Chapter 4

Pairs of Random Variables

4.1 Multiple Random Variables

In the previous two chapters, we have seen one way to define multiple random variables on the same probability space $(\Omega, \mathcal{E}, \mathbb{P})$, by using a function $g(\cdot)$ to map a random variable $X(\omega)$ to a different random variable $Y(\omega) = g(X(\omega))$. However, it is natural in many experiments to generate more than one random variable for each outcome, and for the second random variable not to be derived from the value of the first random variable. Consider an experiment where one rolls two six-sided dice. One random variable, $X(\omega)$, is the value of the first die, and the other random variable, $Y(\omega)$, is the value of the second dice. In this case, notice that $Y(\omega)$ can have multiple values for each value of $X(\omega)$, which means that $Y(\omega)$ is not derived as a function of $X(\omega)$. In this experiment, we expect that the values that $Y(\omega)$ takes and $X(\omega)$ takes are not related, and appear uniformly in $\{1, 2, \dots, 6\}$. We recognize that this experiment was simply the combination of two independent experiments, and that perhaps we can treat X and Y as random variables from different experiments. Thus, it would be sufficient to know the individual probability mass functions $P_X(x), P_Y(y)$ to conduct further analyses.

However, consider an experiment of rolling two dice, but generating two random variables as follows: the first, $X(\omega)$, is the sum of the dice outcomes, and the second, $Y(\omega)$ is the product of the dice outcomes. Now, $X(\omega)$ takes values in $\{2, 3, \dots, 12\}$, and $Y(\omega)$ takes values in a very different discrete set. Furthermore, their values are related in unusual ways: if $X(\omega) = 2$, then $Y(\omega) = 1$. If $X(\omega) = 4$, then $Y(\omega) \in \{3, 4\}$. It is clear that the values of X, Y depend closely on the full outcome ω, and cannot be separated as two independent subexperiments. In essence, the random variables are now a two-dimensional function $\underline{g}(\omega) = (X(\omega), Y(\omega))$, with values in a discrete subset of $\Re^2$. The choice of function $\underline{g}(\cdot)$ defines the range $R_{X,Y}$ and will define a probability mass function in that range.

Note that both experiments use the same underlying probability space $(\Omega, \mathcal{E}, \mathbb{P})$, with the same outcomes Ω and the same discrete probability measure $\mathbb{P}$. However, we defined different random variables in the experiments. We could have generated more than two random variables for the same outcome. **Multiple random variables** are the result of a vector-valued function that assigns multiple real numbers to each outcome in the sample space. Intuitively, we can think of multiple random variables as the observations from an experiment that simultaneously produces two or more numbers for each outcome. The above discussion highlights that the relationship between multiple random variables is more general than what we saw in earlier chapters, where one random variable was derived from the other random variable by a function transformation.

In this chapter, we focus on generalizing the concepts we developed for scalar random variables in Chapter 2 and Chapter 3 to the experiments that generate two random variables $X(\omega), Y(\omega)$ for each outcome. In later chapters, we generalize this to experiments that generate random vectors of higher dimension for each outcome.

4.2 Pairs of Random Variables

Formally, a pair of random variables (X, Y) in a probability space $(\Omega, \mathcal{E}, \mathbb{P})$ consists of a vector-valued function from $\Omega \to \Re^2$. We also refer to such a pair (X, Y) of random variables as bivariate random variables, or joint random variables.

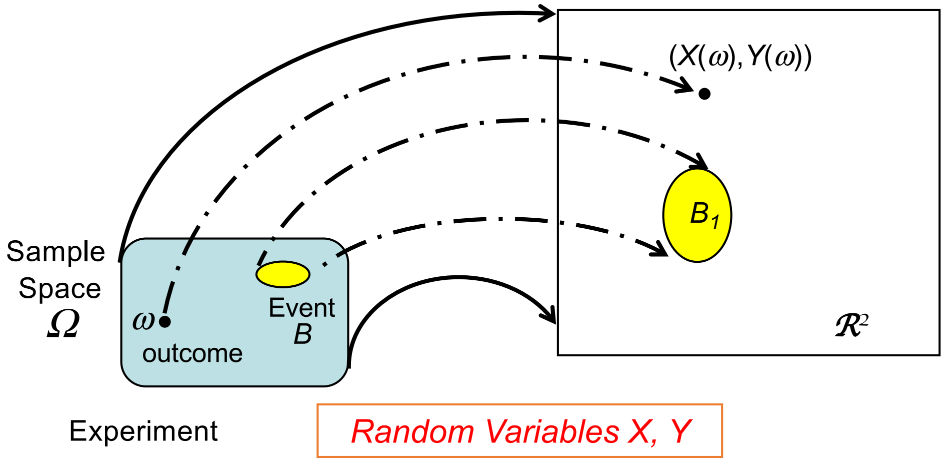


Figure 4.1: Bivariate random variables map single outcomes into two numerical values.

Figure 4.1 illustrates how pairs of random variables map individual outcomes $\omega \in \Omega$ into an ordered pair $(X(\omega), Y(\omega)) \in \Re^2$. We are interested in computing probabilities on the possible values of $X(\omega), Y(\omega)$, such as the probability that $(X(\omega), Y(\omega)) \in B_1 \subset \Re^2$ in Fig. 4.1. Thus, we restrict ourselves to functions where the *inverse image* of reasonable sets such as rectangular subsets of $\Re^2$ generate events $B \in \mathcal{E}$ for which $\mathbb{P}[B]$ is defined.[1] Then, we compute such probabilities as $\mathbb{P}[\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B_1\}] = \mathbb{P}[\{\omega \in B\}]$

For scalar random variables X, we defined the cumulative distribution function $F_X(x)$ as a function that summarized the probability of events defined in terms of intervals of values of X. For bivariate random variables X, Y, each random variable has its own CDF $F_X(x)$ and $F_Y(y)$, defined as in the previous chapters as $F_X(x) = \mathbb{P}[\{\omega \in \Omega : X(\omega) \le x\}], F_Y(y) = \mathbb{P}[\{\omega \in \Omega : Y(\omega) \le y\}]$. However, these CDF functions do not capture how the values of the random variables relate to each other.

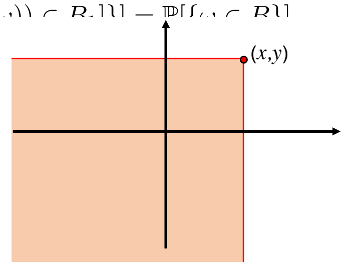


Figure 4.2: The CDF $F_{X,Y}(x,y)$ is the probability that the random variables take values in the shaded area .

