

Chapter 7

Multiple Discrete Random Variables

Thus far, our treatment of probability has been focused on single random variables. It is often convenient or required to model stochastic phenomena using multiple random variables. In this section, we extend some of the concepts developed for single random variables to multiple random variables. We center our exposition of random vectors around the simplest case, pairs of random variables.

7.1 Joint Probability Mass Functions

Consider two discrete random variables X and Y associated with a single experiment. The random pair (X, Y) is characterized by the *joint probability mass function* of X and Y , which we denote by $p_{X,Y}(\cdot, \cdot)$. If x is a possible value of X and y is a possible value of Y , then the probability mass function of (x, y) is denoted by

$$\begin{aligned} p_{X,Y}(x, y) &= \Pr(\{X = x\} \cap \{Y = y\}) \\ &= \Pr(X = x, Y = y). \end{aligned}$$

Note the similarity between the definition of the joint PMF and (5.1).

Suppose that S is a subset of $X(\Omega) \times Y(\Omega)$. We can express the probability

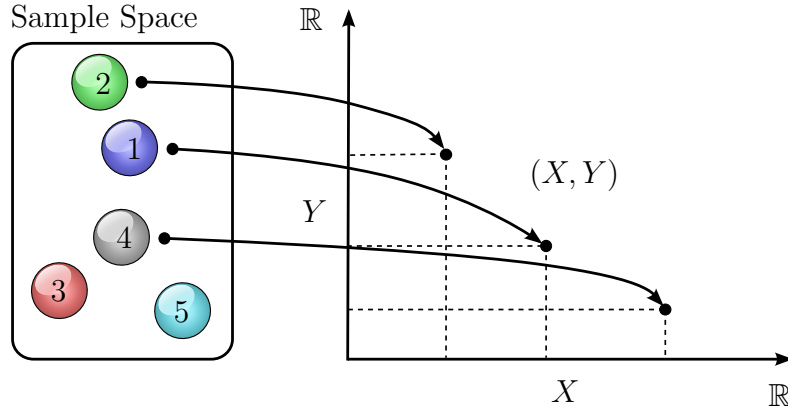


Figure 7.1: The random pair (X, Y) maps every outcome contained in the sample space to a vector in \mathbb{R}^2 .

of S as

$$\begin{aligned} \Pr(S) &= \Pr(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in S\}) \\ &= \sum_{(x,y) \in S} p_{X,Y}(x, y). \end{aligned}$$

In particular, we have

$$\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = 1.$$

To further distinguish between the joint PMF of (X, Y) and the individual PMFs $p_X(\cdot)$ and $p_Y(\cdot)$, we occasionally refer to the latter as *marginal probability mass functions*. We can compute the marginal PMFs of X and Y from the joint PMF $p_{X,Y}(\cdot, \cdot)$ using the formulas

$$\begin{aligned} p_X(x) &= \sum_{y \in Y(\Omega)} p_{X,Y}(x, y), \\ p_Y(y) &= \sum_{x \in X(\Omega)} p_{X,Y}(x, y). \end{aligned}$$

On the other hand, knowledge of the marginal distributions $p_X(\cdot)$ and $p_Y(\cdot)$ is not enough to obtain a complete description of the joint PMF $p_{X,Y}(\cdot, \cdot)$. This fact is illustrated in Examples 56 & 57.

Example 56. An urn contains three balls numbered one, two and three. A random experiment consists of drawing two balls from the urn, without replacement. The number appearing on the first ball is a random variable, which we denote by X . Similarly, we refer to the number inscribed on the second ball as Y . The joint PMF of X and Y is specified in table form below,

$p_{X,Y}(x,y)$	1	2	3
1	0	1/6	1/6
2	1/6	0	1/6
3	1/6	1/6	0

We can compute the marginal PMF of X as

$$p_X(x) = \sum_{y \in Y(\Omega)} p_{X,Y}(x,y) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

where $x \in \{1, 2, 3\}$. Likewise, the marginal PMF of Y is given by

$$p_Y(y) = \begin{cases} 1/3, & \text{if } y \in \{1, 2, 3\} \\ 0, & \text{otherwise.} \end{cases}$$

Example 57. Again, suppose that an urn contains three balls numbered one, two and three. This time the random experiment consists of drawing two balls from the urn with replacement. We use X and Y to denote the numbers appearing on the first and second balls, respectively. The joint PMF of X and Y becomes

$p_{X,Y}(x,y)$	1	2	3
1	1/9	1/9	1/9
2	1/9	1/9	1/9
3	1/9	1/9	1/9

The marginal distributions of X and Y are the same as in Example 56; however, the joint PMFs differ.

