one green ball. A ball is drawn at random and it is returned to the urn, together with another ball of the same color. This process is repeated indefinitely. After the nth play, there will be a total of $n+2$ balls in the urn. Let R_n be the number of these balls that are red, and $X_n = R_n/(n+2)$ the fraction of red balls. We claim that $\{X_n\}$ is a martingale. First, observe that

$$R_{n+1} = \begin{cases} R_n + 1 & \text{with probability } X_n \\ R_n & \text{with probability } 1 - X_n \end{cases}$$

so that

$$E[R_{n+1}|X_n] = R_n + X_n = X_n(2+n+1),$$

and finally,

$$E[X_{n+1}|X_n] = \frac{1}{n+3}E[R_{n+1}|X_n] = \frac{2+n+1}{n+3}X_n = X_n.$$

This verifies the martingale property, and because such a fraction is always nonnegative, indeed, between 0 and 1, there must be a random variable X_∞ to which the martingale converges. We will derive the probability distribution of the random limit. It is immediate that R_1 is equally likely to be 1 or 2, since the first ball chosen is equally likely to be red or green. Continuing,

$$\begin{aligned}\Pr\{R_2 = 3\} &= \Pr\{R_2 = 3|R_1 = 2\}\Pr\{R_1 = 2\} \\ &= \left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{3};\end{aligned}$$

$$\begin{aligned}\Pr\{R_2 = 2\} &= \Pr\{R_2 = 2|R_1 = 1\}\Pr\{R_1 = 1\} \\ &\quad + \Pr\{R_2 = 2|R_1 = 2\}\Pr\{R_1 = 2\} \\ &= \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{3};\end{aligned}$$

and since the probabilities must sum to 1,

$$\Pr\{R_2 = 1\} = \frac{1}{3}.$$

By repeating these simple calculations, it is easy to see that

$$\Pr\{R_n = k\} = \frac{1}{\mathrm{n}+1} \quad \text{for } k = 1, 2, \ldots, n+1,$$

and that, therefore, X_n is uniformly distributed over the values $1/(n+2)$, $2/(n+2), \ldots, (n+1)/(n+2)$. This uniform distribution must prevail in the limit, which leads to

$$\Pr\{X_\infty \le x\} = x \quad \text{for } 0 < x < 1.$$

Think about this remarkable result for a minute! If you sit down in front of such an urn and play this game, eventually the fraction of red balls in your urn will stabilize in the near vicinity of some value, call it U. If I play the game, the fraction of red balls in my urn will stabilize also, but at another value, U'. Anyone who plays the game will find the fraction of red balls in the urn tending toward some limit, but everyone will experience a different limit. In fact, each play of the game generates a fresh, uniformly distributed random variable, in the limit. Of course, there may be faster and simpler ways to generate uniformly distributed random variables.

Martingale implications include many more inequalities and convergence theorems. As briefly mentioned at the start, there are so-called *systems theorems* that delimit the conditions under which a gambling system, such as doubling the bets until a win is secured, can turn a fair game into a winning game. A deeper discussion of martingale theory would take us well beyond the scope of this introductory text, and our aim must be limited to building an enthusiasm for further study. Nevertheless, a large variety of important martingales will be introduced in the Problems at the end of each section in the remainder of the book.

Exercises

2.5.1 Let X be an exponentially distributed random variable with mean $E[X] = 1$. For $x = 0.5, 1$, and 2, compare $\Pr\{X > x\}$ with the Markov inequality bound $E[X]/x$.

2.5.2 Let X be a Bernoulli random variable with parameter p. Compare $\Pr\{X \ge 1\}$ with the Markov inequality bound.

2.5.3 Let ξ be a random variable with mean μ and standard deviation σ. Let $X = (\xi - \mu)^2$. Apply Markov's inequality to X to deduce Chebyshev's inequality:

$$\Pr\{|\xi - \mu| \ge \varepsilon\} \le \frac{\sigma^2}{\varepsilon^2} \quad \text{for any } \varepsilon > 0.$$

Problems

2.5.1 Use the law of total probability for conditional expectations $E[E\{X|Y,Z\}|Z] = E[X|Z]$ to show

$$E[X_{n+2}|X_0, \ldots, X_n] = E[E\{X_{n+2}|X_0, \ldots, X_{n+1}\}|X_0, \ldots, X_n].$$

Conclude that when X_n is a martingale,

$$E[X_{n+2}|X_0, \ldots, X_n] = X_n.$$

2.5.2 Let $U_1, U_2, \ldots$ be independent random variables each uniformly distributed over the interval $(0, 1]$. Show that $X_0 = 1$ and $X_n = 2^n U_1 \cdots U_n$ for $n = 1, 2, \ldots$ defines a martingale.

2.5.3 Let $S_0 = 0$, and for $n \geq 1$, let $S_n = \varepsilon_1 + \cdots + \varepsilon_n$ be the sum of n independent random variables, each exponentially distributed with mean $E[\varepsilon] = 1$. Show that

$$X_n = 2^n \exp(-S_n), \quad n \geq 0$$

defines a martingale.

2.5.4 Let $\xi_1, \xi_2, \ldots$ be independent Bernoulli random variables with parameter $p, 0 < p < 1$. Show that $X_0 = 1$ and $X_n = p^{-n}\xi_1 \cdots \xi_n, n = 1, 2, \ldots$, defines a nonnegative martingale. What is the limit of X_n as $n \to \infty$?

2.5.5 Consider a stochastic process that evolves according to the following laws: If $X_n = 0$, then $X_{n+1} = 0$, whereas if $X_n > 0$, then

$$X_{n+1} = \begin{cases} X_n + 1 & \text{with probability } \dfrac{1}{2} \\[2ex] X_n - 1 & \text{with probability } \dfrac{1}{2}. \end{cases}$$

(a) Show that X_n is a nonnegative martingale.

(b) Suppose that $X_0 = i > 0$. Use the maximal inequality to bound

$$\Pr\{X_n \geq N \quad \text{for some } n \geq 0 | X_0 = i\}.$$

Note: X_n represents the fortune of a player of a fair game who wagers \$1 at each bet and who is forced to quit if all money is lost ($X_n = 0$). This *gambler's ruin* problem is discussed fully in Chapter 3, Section 3.5.3.