To capture the probabilistic relationship between the two random variables, we define the **joint cumulative distribution function (CDF)** for values $(x, y) \in \Re^2$ as

$$F_{X,Y}(x,y) = \mathbb{P}[\{\omega \in \Omega : X(\omega) \le x,\ Y(\omega) \le y\}] = \mathbb{P}\big[\{\omega \in \Omega : X(\omega) \le x\} \cap \{\omega \in \Omega : Y(\omega) \le y\}\big].$$

That is, the joint CDF $F_{X,Y}(x,y)$ measures the probability of the event of outcomes where the random variables take values in the semi-infinite rectangle $(-\infty, x] \times (-\infty, y]$. This is illustrated in Figure 4.2. Note that this definition of CDF makes no distinction as to whether the joint random variables X, Y are discrete-valued or continuous-valued.

The joint CDF satisfies the following basic properties:

- **Non-negativity:** $0 \le F_{X,Y}(x,y)$.
- **Normalization:** $\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$.
- **Non-decreasing:** For any $x \le \tilde{x}$ and $y \le \tilde{y}$, $F_{X,Y}(x,y) \le F_{X,Y}(\tilde{x},\tilde{y})$.

[1] Formally, we define the Borel σ-field in $\Re^2$ as that generated from two-dimensional intervals by countable unions, intersections and complementation, and we require the function $\underline{g}(\omega) = (X(\omega), Y(\omega))$ to be measurable, so inverse images of Borel sets are events in $\mathcal{E}$.

- **Marginalization:** $\lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x)$ and $\lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y)$.
- $\lim_{x,\to-\infty} F_{X,Y}(x,y) = 0$ and $\lim_{y\to-\infty} F_{X,Y}(x,y) = 0$.
- $F_{X,Y}(x,y)$ is a right-continuous function of x for each y and a right-continuous function of y for each x. That is,

$$\lim_{\epsilon>0,\epsilon\to 0} F_{X,Y}(x+\epsilon,y) = F_{X,Y}(x,y); \quad \lim_{\epsilon>0,\epsilon\to 0} F_{X,Y}(x,y+\epsilon) = F_{X,Y}(x,y).$$

Note that the marginal CDFs $F_X(x), F_X(y)$ can be derived from the joint CDF. The converse, however, is not true, as is clear from the dice discussion earlier.

Using the joint CDF, we can perform computations of the probabilities that the bivariate random variables take values in certain intervals, as illustrated in the following example.

Example 4.1
Compute the following probabilities using the joint CDF:

(a) $\mathbb{P}[\{X > x\} \cup \{Y > y\}]$ (b) $\mathbb{P}[\{X \le x\} \cup \{Y \le y\}]$
(c) $\mathbb{P}[\{X \le x\} \cup \{Y > y\}]$ (d) $\mathbb{P}[\{\omega \in \Omega : X(\omega) \in (x,x'], Y(x) \in (y,y']\}$

Answer: Figure 4.3 shows the areas in $\Re^2$ for the questions, with some ambiguity as to whether the red boundaries are part of the region of interest. Notice the specific choice of the questions, to determine the type of interval required, as that determines whether the boundary is included.

For (a), we see the answer is the complement of the joint CDF, as $\mathbb{P}[\{X > x\} \cup \{Y > y\}] = 1 - F_{X,Y}(x,y)$.

For (b), the answer is a little more complex:

$$\begin{aligned}\mathbb{P}[\{X \le x\} \cup \{Y \le y\}] &= \mathbb{P}[\{X \le x\}] + \mathbb{P}[\{Y \le y\}] - \mathbb{P}[[\{X \le x\} \cap \{Y \le y\}] \\ &= F_{X,Y}(x,\infty) + F_{X,Y}(\infty,y) - F_{X,Y}(x,y) = F_X(x) + F_Y(y) - F_{X,Y}(x,y)\end{aligned}$$

For (c), we see that $\{X \le x, y, Y \le y\} \cap \{Y > y\} = \emptyset$, so

$$\mathbb{P}[\{X \le x\} \cup \{Y > y\}] = F_{X,Y}(x,y) + (1 - F_{X,Y}(\infty,y)) = F_{X,Y}(x,y) + (1 - F_Y(y))$$

For (d), we have

$$\mathbb{P}[\{\omega \in \Omega : X(\omega) \in (x,x'], Y(x) \in (y,y']\} = F_{X,Y}(x',y') - F_{X,Y}(x,y') - F_{X,Y}(x',y) + F_{X,Y}(x,y)$$

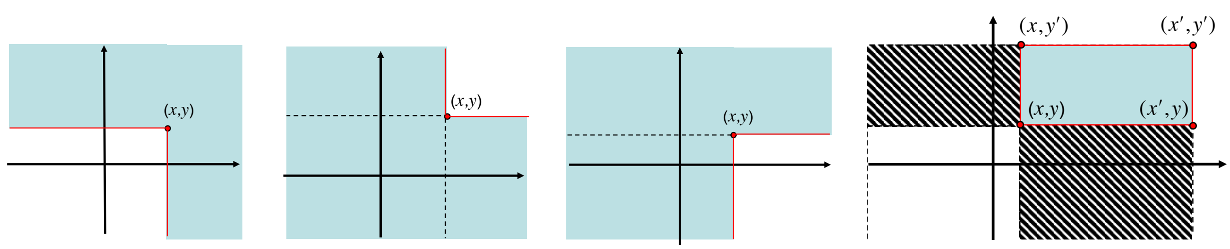


Figure 4.3: Regions of interest for the questions in 4.1.