7.2 Functions and Expectations

Let X and Y be two random variables with joint PMF $p_{X,Y}(\cdot, \cdot)$. Consider a third random variable defined by $V = g(X, Y)$, where $g(\cdot, \cdot)$ is a real-valued

function. We can obtain the PMF of V by computing

$$p_V(v) = \sum_{\{(x,y)|g(x,y)=v\}} p_{X,Y}(x,y). \quad (7.1)$$

This equation is the analog of (5.3) for pairs of random variables.

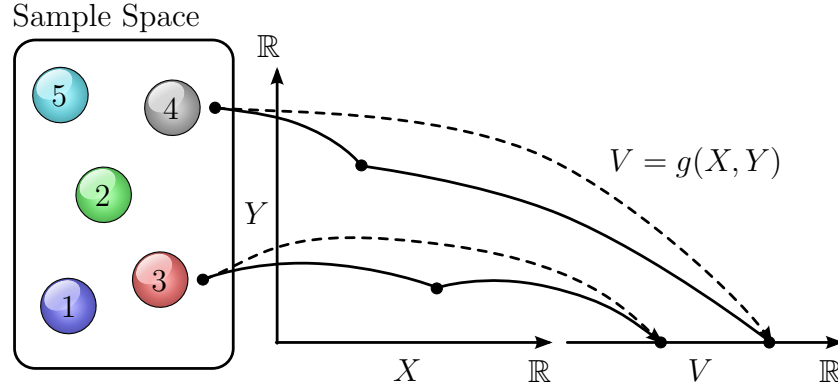


Figure 7.2: A real-valued function of two random variables, X and Y , is also a random variable. Above, $V = g(X, Y)$ maps elements of the sample space to real numbers.

Example 58. *Two dice, a blue die and a red one, are rolled simultaneously. The random variable X represents the number of dots that appears on the top face of the blue die, whereas Y denotes the number of dots on the red die. We can form a random variable U that describes the sum of these two dice, $U = X + Y$.*

The lowest possible value for U is two, and its maximum value is twelve. The PMF of U , as calculated using (7.1), appears in table form below

k	2, 12	3, 11	4, 10	5, 9	6, 8	7
$p_U(k)$	1/36	1/18	1/12	1/9	5/36	1/6

The definition of the expectation operator can be extended to multiple random variables. In particular, the expected value of $g(X, Y)$ is obtained by computing

$$E[g(X, Y)] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x, y) p_{X,Y}(x, y). \quad (7.2)$$

Example 59. *An urn contains three balls numbered one, two and three. Two balls are selected from the urn at random, without replacement. We employ X and Y to represent the numbers on the first and second balls, respectively. We wish to compute the expected value of the function $X + Y$.*

Using (7.2), we compute the expectation of $g(X, Y) = X + Y$ as

$$\begin{aligned} E[g(X, Y)] &= E[X + Y] \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (x + y)p_{X,Y}(x, y) \\ &= \sum_{x \in X(\Omega)} xp_X(x) + \sum_{y \in Y(\Omega)} yp_Y(y) = 4. \end{aligned}$$

The expected value of $X + Y$ is four.

7.3 Conditional Random Variables

Many events of practical interest are dependent. That is, knowledge about event A may provide partial information about the realization of event B . This inter-dependence is captured by the concept of conditioning, which was first discussed in Chapter 4. In this section, we extend the concept of conditioning to multiple random variables. We study the probability of events related to the random variable Y given knowledge of some event S or information about another random variable X .

7.3.1 Conditioning on Events

First, we consider the conditional PMF of a random variable X conditioned on the event S , where $\Pr(X \in S) > 0$. Let X be a random variable associated with a particular experiment, and let S be a non-trivial event corresponding to this experiment. The conditional PMF of X given S is defined by

$$p_{X|S}(x) = \Pr(X = x|S) = \frac{\Pr(\{X = x\} \cap S)}{\Pr(S)}. \quad (7.3)$$

Note that the events $\{\omega \in \Omega | X(\omega) = x\}$ form a partition of Ω as x ranges over all the values in $X(\Omega)$. Using the total probability theorem, we gather that

$$\sum_{x \in X(\Omega)} \Pr(\{X = x\} \cap S) = \Pr(S)$$

and, consequently, we get

$$\begin{aligned} \sum_{x \in X(\Omega)} p_{X|S}(x) &= \sum_{x \in X(\Omega)} \frac{\Pr(\{X = x\} \cap S)}{\Pr(S)} \\ &= \frac{\sum_{x \in X(\Omega)} \Pr(\{X = x\} \cap S)}{\Pr(S)} = 1. \end{aligned}$$

We conclude that $p_{X|S}(\cdot)$ is a valid PMF.