4.3 Pairs of Discrete Random Variables

4.3.1 Joint Probability Mass Function

A pair of random variables X, Y is discrete if X and Y are discrete random variables. For discrete bivariate random variables, we define the joint probability mass function $P_{X,Y}(x,y)$ as follows:

Definition 4.1
The **joint probability mass function (PMF)** of a pair of discrete random variables X and Y is

$$P_{X,Y}(x,y) = \mathbb{P}[\{\omega \in \Omega : X(\omega) = x,\ Y(\omega) = y\}] = \mathbb{P}\big[\{X = x\} \cap \{Y = y\}\big].$$

The joint PMF is zero except at a discrete number of points in $\Re^2$, each of which has positive probability mass. The **range** $R_{X,Y}$ of a pair of discrete random variables is the set of all possible pairs of values,

$$R_{X,Y} = \{(x,y) : P_{X,Y}(x,y) > 0\}.$$

The joint PMF satisfies the following properties:

- **Non-negativity:** $P_{X,Y}(x,y) \geq 0$.
- **Normalization:** $\displaystyle\sum_{(x,y)\in R_{X,Y}} P_{X,Y}(x,y) = 1.$
- **Probability of an event:** Suppose we have a set $B \subset R_{x,y}$. Then,

$$\mathbb{P}[\{(x,y) \in B\}] = \mathbb{P}\big[\{(X,Y) \in B\}\big] = \sum_{(x,y)\in B} P_{X,Y}(x,y).$$

When the range sets R_X, R_Y of the two random variables are finite, we can visualize the joint PMF as an array of probability masses. Let $R_X = \{x_1, x_2, \ldots, x_n\}, R_Y = \{y_1, y_2, \ldots, y_m\}$, as illustrated in Table 4.1. Note that some of the numbers in the array can be zero, as the joint range $R_{X,Y}$ is often not equal to the cross product $R_X \times R_Y$.

$Y \backslash X$	x_1	x_2	$\cdots$	x_{n-1}	x_n
y_1	$P_{X,Y}(x_1,y_1)$	$P_{X,Y}(x_2,y_1)$	$\cdots$	$P_{X,Y}(x_{n-1},y_1)$	$P_{X,Y}(x_n,y_1)$
y_2	$P_{X,Y}(x_1,y_2)$	$P_{X,Y}(x_2,y_2)$	$\cdots$	$P_{X,Y}(x_{n-1},y_2)$	$P_{X,Y}(x_n,y_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{m-1}	$P_{X,Y}(x_1,y_{m-1})$	$P_{X,Y}(x_2,y_{m-1})$	$\cdots$	$P_{X,Y}(x_{n-1},y_{m-1})$	$P_{X,Y}(x_n,y_{m-1})$
y_m	$P_{X,Y}(x_1,y_m)$	$P_{X,Y}(x_2,y_m)$	$\cdots$	$P_{X,Y}(x_{n-1},y_m)$	$P_{X,Y}(x_n,y_m)$

Table 4.1: Visualization of joint PMF as a table of probability masses

From the joint PMF, we can obtain marginal PMFs for each random variable X, Y as follows: The **marginal PMF** $P_X(x)$ is just the PMF of X, and can be obtained from the joint PMF $P_{X,Y}(x,y)$ as:

$$P_X(x) = \sum_{y\in R_Y} P_{X,Y}(x,y).$$

Note we sum over all the possible values of the variable that we are trying to eliminate, Y. Similarly, the the marginal PMF $P_Y(y)$ is just the PMF of Y, obtained from the joint PMF as

$$P_Y(y) = \sum_{x\in R_X} P_{X,Y}(x,y).$$

In terms of the array representation in Table 4.1, $P_X(x)$ is obtained by summing the elements of the column corresponding to $X = x$, and $P_Y(y)$ is obtained by summing the elements of the row corresponding to $Y = y$.

Example 4.2
Given the details of an experiment, we can compute the joint PMF of a pair of discrete random variables X, Y by using the underlying probability measure $\mathbb{P}$ on the probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Specifically,

$$P_{X,Y}(x,y) = \mathbb{P}[\{\omega \in \Omega : X(\omega) = x,\ Y(\omega) = y\}].$$

We illustrate this below.

Let the experiment consist of rolling two six-sided dice: an outcome ω is an ordered pair of numbers (a, b), with $a, b \in \{1, 2, \dots, 6\}$. Each outcome in Ω is equally likely. We define the discrete random variables X, Y as follows:

$$X(\omega) = \begin{cases} 1 & \text{if the sum of dice rolls is odd,} \\ 0 & \text{otherwise;} \end{cases} \qquad Y(\omega) = \begin{cases} 1 & \text{if the product of dice rolls is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

The range of X, Y is $R_{X,Y} = \{(0,0), (0,1), (1,0), (1,1)\}$. To compute the PMF, for each value $(x, y) \in R_{X,Y}$, we compute the set of outcomes $\omega \in \Omega$ that map to that value, and compute the probability of that set. For instance,

$$\{\omega \in \Omega : X(\omega) = 0, Y(\omega) = 0\} = \{(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)\}.$$

Thus,

$$P_{X,Y}(0,0) = \mathbb{P}[\{\omega \in \Omega : X(\omega) = 0, Y(\omega) = 0\}] = \frac{9}{36} = \frac{1}{4}.$$

What about $P_{X,Y}(1,1)$? The set $\{\omega \in \Omega : X(\omega) = 1, Y(\omega) = 1\} = \emptyset$, because no pair of dice outcome can have an odd sum and an odd product! Thus, $P_{X,Y}(1,1) = 0$.

To complete the PMF, note that

$$\{\omega \in \Omega : X(\omega) = 0, Y(\omega) = 1\} = \{(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5)\},$$

so $P_{X,Y}(0,1) = \frac{1}{4}$ also. Hence, by normalization, we must have $P_{X,Y}(1,0) = \frac{1}{2}$. We can verify this, as $\{\omega \in \Omega : X(\omega) = 1, Y(\omega) = 0\}$ will have the remaining 18 outcomes.

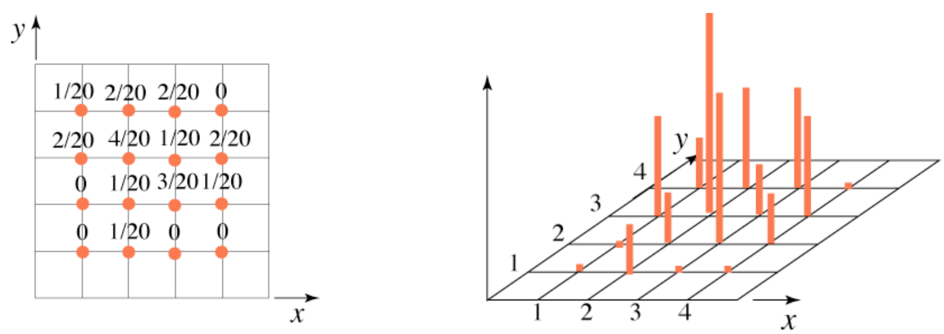


Figure 4.4: Figure for example 4.3.

Example 4.3
Consider a pair of random variables X, Y with joint PMF as illustrated in Figure 4.4, where the array representation of the joint PMF is shown on the left. Compute the probability that (X, Y) take values in the set $B = \{(x, y) : x \in [1, 2], y \in [2, 3]\}$. Also, compute the marginal PMF functions $P_X(x)$ and $P_X(y)$.