Example 60. *A data packet is sent to a destination over an unreliable wireless communication link. Data is successfully decoded at the receiver with probability p , and the transmission fails with probability $(1 - p)$, independently of previous trials. When initial decoding fails, a retransmission is requested immediately. Let X denote the number of trials required for success.*

Consider another system which drops packets. If packet transmission fails n consecutive times, then the data is dropped from the queue altogether, with no further transmission attempt. Let Y denote the number of trials for this system, and let S be the event that the data packet is successfully transmitted. We wish to compute the conditional PMF of Y given S .

First, we note that a successful transmission can happen at any of n possible instants. These outcomes being mutually exclusive, the probability of S can be written as

$$\Pr(S) = \Pr(\{X \leq n\}) = \sum_{k=1}^n (1-p)^{k-1} p = 1 - (1-p)^n.$$

Thus, the conditional PMF $p_{Y|S}(\cdot)$ is equal to

$$p_{Y|S}(k) = p_{X|X \leq n}(k) = \frac{(1-p)^{k-1} p}{1 - (1-p)^n},$$

where $k = 1, 2, \dots, n$.

7.3.2 Conditioning on Random Variables

Let X and Y be two random variables associated with a same experiment. The *conditional probability mass function* of Y given $X = x$, which we write

$p_{Y|X}(\cdot|\cdot)$, is defined by

$$\begin{aligned} p_{Y|X}(y|x) &= \Pr(Y = y|X = x) \\ &= \frac{\Pr(\{Y = y\} \cap \{X = x\})}{\Pr(X = x)} \\ &= \frac{p_{X,Y}(x, y)}{p_X(x)}, \end{aligned} \tag{7.4}$$

provided that $p_X(x) \neq 0$. Note that conditioning on $X = x$ is not possible when $p_X(x)$ vanishes, as it has no relevant meaning. This is similar to conditional probabilities being only defined for conditional events with non-zero probabilities.

Let x be fixed with $p_X(x) > 0$. The conditional PMF introduced above is a valid PMF since it is nonnegative and

$$\begin{aligned} \sum_{y \in Y(\Omega)} p_{Y|X}(y|x) &= \sum_{y \in Y(\Omega)} \frac{p_{X,Y}(x, y)}{p_X(x)} \\ &= \frac{1}{p_X(x)} \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = 1. \end{aligned}$$

The probability that Y belongs to S , given $X = x$, is obtained by summing the conditional probability $p_{Y|X}(\cdot|x)$ over all outcomes included in S ,

$$\Pr(Y \in S|X = x) = \sum_{y \in S} p_{Y|X}(y|x).$$

It turns out that conditioning on an event, as described in (7.3), is essentially equivalent to conditioning on the value of another random variable, as in (7.4). First, we observe that $p_{X|S}$ can be written in term of a conditional PMF. Let $Y = \mathbf{1}_S(\cdot)$ symbolize the indicator function of S , with

$$\mathbf{1}_S(\omega) = \begin{cases} 1, & \omega \in S \\ 0, & \omega \notin S. \end{cases}$$

Then $p_{X|S}(x) = p_{X|Y}(x|1)$, where $p_{X|Y}$ is the conditional probability mass function (PMF) of X given Y and will be introduced next. In this sense, conditioning on an event can be seen as a special case of a conditional PMF. Likewise, the conditional PMF of X given Y can be expressed as the conditional probability of X given the event $Y = y$ using the notation

$$p_{X|Y}(x|y) = p_{X|Y=y}(x).$$

Example 61 (*Hypergeometric Random Variable). A wireless communication channel is employed to transmit a data packet that contains a total of m bits. The first n bits of this message are dedicated to the packet header, which stores very sensitive information. The wireless connection is unreliable; every transmitted bit is received properly with probability $1 - p$ and erased with probability p , independently of other bits. We wish to derive the conditional probability distribution of the number of erasures located in the header, given that the total number of corrupted bits in the packet is equal to c .

Let H represent the number of erasures in the header, and denote the number of corrupted bits in the entire message by C . The conditional probability mass function of H given $C = c$ is equal to

$$\begin{aligned} p_{H|C}(h|c) &= \frac{p_{H,C}(h,c)}{p_C(c)} \\ &= \frac{\binom{n}{h}(1-p)^{n-h}p^h \binom{m-n}{c-h}(1-p)^{(m-n)-(c-h)}p^{c-h}}{\binom{m}{c}(1-p)^{m-c}p^c} \\ &= \frac{\binom{n}{h} \binom{m-n}{c-h}}{\binom{m}{c}}, \end{aligned}$$

where $h = 0, 1, \dots, \min\{c, n\}$. Clearly, the conditioning affects the probability distribution of the number of corrupted bits in the header. In general, a random variable with such a distribution is known as a hypergeometric random variable.

The definition of conditional PMF can be rearranged to obtain a convenient formula to calculate the joint distribution of X and Y , namely

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x) = p_{X|Y}(x|y)p_Y(y).$$

This formula can be used to compute the joint PMF of X and Y sequentially.

Example 62 (Splitting Property of Poisson PMF). A digital communication system sends out either a one with probability p or a zero with probability $1 - p$, independently of previous transmissions. The number of transmitted binary digits within a given time interval has a Poisson PMF with parameter λ . We wish to show that the number of ones sent in that same time interval has a Poisson PMF with parameter $p\lambda$.

Let M denote the number of ones within the stipulated interval, N be the number of zeros, and $K = M + N$ be the total number of bits sent during the same interval. The number of ones given that the total number of transmissions is k is given by

$$p_{M|K}(m|k) = \binom{k}{m} p^m (1-p)^{k-m}, \quad m = 0, 1, \dots, k.$$

The probability that M is equal to m is therefore equal to

$$\begin{aligned} p_M(m) &= \sum_{k=0}^{\infty} p_{K,M}(k, m) = \sum_{k=0}^{\infty} p_{M|K}(m|k) p_K(k) \\ &= \sum_{k=m}^{\infty} \binom{k}{m} p^m (1-p)^{k-m} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{u=0}^{\infty} \binom{u+m}{m} p^m (1-p)^u \frac{\lambda^{u+m}}{(u+m)!} e^{-\lambda} \\ &= \frac{(\lambda p)^m}{m!} e^{-\lambda} \sum_{u=0}^{\infty} \frac{((1-p)\lambda)^u}{u!} = \frac{(\lambda p)^m}{m!} e^{-p\lambda}. \end{aligned}$$

Above, we have used the change of variables $k = u + m$. We have also rearranged the sum into a familiar form, leveraging the fact that the summation of a Poisson PMF over all possible values is equal to one. We can see that M has a Poisson PMF with parameter $p\lambda$.

7.4 Conditional Expectations

The *conditional expectation* of Y given $X = x$ is simply the expectation of Y with respect to the conditional PMF $p_{Y|X}(\cdot|x)$,

$$\mathbb{E}[Y|X = x] = \sum_{y \in Y(\Omega)} y p_{Y|X}(y|x).$$

This conditional expectation can be viewed as a function of x ,

$$h(x) = \mathbb{E}[Y|X = x].$$

It is therefore mathematically accurate and sometimes desirable to talk about the random variable $h(X) = \mathbb{E}[Y|X]$. In particular, if X and Y are two random variables associated with an experiment, the outcome of this experiment

determines the value of X , say $X = x$, which in turn yields the conditional expectation $h(x) = E[Y|X = x]$. From this point of view, the conditional expectation $E[Y|X]$ is simply an instance of a random variable.

Not too surprisingly, the expectation of $E[Y|X]$ is equal to $E[Y]$, as evinced by the following derivation,

$$\begin{aligned} E[E[Y|X]] &= \sum_{x \in X(\Omega)} E[Y|X = x]p_X(x) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} yp_{Y|X}(y|x)p_X(x) \\ &= \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega)} yp_{X,Y}(x, y) \\ &= \sum_{y \in Y(\Omega)} yp_Y(y) = E[Y]. \end{aligned}$$

Using a similar argument, it is straightforward to show that

$$E[E[g(Y)|X]] = E[g(Y)].$$

Example 63. *An entrepreneur opens a small business that sells two kinds of beverages from the Brazos Soda Company, cherry soda and lemonade. The number of bottles sold in an hour at the store is found to be a Poisson random variable with parameter $\lambda = 10$. Every customer selects a cherry soda with probability p and a lemonade with probability $(1 - p)$, independently of other customers. We wish to find the conditional mean of the number of cherry sodas purchased in an hour given that ten beverages were sold to customers during this time period.*

Let B represent the number of bottles sold during an hour. Similarly, let C and L be the number of cherry sodas and lemonades purchased during the same time interval, respectively. We note that $B = C + L$. The conditional PMF of C given that the total number of beverages sold equals ten is

$$p_{C|B}(k|10) = \binom{10}{k} p^k (1 - p)^{10-k}. \quad (7.5)$$

This follows from the fact that every customer selects a cherry soda with probability p , independently of other customers. The conditional mean is then seen to equal

$$E[C|B = 10] = \sum_{k=0}^{10} k \binom{10}{k} p^k (1 - p)^{10-k} = 10p.$$

The same approach can be used to show that $E[C|B] = pB$ (i.e., the conditional expectation is a random variable).