To compute $\mathbb{P}[\{(x, y) \in B\}]$, we use the joint PMF and add the probability over the masses at the points in B:

$$\mathbb{P}[\{(x,y) \in B\}] = P_{X,Y}(1,2) + P_{X,Y}(1,3) + P_{X,Y}(2,2) + P_{X,Y}(2,3) = 0 + \frac{2}{20} + \frac{1}{20} + \frac{4}{20} = \frac{7}{20}.$$

For the marginal PMFs, we first compute the PMF of X:

$$P_X(1) = \sum_{y \in R_Y} P_{X,Y}(1,y) = P_{X,Y}(1,3) + P_{X,Y}(1,4) = \frac{2}{20} + \frac{1}{20} = \frac{3}{20}$$

$$P_X(2) = \sum_{y \in R_Y} P_{X,Y}(2,y) = \frac{1}{20} + \frac{1}{20} + \frac{4}{20} + \frac{2}{20} = \frac{8}{20}$$

$$P_X(3) = \sum_{y \in R_Y} P_{X,Y}(3,y) = \frac{3}{20} + \frac{1}{20} + \frac{2}{20} = \frac{6}{20}$$

$$P_X(4) = \sum_{y \in R_Y} P_{X,Y}(4,y) = \frac{1}{20} + \frac{2}{20} = \frac{3}{20}$$

For the marginal PMF of Y, we compute:

$$P_Y(1) = \sum_{x \in R_X} P_{X,Y}(x,1) = P_{X,Y}(2,1) = \frac{1}{20}$$
$$P_Y(2) = \sum_{x \in R_X} P_{X,Y}(x,2) = P_{X,Y}(2,2) + P_{X,Y}(3,2) + P_{X,Y}(4,2) = \frac{5}{20}$$
$$P_Y(3) = \sum_{x \in R_X} P_{X,Y}(x,3) = P_{X,Y}(1,3) + P_{X,Y}(2,3) + P_{X,Y}(3,3) + P_{X,Y}(4,3) = \frac{9}{20}$$
$$P_Y(4) = \sum_{x \in R_X} P_{X,Y}(x,4) = P_{X,Y}(1,4) + P_{X,Y}(2,4) + P_{X,Y}(3,4) = \frac{5}{20}$$

4.3.2 Conditional PMF

For discrete random variables X in a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, we defined the conditional probability mass function of X given observation of an event $B \in \mathcal{E}$ as

$$P_{X|B}(x) = \begin{cases} \dfrac{\mathbb{P}[\{s : X(\omega) = x\} \cap B]}{\mathbb{P}[B]}, & \mathbb{P}[B] > 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

When we have a pair of random variables X, Y, the set B can be defined in terms of the random variable Y as $B = \{\omega \in \Omega : Y(\omega) = y\}$, which we write as an abbreviated $\{Y = y\}$. For this case, the following relationships hold:

$$\mathbb{P}[\{X = x\} \cap B] = \mathbb{P}[\{X = x\} \cap \{Y = y\}] = P_{X,Y}(x,y); \quad \mathbb{P}[B] = \mathbb{P}[\{Y = y\}] = P_Y(y)$$

Thus, we define the **conditional PMF** that $X = x$ given that $Y = y$ is observed as

$$P_{X|Y}(x|y) = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P_Y(y)}, & P_Y(y) > 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Note the following: if the numerator $P_{X,Y}(x,y) > 0$ for some x, y, then $P_Y(y) > 0$ and $P_X(x) > 0$ as obtained by marginalization. Hence, the reason for the *undefined* clause in the above equation is to handle the case when both numerator and denominator in the ratio are zero, in which case we don't define that conditional probability.

Similarly, we define the conditional PMF that $Y = y$ given that $X = x$ is observed as

$$P_{Y|X}(y|x) = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P_X(x)}, & P_X(x) > 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

In essence, the conditional probability mass function is the ratio of the joint PMF to the marginal PMF of the variable being observed. When both $P_X(x) > 0, P_Y(y) > 0$, we can also represent the joint CMF as the product of the conditional CMF and the marginal CMF, as

$$P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y) = P_{Y|X}(y|x)P(x).$$

We refer to this property as the **Multiplication Rule.**

The conditional PMF $P_{X|Y}(x|y)$ is a valid probability mass function on R_X, and thus satisfies the following basic properties of probability mass functions:

- **Non-negativity:** $P_{X|Y}(x|y) \geq 0$ and $P_{Y|X}(y|x) \geq 0$ for all $x \in R_X$, $y \in R_Y$.

- **Normalization:** $\sum_{x \in R_X} P_{X|Y}(x|y) = 1$ for any $y \in R_Y$ and $\sum_{y \in R_Y} P_{Y|X}(y|x) = 1$ for any $x \in R_X$.

- **Additivity:** For any event $B \subset R_X$, the probability that X falls in B given $Y = y$ is

$$\mathbb{P}[B|\{Y = y\}] = \sum_{x \in B} P_{X|Y}(x|y) \text{ for } y \in R_Y.$$

For any event $B \subset R_Y$, the probability that Y falls in B given $X = x$ is

$$\mathbb{P}[B|\{X = x\}] = \sum_{y \in B} P_{Y|X}(y|x) \text{ for } x \in R_X.$$

Example 4.4
Consider two random variables X, Y with the joint PMF function used in the previous example 4.3, illustrated in Figure 4.4. Compute $P_{X|Y}(x|y)$ for $y = 2, y = 3$. The table below has the joint PMF of X, Y for this example. To compute

$Y \backslash X$	1	2	3	4
1	0	$\frac{1}{20}$	0	0
2	0	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	0

Table 4.2: Visualization of joint PMF as a table of probability masses

$P_{X|Y}(x|3)$, we **restrict** the the value of Y to the row $Y = 3$. We sum the probability masses in that row to get $P_Y(3) = \frac{9}{20}$. We use these to rescale the values in that row to get:

$$P_{X|Y}(1|3) = \frac{P_{X,Y}(1,3)}{P_Y(3)} = \frac{\frac{2}{20}}{\frac{9}{20}} = \frac{2}{9}$$
$$P_{X|Y}(2|3) = \frac{P_{X,Y}(2,3)}{P_Y(3)} = \frac{\frac{4}{20}}{\frac{9}{20}} = \frac{4}{9}$$
$$P_{X|Y}(3|3) = \frac{P_{X,Y}(3,3)}{P_Y(3)} = \frac{\frac{1}{20}}{\frac{9}{20}} = \frac{1}{9}$$
$$P_{X|Y}(4|3) = \frac{P_{X,Y}(4,3)}{P_Y(3)} = \frac{\frac{2}{20}}{\frac{9}{20}} = \frac{2}{9}$$

Notice that $P_{X|Y}(x|3)$ is proportional to the row $P_{X,Y}(x,3)$, **rescaled** by dividing by $P_Y(3)$ so that $\sum_{x \in R_X} P_{X|Y}(x|3) = 1$.

Similarly, $P_{X|Y}(x|2)$ is computed as follows: $P_Y(2) = \sum_{x \in R_X} P_{X,Y}(x,2) = \frac{5}{20}$. Then,

$$P_{X|Y}(1|2) = \frac{P_{X,Y}(1,2)}{P_Y(2)} = \frac{0}{\frac{5}{20}} = 0$$
$$P_{X|Y}(2|2) = \frac{P_{X,Y}(2,2)}{P_Y(2)} = \frac{\frac{1}{20}}{\frac{5}{20}} = \frac{1}{5}$$
$$P_{X|Y}(3|2) = \frac{P_{X,Y}(3,2)}{P_Y(2)} = \frac{\frac{3}{20}}{\frac{5}{20}} = \frac{3}{5}$$
$$P_{X|Y}(4|2) = \frac{P_{X,Y}(4,2)}{P_Y(2)} = \frac{\frac{1}{20}}{\frac{5}{20}} = \frac{1}{5}$$

The techniques we developed for conditional probabilities can be extended for conditional PMF functions, as follows. Let $R_Y = \{y_1, y_2, \ldots\}$ denote the discrete range of the random variable Y. Then, the events $\{\omega \in \Omega : Y(\omega) = y_1\}, \{\omega \in \Omega : Y(\omega) = y_2\}, \ldots$ are mutually disjoint if $y_1 \neq y_2$, because Y is a function, so there is only one value of y associated with an outcome $\omega \in \Omega$. Furthermore, they are collectively exhaustive, because every $\omega \in \Omega$ must be mapped to some $y \in R_Y$. Thus, we can derive a version of the Law of Total Probability for pairs of discrete random variables X, Y , which is:

Law of Total Probability:

$$P_X(x) = \sum_{y \in R_Y} P_{X|Y}(x|y) P_Y(y)$$
$$P_Y(y) = \sum_{x \in R_X} P_{Y|X}(y|x) P_X(x) \; .$$

We can also develop a version of Bayes' Rule for pairs of discrete random variables, as:

Bayes' Rule:

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)} = \frac{P_{Y|X}(y|x)P_X(x)}{\sum_{x' \in R_X} P_{Y|X}(y|x')P_X(x')}$$

Example 4.5

Consider an X-ray source that generates photons with a specified rate λ photons per unit time. The emitted photons go through a mask that absorbs each photon with probability p, independently for each photon. For instance, in computed tomography machines, X-ray sources are typically modulated with masks to attenuate low-energy rays during X-ray imaging, as they contribute little to the quality of the image and get absorbed in body tissues.

Assume we operate the X-ray source for a single unit of time. The number of photons emitted is represented as a Poisson random variable with parameter λ, denoted by N. That is,

$$\mathbb{P}[\{N = n\}] \equiv P_N(n) = \frac{\lambda^n}{n!} e^{-\lambda}, \; n = 0, 1, 2, \ldots$$

We are interested in the number of photons that make it through the mask. That is a second random variable X. Note that if we know that $N = n$, then we can characterize the type of random variable that X is: There are n independent trials for photons to go through, and the success rate of each trial is $(1-p)$. Thus, conditioned on $N = n$, X is a binomial random variable with parameters n, $(1-p)$. That is,

$$\mathbb{P}[\{X = k\}|\{N = n\}] \equiv P_{X|N}(k|n) = \binom{n}{k} (1-p)^k p^{n-k}, \; k = 1, 2, \ldots, n.$$

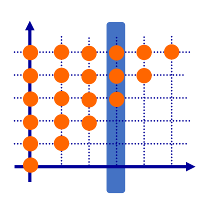
With these ideas, we can define the joint PMF of N, X as the product of a conditional probability and a marginal probability, as

$$P_{N,X}(n,x) = P_{X|N}(x|n) P_N(n) = \binom{n}{x} (1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda}.$$

The range of values for N, X require that $X \leq N$, so it is

$$R_{N,X} = \big\{(n,x) : n \in \{0, 1, 2, \ldots\}, x \in \{0, 1, \ldots, n\}\big\}$$

We can now perform computations that would be of interest, such as finding the marginal probability of X, the number of photons that make it through the mask. We get the marginal probability of X from the joint probability of N, X by marginalization over the values of N. Notice that, for a particular value of $X = x$, we have $P_{N,X}(n,x) = 0, n < x$, as illustrated in the figure on the right.



Thus, the marginal probability of X is computed as:

$$\begin{aligned}
P_X(x) &= \sum_{n=0}^{\infty} P_{N,X}(n,x) = \sum_{n=x}^{\infty} \binom{n}{x} (1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} \text{ where the lower sum limit is } x \text{ because of the range } R_{N,X} \\
&= \sum_{n=x}^{\infty} \frac{n!}{x!(n-x)!} (1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=x}^{\infty} \frac{\lambda^n}{x!(n-x)!} (1-p)^x p^{n-x} e^{-\lambda} \text{ (cancel the } n! \text{ terms)} \\
&= \frac{\big(\lambda(1-p)\big)^x}{x!} e^{-\lambda} \sum_{n'=0}^{\infty} \frac{(\lambda p)^{n'}}{(n')!} \text{ (substitute } n-x=n'\text{)} \\
&= \frac{\big(\lambda(1-p)\big)^x}{x!} e^{-\lambda} e^{\lambda p} = \frac{\big(\lambda(1-p)\big)^x}{x!} e^{-\lambda(1-p)} \text{ (recognizing the sum is an exponential.)}
\end{aligned}$$

Remarkably, we have just proven that the number of photons that make it through the mask is also a Poisson random variable, with parameter $\lambda(1-p)$, which is the product of the incoming photon intensity times the probability that each photon makes it through. Thus, we know that the expected number of photons that make it through the mask is $\lambda(1-p)$, and the variance of the number of photons that make it through the mask is also $\lambda(1-p)$.