We can define the expectation of X conditioned on event S in an analogous fashion. Let S be an event such that $\Pr(S) > 0$. The conditional expectation of X given S is

$$E[X|S] = \sum_{x \in X(\Omega)} xp_{X|S}(x),$$

where $p_{X|S}(\cdot)$ is defined in (7.3). Similarly, we have

$$E[g(X)|S] = \sum_{x \in X(\Omega)} g(x)p_{X|S}(x),$$

Example 64. *Spring break is coming and a student decides to renew his shirt collection. The number of shirts purchased by the student is a random variable denoted by N . The PMF of this random variable is a geometric distribution with parameter $p = 0.5$. Any one shirt costs \$10, \$20 or \$50 with respective probabilities 0.5, 0.3 and 0.2, independently of other shirts. We wish to compute the expected amount of money spent by the student during his shopping spree. Also, we wish to compute the expected amount of money disbursed given that the student buys at least five shirts.*

Let C_i be the cost of the i th shirt. The total amount of money spent by the student, denoted by T , can be expressed as

$$T = \sum_{i=1}^N C_i.$$

The mean of T can be computed using nested conditional expectation. It is equal to

$$\begin{aligned} E[T] &= E \left[\sum_{i=1}^N C_i \right] = E \left[E \left[\sum_{i=1}^N C_i \middle| N \right] \right] \\ &= E \left[\sum_{i=1}^N E[C_i|N] \right] = 21E[N] = 42. \end{aligned}$$

The student is expected to spend \$42. Given that the student buys at least five

shirts, the conditional expectation becomes

$$\begin{aligned} \mathbf{E}[T|N \geq 5] &= \mathbf{E}\left[\sum_{i=1}^N C_i \mid N \geq 5\right] \\ &= \mathbf{E}\left[\mathbf{E}\left[\sum_{i=1}^N C_i \mid N\right] \mid N \geq 5\right] \\ &= 21\mathbf{E}[N|N \geq 5] = 126. \end{aligned}$$

Conditioned on buying more than five shirts, the student is expected to spend \$126. Note that, in computing the conditional expectation, we have utilized the memoryless property of the geometric random variable,

$$\mathbf{E}[N|N \geq 5] = \mathbf{E}[N|N > 4] = 4 + \mathbf{E}[N] = 6.$$

7.5 Independence

Let X and Y be two random variables associated with a same experiment. We say that X and Y are *independent random variables* if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

for every $x \in X(\Omega)$ and every $y \in Y(\Omega)$.

There is a clear relation between the concept of independence introduced in Section 4.4 and the independence of two random variables. Random variables X and Y are independent if and only if the events $\{X = x\}$ and $\{Y = y\}$ are independent for every pair (x, y) such that $x \in X(\Omega)$ and $y \in Y(\Omega)$.

Proposition 6. *If X and Y are independent random variables, then*

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y].$$

Proof. Assume that both $\mathbf{E}[X]$ and $\mathbf{E}[Y]$ exist, then

$$\begin{aligned} \mathbf{E}[XY] &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xyp_{X,Y}(x, y) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xyp_X(x)p_Y(y) \\ &= \sum_{x \in X(\Omega)} xp_X(x) \sum_{y \in Y(\Omega)} yp_Y(y) = \mathbf{E}[X]\mathbf{E}[Y], \end{aligned}$$

where we have used the fact that $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for independent random variables. \square

We saw in Section 7.3 that a random variable can also be conditioned on a specific event. Let X be a random variable and let S be a non-trivial event. The variable X is *independent* of S if

$$\Pr(\{X = x\} \cap S) = p_X(x) \Pr(S)$$

for every $x \in X(\Omega)$. In particular, if X is independent of event S then

$$p_{X|S}(x) = p_X(x)$$

for all $x \in X(\Omega)$.

Example 65. *Two dice of different colors are rolled, as in Example 58. We wish to compute the expected value of their product. We know that the mean of each role is $E[X] = E[Y] = 3.5$. Furthermore, it is straightforward to show that X and Y , the numbers of dots on the two dice, are independent random variables. The expected value of their product is then equal to the product of the individual means, $E[XY] = E[X]E[Y] = 12.25$.*

We can parallel the proof of Proposition 6 to show that

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)] \quad (7.6)$$

whenever X and Y are independent random variables and the corresponding expectations exist.

Example 66. *Let X, Y be random variables and $Z = X + Y$ be their sum. We have already seen that the expectation satisfies $E[Z] = E[X] + E[Y]$. What about the variance of Z ? In this case, we find that*

$$\begin{aligned} \text{Var}[Z] &= E[(X + Y - E[X + Y])^2] \\ &= E[((X - E[X]) + (Y - E[Y]))^2] \\ &= E[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2] \\ &= \text{Var}[X] + 2E[(X - E[X])(Y - E[Y])] + \text{Var}[Y]. \end{aligned}$$

So, the variance of the sum is equal to the sum of the variance and an extra cross-term. But, if the random variables X, Y are independent, then one finds that the cross term is

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]] \mathbb{E}[Y - \mathbb{E}[Y]] = 0.$$

An important consequence of this fact is that the variance of the sum equals the sum of the variance for independent random variables.

We turn to the question of determining the distribution of a sum of two independent random variables in terms of the marginal PMF of the summands. Suppose X and Y are independent random variables that take on integer values. Let $p_X(\cdot)$ and $p_Y(\cdot)$ be the PMFs of X and Y , respectively. We wish to determine the distribution $p_U(\cdot)$ of U , where $U = X + Y$. To accomplish this task, it suffices to compute the probability that U assumes a value k , where k is an arbitrary integer,

$$\begin{aligned} p_U(k) &= \Pr(U = k) = \Pr(X + Y = k) \\ &= \sum_{\{(x,y) \in X(\Omega) \times Y(\Omega) | x+y=k\}} p_{X,Y}(x, y) \\ &= \sum_{m \in \mathbb{Z}} p_{X,Y}(m, k - m) \\ &= \sum_{m \in \mathbb{Z}} p_X(m) p_Y(k - m). \end{aligned}$$

The latter operation is called a *discrete convolution*. In particular, the PMF of $U = X + Y$ is the discrete convolution of $p_X(\cdot)$ and $p_Y(\cdot)$ given by

$$p_U(k) = (p_X * p_Y)(k) = \sum_{m \in \mathbb{Z}} p_X(m) p_Y(k - m). \quad (7.7)$$

It is easy to verify that discrete convolution is commutative and associative.

7.6 Ordinary Generating Functions

In the special yet important case where $X(\Omega)$ is a subset of the non-negative integers, it is occasionally useful to employ the *ordinary generating function*. This function bears a close resemblance to the z -transform and is defined by

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k=0}^{\infty} z^k p_X(k). \quad (7.8)$$

It is also called the *probability-generating function* because of the following property. The probability $\Pr(X = k) = p_X(k)$ can be recovered from the corresponding generating function $G_X(z)$ through Taylor series expansion. Within the radius of convergence of $G_X(z)$, we have

$$G_X(z) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{d^k G_X}{dz^k}(0) \right) z^k.$$

Comparing this equation to (7.8), we conclude that

$$p_X(k) = \frac{1}{k!} \frac{d^k G_X}{dz^k}(0).$$

We note that, for $|z| \leq 1$, we have

$$\left| \sum_{k=0}^{\infty} z^k p_X(k) \right| \leq \sum_{k=0}^{\infty} |z|^k p_X(k) \leq \sum_{k=0}^{\infty} p_X(k) = 1$$

and hence the radius of convergence of any probability-generating function must include one.

The ordinary generating function plays an important role in dealing with sums of discrete random variables. As a preview of what lies ahead, we compute ordinary generating functions for Bernoulli and Binomial random variables below.

Example 67. Let X be a Bernoulli random variable with parameter p . The ordinary generating function of X is given by

$$G_X(z) = p_X(0) + p_X(1)z = 1 - p + pz.$$

Example 68. Let S be a binomial random variable with parameters n and p . The ordinary generating function of S can be computed as

$$\begin{aligned} G_S(z) &= \sum_{k=0}^n z^k p_S(k) = \sum_{k=0}^n z^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (pz)^k (1-p)^{n-k} = (1-p + pz)^n. \end{aligned}$$

We know, from Section 5.2.2, that one way to create a binomial random variable is to sum n independent and identically distributed Bernoulli random

variables, each with parameter p . Looking at the ordinary generating functions above, we notice that the $G_S(z)$ is the product of n copies of $G_X(z)$. Intuitively, it appears that the sum of independent discrete random variables leads to the product of their ordinary generating functions, a relation that we will revisit shortly.