The above result can be stated generally as: A Poisson random variable with intensity λ that undergoes independent sampling for each instance with probability p remains a Poisson random variable with a reduced intensity $p(1-\lambda)$. This result has many applications in engineering: For instance, consider a fork in a traffic road, where cars randomly choose with probability p to take the left fork and with probability $(1-p)$ to take the right fork. If the number of arrivals to the fork is modeled as a Poisson random variable with intensity λ, the number of departures on the left fork will be a Poisson random variable with intensity λp. Similarly, the number of departures on the right fork will be a Poisson random variable with intensity $\lambda(1-p)$.

Many sensor systems that count particles using physical mechanisms are modeled similarly. For instance, Geiger counters for radiation detection interact with α-particles, and detect each particle with a given probability. X-ray detector panels use scintillating materials that interact with incoming X-ray photons, and convert each photon to electrons with a given probability. If the arrival of particles is modeled as a Poisson random variable, the measured counts in these systems will also be Poisson random variables, albeit with reduced intensity.

Example 4.6
Consider the model of the previous example 4.5 for the pair of random variables N, X. Assume we observe that $X = 5$. What is the conditional probability distribution of N, given the information that $X = 5$?

To solve this, we apply Bayes' Rule for discrete random variables, as $P_{N|X}(n|x) = \dfrac{P_{N,X}(n,x)}{P_X(x)}$.

Fortunately, we have expressions for all of these from the previous problem:

$$P_{N,X}(n,x) = \binom{n}{x} (1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda}; \; P_X(x) = \frac{\big(\lambda(1-p)\big)^x}{x!} e^{-\lambda(1-p)} .$$

Hence, for observing $X = x$,

$$P_{N|X}(n|x) = \frac{P_{N,X}(n,x)}{P_X(x)} = \frac{\binom{n}{x}(1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda}}{\frac{(\lambda(1-p))^x}{x!} e^{-\lambda(1-p)}} = \frac{I_{\{n\geq x\}} \frac{n!}{x!(n-x)!}(1-p)^x p^{n-x} \frac{\lambda^n}{n!} e^{-\lambda}}{\frac{(\lambda(1-p))^x}{x!} e^{-\lambda(1-p)}}$$

where the indicator function $I_{\{n\geq x\}}$ is 1 if the condition is true ($n \geq x$), and zero otherwise. Canceling the appropriate factors in the numerator and denominator, we get

$$P_{N|X}(n|x) = \frac{I_{\{n\geq x\}} \frac{\lambda^n}{(n-x)!}(1-p)^x p^{n-x} e^{-\lambda}}{\big(\lambda(1-p)\big)^x e^{-\lambda(1-p)}} = I_{\{n\geq x\}} \frac{\lambda^{n-x}}{(n-x)!} p^{n-x} e^{-\lambda p}$$

Substituting $x = 5$ above gives the desired conditional PMF for N.

We can recognize what type of conditional distribution is $P_{N|X}(n|x)$ by defining a derived random variable, conditioned on knowing $X = x$, as $N' = N - x$. Note that

$$P_{N|X}(n|x) = I_{\{n\geq x\}} \frac{(\lambda p)^{n-x}}{(n-x)!} e^{-\lambda p} = I_{\{n'\geq 0\}} \frac{(\lambda p)^{n'}}{(n')!} e^{-\lambda p} = P_{N'|X}(n'|x).$$

Thus, conditioned on $X = x$, N has the PMF of the sum of x and a Poisson random variable with intensity λp. Notice that the gap from $X = x$ to $N = n$ corresponds to absorbed photons, and the probability of absorption for each photon is p. As discussed in example 4.5, the number of absorbed photons is a Poisson random variable with intensity λp. Using this reasoning, we could have obtained the above answer directly with no computation.

4.4 Pairs of Continuous Random Variables

For discrete scalar random variables X, the PMF $P_X(x)$ described how probability mass accumulated in discrete points in the real line $\Re$. In contrast, a continuous random variable X spreads its probability over the real line $\Re$ so that there is no probability mass at any point, but instead we have a probability density function (PDF) $f_X(x)$, measured in probability per unit length, that describes how probability is accumulated. Indeed, for a random variable X to be continuous, its cumulative distribution function (CDF) $F_X(x)$ must be differentiable almost everywhere, and

$$f_X(x) = \begin{cases} \frac{d}{dx} F_X(x) & \text{if } F_X(x) \text{ is differentiable at } x, \\ \text{arbitrary} & \text{elsewhere.} \end{cases}$$

We want to extend these concepts to bivariate random variables (X, Y) defined on a common probability space $(\Omega, \mathcal{E}, \mathbb{P})$. In the previous section, we saw how we defined discrete bivariate random variables, and characterized their properties using the joint PMF function $P_{X,Y}(x, y)$. We define the concept of **jointly continuous** bivariate random variables as follows.

Definition 4.2
A pair of random variables X, Y are said to be **jointly continuous** if their joint CDF $F_{X,Y}(x, y)$ is continuous, and differentiable almost everywhere, so that there exists a joint probability density function $f_{X,Y}(x, y)$ such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y') d\, x' d\, y' \ .$$

An implication of this definition is that there is no region $B \subset \Re^2$ where the area of B is zero, and the probability $\mathbb{P}[\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}] > 0$. Thus, there are no points which have positive probability masses, and there are no lines or curves with zero area that have positive probability of occurring.

4.4.1 Joint Probability Density Function

From the above definition, the joint probability density function (PDF) $f_{X,Y}(x, y)$ is computed as

$$f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2}{\partial x\ \partial y} F_{X,Y}(x, y) \partial x \partial y & \text{if } F_{X,Y}(x, y) \text{ is differentiable at } (x, y), \\ \text{arbitrary} & \text{otherwise.} \end{cases}$$

The **range** $R_{X,Y}$ of a pair of continuous random variables is the set of all possible pairs of values,

$$R_{X,Y} = \big\{(x, y) : f_{X,Y}(x, y) > 0\big\}.$$

The joint PDF has some structural properties that we highlight below:

- $f_{X,Y}(x, y) \geq 0$ for all $(x, y) \in \Re^2$. This follows from the fact that the joint CDF $F_{X,Y}(x, y)$ is monotone non-decreasing, and thus has non-negative derivatives.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x', y') d\, x' d\, y' = F_{X,Y}(\infty, \infty) = 1$. That is, the total volume between the surface map of $f_{X,Y}(x, y)$ and the x-y plane must be equal to 1.

Being a density and not a probability, the joint PDF can take positive values greater than 1.