The mean and second moment of X can be computed based on its ordinary generating function. In particular, we have

$$E[X] = \lim_{z \uparrow 1} \frac{dG_X}{dz}(z).$$

Similarly, the second moment of X can be derived as

$$E[X^2] = \lim_{z \uparrow 1} \left(\frac{d^2 G_X}{dz^2}(z) + \frac{dG_X}{dz}(z) \right).$$

This can be quite useful, as illustrated in the following example.

Example 69 (Poisson Random Variable). *Suppose that X has a Poisson distribution with parameter $\lambda > 0$. The function $G_X(s)$ can be computed using the distribution of X ,*

$$G_X(z) = \sum_{k=0}^{\infty} z^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}.$$

The first two moments of X are given by

$$\begin{aligned} E[X] &= \lim_{z \uparrow 1} \frac{dG_X}{dz}(z) = \lim_{z \uparrow 1} \lambda e^{\lambda(z-1)} = \lambda \\ E[X^2] &= \lim_{z \uparrow 1} \left(\frac{d^2 G_X}{dz^2}(z) + \frac{dG_X}{dz}(z) \right) = \lim_{z \uparrow 1} (\lambda^2 + \lambda) e^{\lambda(z-1)} = \lambda^2 + \lambda. \end{aligned}$$

This provides a very efficient way to compute the mean and variance of X , which are both equal to λ . It may be helpful to compare this derivation with Example 54.

7.6.1 The Sum of Two Independent Random Variables

Now, we revisit the distribution of the sum of two independent random variables. One interesting application of (7.6) occurs when dealing with sums of

independent random variables. Suppose X and Y are two independent random variables that take on integer values and, again, let $U = X + Y$. The ordinary generating function of U , as defined in Section 7.6, is given by

$$\begin{aligned} G_U(z) &= \mathbf{E} [z^U] = \mathbf{E} [z^{X+Y}] \\ &= \mathbf{E} [z^X z^Y] = \mathbf{E} [z^X] \mathbf{E} [z^Y] \\ &= G_X(z)G_Y(z). \end{aligned}$$

That is, the generating function of a sum of independent random variables is equal to the product of the individual ordinary generating functions.

Example 70 (Sum of Poisson Random Variables). *Let X be a Poisson random variable with parameter α and let Y be a Poisson random variable with parameter β . We wish to find the probability mass function of $U = X + Y$.*

To solve this problem, we use ordinary generating functions. First, recall that the generating function of a Poisson random variable with parameter λ is $e^{\lambda(z-1)}$. The ordinary generating functions of X and Y are therefore equal to

$$\begin{aligned} G_X(z) &= e^{\alpha(z-1)} \\ G_Y(z) &= e^{\beta(z-1)}. \end{aligned}$$

As such, the ordinary generating function of $U = X + Y$ is

$$G_U(z) = G_X(z)G_Y(z) = e^{\alpha(z-1)}e^{\beta(z-1)} = e^{(\alpha+\beta)(z-1)}.$$

We conclude, by the uniqueness of generating functions, that U is a Poisson random variable with parameter $\alpha + \beta$ and, accordingly, we get

$$p_U(k) = \frac{(\alpha + \beta)^k}{k!} e^{-(\alpha+\beta)}, \quad k = 0, 1, 2, \dots$$

This method of finding the PMF of U is more concise than using the discrete convolution.

7.7 Numerous Random Variables

The notion of a joint distribution can be applied to any number of random variables. Let X_1, X_2, \dots, X_n be random variables; their joint PMF is defined

by

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \Pr \left(\bigcap_{k=1}^n \{X_k = x_k\} \right),$$

or, in vector form, $p_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x})$. When the random variables $\{X_k\}$ are independent, their joint PMF reduces to

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{k=1}^n p_{X_k}(x_k).$$

Although not discussed in details herein, we emphasize that most concepts introduced in this chapter can be extended to multiple random variables in a straightforward manner. Also, we note that, from an abstract perspective, the space defined by vectors of the form $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$, $\omega \in \Omega$, is itself a sample space on which random variables can be defined.