Figure 4.5 illustrates the joint PDF of a pair of jointly continuous random variables X, Y. The top figure shows a small area ΔA around a point (x, y). the probability that (X, Y) take on values in ΔA is approximately computed as

$$\mathbb{P}[\{(X, Y) \in \Delta A\}] \approx f_{X,Y}(x, y)|\Delta A|,$$

where $|\Delta A|$ is the area of the region ΔA. This is approximately the volume of a column over ΔA, with height $f_{X,Y}(x, y)$, that is, the volume under the $f_{X,Y}$ graph that is above the area ΔA.
For any subset $A \in \Re^2$ with positive area, we compute the probability that (X, Y) take values in A using the joint PDF as

$$\mathbb{P}[\{(X, Y) \in A\}] = \iint_A f_{X,Y}(x, y) d\, x d\, y.$$

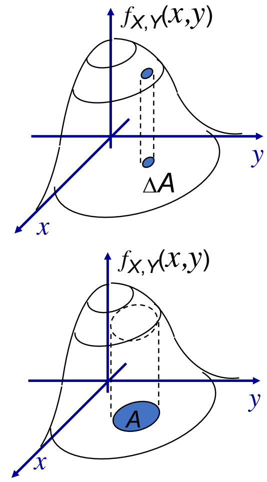


Figure 4.5: Illustration of joint PDF used for computation of probabilities.

Example 4.7
Consider a continuous random variable X defined on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. For a continuous function $g : \Re \to \Re$, we define the random variable $Y = g(X)$. Is the pair (X, Y) a jointly continuous pair of random variables?

The answer is no. To make this discussion simpler, let $g(x) = x$, and define the region $B = \{(x, y) \in \Re : x = y\}$. Note that this region is a line in the x-y plane, and has no area: $|B| = 0$. However, it is clear that $\mathbb{P}[\{(X, Y) \in B\}] = 1$, so that there is probability mass for a set of zero area. Hence, the pair of random variables is not jointly continuous.

You can extend this argument for any continuous function $g(\cdot)$. Basically, the set $B = \{(x, y) \in \Re : x = y\}$ represents a continuous line in the x-y plane which has zero area, and the probability that (X, Y) take values in B is one. This argument can also be extended to discontinuous functions $g(\cdot)$.

Example 4.8
Consider a pair of random variables (X, Y) with joint PDF given by

$$f_{X,Y}(x, y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the plot of this joint PDF is a cube of height 1 over the rectangle of area 1, and hence this join PDF satisfies the properties highlighted above: It is nonegative, and it integrates to 1, as the volume under the graph is 1.

What is the joint CDF of (X, Y)? By definition, this is $F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{\infty}^{y} f_{X,Y}(x, y) d\, x d\, y$.

Note that this integral is zero as long as either $x \le 0$ or $y \le 0$, as the integral takes place over a region where $f_{X,Y}(x, y) = 0$. Furthermore, if $x > 1$ and $y > 1$, then $F_{X,Y}(x, y) = 1$, because we integrate over the entire region where $f_{X,Y}(x, y) > 0$, namely the range $R_{X,Y}$. Elsewhere, we integrate to compute $F_{X,Y}(x, y)$. To make this easier, lets rewrite the joint PDF of X, Y using indicator functions, as: $f_{X,Y}(x, y) = 1 I_{\{x \in [0,1]\}} I_{\{y \in [0,1]\}}$. Then, for $x > 0, y > 0$, we have

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} I_{\{x \in [0,1]\}} I_{\{y \in [0,1]\}} d\, x d\, y \\ &= (\int_{0}^{x} I_{\{x \in [0,1]\}} d\, x)(\int_{0}^{y} I_{\{y \in [0,1]\}} d\, y) = \min(x, 1) \min(y, 1). \end{aligned}$$

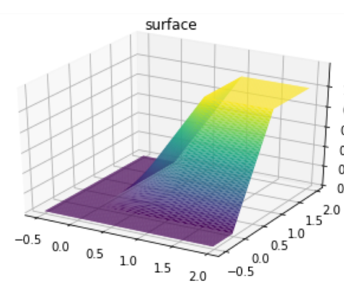


Figure 4.6: Joint CDF for Example 4.8.

Figure 4.6 shows a plot of the resulting CDF.

Putting all the equations together yields

$$F_{X,Y}(x,y) = \begin{cases} \min(x,1)\min(y,1), & x \geq 0, y \geq 0 \\ -0 & \text{otherwise.} \end{cases}$$

Example 4.9
Consider a pair of random variables (X, Y) with joint PDF given by

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \leq x \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute the following probabilities: What is the probability that $X + Y > 1$?. What is the probability that $(X - 0.5)^2 + (Y - 0.5)^2 < 0.25$?

A diagram is helpful to identify the sets involved in the answering the questions: First, the range $R_{X,Y}$ is a triangle in the plane, with corners (0,0), (1,1), (0,1). This is illustrated in Figure 4.7, which shows a plot of the joint PDF of (X, Y) with the range $R_{X,Y}$ outlined in the x-y plane. The intersection of the region $\{(x, y) : x + y \geq 1\}$ is highlighted in orange. Let $A = \{(x, y) : x + y \geq 1, 0 \leq x \leq y \leq 1\}$ denote that intersection region. The probability that $X + Y > 1$ is the probability that (X, Y) take values in A, which is computed from the join PDF as $\mathbb{P}[\{(X, Y) \in A\}] = \iint_A f_{X,Y}(x, y)d\ xd\ y$.

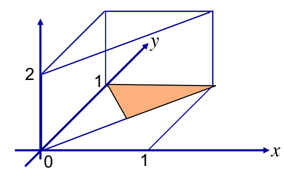


Figure 4.7: Joint PDF for Example 4.9.

Fortunately, the joint probability density function is constant $(= 2)$ in the region A, so we can compute the integral using simple geometric ideas: The volume of the region between the graph of the joint PDF and the area A is just the height times the area of the triangular base. The height is 2, and the triangle is seen to have a base of 1, height 0.5 so its area is 0.25. Hence, $\mathbb{P}[\{(X, Y) \in A\}] = 2 \times 0.25 = 0.5$.

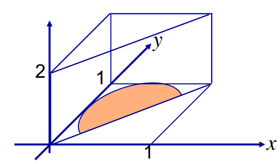


Figure 4.8: Joint PDF for Example 4.9.