Consider the empirical sums

$$S_n = \sum_{k=1}^n X_k \quad n \geq 1,$$

where X_1, X_2, \dots are independent integer-valued random variables, each with marginal PMF $p_X(\cdot)$. Obviously, the distribution of S_1 is simply $p_X(\cdot)$. More generally, the PMF of S_n can be obtained recursively using the formula

$$S_n = S_{n-1} + X_n.$$

This leads to the PMF

$$p_{S_n}(k) = (p_X * p_X * \dots * p_X)(k),$$

which is the n -fold convolution of $p_X(\cdot)$. Furthermore, the ordinary generating function of S_n can be written as

$$G_{S_n}(z) = (G_X(z))^n.$$

Example 71 (Binomial Random Variables). *Suppose that S_n is a sum of n independent and identically distributed Bernoulli random variables, each with parameter $p \in (0, 1)$. The PMF of S_1 is equal to*

$$p_{S_1}(k) = \begin{cases} 1-p & k=0 \\ p & k=1. \end{cases}$$

Assume that the PMF of S_{n-1} is given by

$$p_{S_{n-1}}(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}, \quad k = 0, 1, \dots, n-1.$$

Then, the distribution of S_n can be computed recursively using the discrete convolution,

$$\begin{aligned} p_{S_n}(k) &= \sum_{m=-\infty}^{\infty} p_{S_{n-1}}(m) p_X(k-m) \\ &= p_{S_{n-1}}(k)(1-p) + p_{S_{n-1}}(k-1)p \\ &= \binom{n-1}{k} p^k (1-p)^{n-k} + \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

where $k = 0, 1, \dots, n$. Thus, by the principle of mathematical induction, we gather that the sum of n independent Bernoulli random variables, each with parameter p , is a binomial random variable with parameters n and p .

The convolution of two binomial distributions, one with parameter m and p and the other with parameters n and p , is also a binomial random distribution with parameters $(m+n)$ and p . This fact follows directly from the previous argument.

Example 72 (Negative Binomial Random Variable*). Suppose that a Bernoulli trial is repeated multiple times until r ones are obtained. We denote by X the random variable that represents the number of zeros observed before completion of the process. The distribution of X is given by

$$p_X(k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

where p is the common parameter of the Bernoulli trials. In general, a random variable with such a distribution is known as a negative binomial random variable.

We wish to show that X can be obtained as a sum of r independent random variables, a task which we perform by looking at its ordinary generating function. We use properties of differential equations to derive $G_X(z)$ from $p_X(\cdot)$.

By definition, we have

$$G_X(z) = \sum_{k=0}^{\infty} z^k \binom{k+r-1}{r-1} p^r (1-p)^k.$$

Consider the derivative of $G_X(z)$,

$$\begin{aligned} \frac{dG_X}{dz}(z) &= \sum_{k=0}^{\infty} k z^{k-1} \binom{k+r-1}{r-1} p^r (1-p)^k \\ &= \sum_{k=1}^{\infty} z^{k-1} \frac{(k+r-1)!}{(r-1)!(k-1)!} p^r (1-p)^k \\ &= (1-p) \sum_{\ell=0}^{\infty} (\ell+r) z^{\ell} \binom{\ell+r-1}{r-1} p^r (1-p)^{\ell} \\ &= (1-p) \left(z \frac{dG_X}{dz}(z) + r G_X(z) \right). \end{aligned}$$

It follows from this equation that

$$\frac{1}{G_X(z)} \frac{dG_X}{dz}(z) = \frac{(1-p)r}{1-(1-p)z}$$

or, alternatively, we can write

$$\frac{d}{dz} \log(G_X(z)) = \frac{(1-p)r}{1-(1-p)z}.$$

Integrating both sides yields

$$\log(G_X(z)) = -r \log(1-(1-p)z) + c,$$

where c is an arbitrary constant. Applying boundary condition $G_X(1) = 1$, we get the ordinary generating function of X as

$$G_X(z) = \left(\frac{p}{1-(1-p)z} \right)^r.$$

From this equation, we can deduce that X is the sum of r independent random variables, Y_1, \dots, Y_r , each with ordinary generating function

$$G_{Y_i}(z) = \frac{p}{1-(1-p)z}.$$

In particular, the distribution of Y_i is given by

$$p_{Y_i}(m) = \frac{1}{m!} \frac{d^m G_{Y_i}}{dz^m}(0) = p(1-p)^m \quad m = 0, 1, \dots$$

It may be instructive to compare this distribution with the PMF of a geometric random variable.

Further Reading

1. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Sections 2.5–2.7.
2. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Sections 6.1–6.4.
3. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Sections 3.4–3.5.