Similarly, Let $B = \{(x, y) : (x - 0.5)^2 + (y - 0.5)^2 < 0.25\} \cap R_{X,Y}$. This area is highlighted in Figure 4.8. Then the probability that $(X - 0.5)^2 + (Y - 0.5)^2 < 0.25$ is the probability that (X, Y) takes values in B, which is $\mathbb{P}[\{(X, Y) \in B\}] = \iint_B f_{X,Y}(x, y)d\ xd\ y$. Reasoning as above, this is 2 times the area of B, which can be recognized from Figure 4.8 to be a half circle with radius 0.5. Hence,

$$\mathbb{P}[\{(X, Y) \in B\}] = 2 \times (0.5\pi(0.5)^2) = \frac{\pi}{4}\ .$$

4.4.2 Marginal PDF

If X, Y are jointly continuous random variables, then X and Y are continuous random variables individually, and have probability density functions $f_X(x)$ and $f_Y(y)$, called the marginal probability density functions. These can be computed from the joint CDF of (X, Y), by computing the marginal CDFs of X, Y and differentiating to obtain the marginal pdfs:

$$F_X(x) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(x', y)d\ ydx'$$

$$f_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$F_Y(y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(x, y')d\ xdy'$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x', y)d\ x'\ .$$

Alternatively, we obtain the marginal PDF of X at $X = x$ directly from the joint PDF by integrating the joint PDF over all values y such that $(x, y) \in R_{X,Y}$. The result is still a density, not a probability.

Example 4.10
Let X, Y be jointly continuous random variables with PDF given by $f_{X,Y}(x,y) = \begin{cases} c(1-x-y) & 0 \le x \le (1-y) \le 1 \\ 0 & \text{otherwise,} \end{cases}$ where c is a constant that needs to be determined so that $\int_{-\infty}^{\infty} f_{X,Y}(x,y)d\,xd\,y = 1$. Find the value of c, and then the marginal PDF $f_X(x)$. Also, compute the probability that $X < Y$.

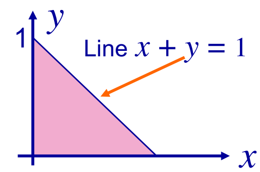


Figure 4.9: Range $R_{X,Y}$.

To begin with, it is always useful to visualize the range $R_{X,Y}$ where the joint PDF is non-zero. In this case, it is a triangle defined by the inequalities $0 \le x \le (1-y) \le 1$. Figure 4.9 shows this area, a triangle defined by the three inequalities $x \ge 0, y \ge 0, x + y \le 1$. This will help us evaluate the limits of integration for computing the marginal densities or the constant of integration. Let's compute c first. We can do this with geometry if we visualize the graph of the joint PDF as a pyramid, as shown in Figure 4.10 because the joint PDF is a linear function of x, y. Since we know the volume of a pyramid is $(1/3)\times$ base area $\times$ height, and the height is c, the volume under the joint PDF is

$$\iint_{R_{X,Y}} f_{X,Y}(x,y)dxdy = \frac{1}{3} \times \frac{1}{2} \times c = \frac{c}{6} = 1,$$

which implies that $c = 6$.

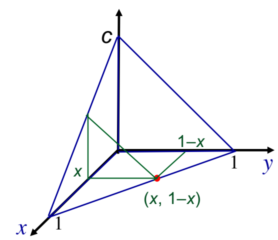


Figure 4.10: Joint PDF.

Alternatively, we compute this more generally from the double integral directly. Using the diagrams of Figures 4.9 and 4.10 to set the limits of integration, we obtain:

$$\iint_{R_{X,Y}} f_{X,Y}(x,y)dxdy = \int_0^1 \left(\int_0^{1-x} c(1-x-y)\,dy \right) dx$$
$$= \int_0^1 \frac{c(1-x)^2}{2} dx = \frac{c}{6}$$

and we get the same answer, $c = 6$.

To compute the marginal $f_X(x)$, we integrate the joint PDF over the range of possible values of Y with nonzero joint PDF for a given value $X = x$. Using the diagram of Figure 4.9 to set limits, we see that the range of values of Y for a given $X = x$ is $y \in [0, 1-x]$. Thus,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^{1-x} 6(1-x-y)dy = 3(1-x)^2 & x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\int_{-\infty}^{\infty} f_X(x)dx = 1$, which is the normalization property of PDFs. By symmetry, we also get that the marginal PDF of Y is $f_Y(y) = \begin{cases} 3(1-y)^2 & y \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$

Finally, we compute the probability that $X \le Y$. If we are really clever, we see that the line $X = Y$ bisects $R_{X,Y}$ into two equal regions, so the volume under the joint PDF in the region $\{(x,y) : (x,y) \in R_{X,Y}, x \le y\}$ is exactly 1/2. However, let's avoid cleverness and compute this as an integral, as one would have to do in a more general setting. The key is to visualize the region, and set the right limits for the integrals. We note that the maximum value for X such that $X \le Y$ is 1/2, and that, for each value $X = x$, the region of values of y that we are interested is $y \in [x, 1-x]$. Then,

$$\mathbb{P}[\{X \le Y\}] = \int_0^{1/2} (\int_x^{1-x} 6(1-x-y)dy)dx = \int_0^{1/2} \frac{3}{2}(1-2x)^2 dx = \frac{1}{2}.$$

Example 4.11
Consider a pair of continuous random variables X, Y, uniformly distributed on the unit disk with radius 1, centered at (0,0). Thus, the joint PDF of X, Y is given by

$$f_{X,Y} = \begin{cases} \frac{1}{\pi} & 0 \le x^2 + y^2 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The joint PDF of X, Y is illustrated in Figure 4.11.
For this problem, we want to compute $\mathbb{E}[X], \mathbb{E}[Y]$. We also want to compute the marginal PDFs $f_X(x), f_Y(y)$. Note that, by symmetry of the density, both $\mathbb{E}[X] = 0, \mathbb{E}[Y] = 0$.

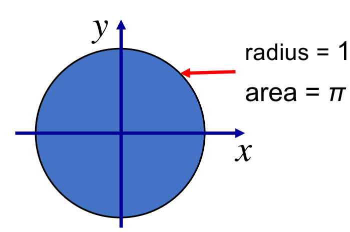


Figure 4.11: Example 4.11.

To compute the marginal density, it is useful to examine the range $R_{X,Y}$ illustrated in Fig. 4.11. For a given value of x, the values of y range from $-\sqrt{1-x^2}$ to $\sqrt{1-x^2}$. Then, the marginal density $f_X(x)$ is computed as:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy = \frac{2\sqrt{1-x^2}}{\pi}.$$

By symmetry, it is clear that

$$f_Y(y) = \frac{2\sqrt{1-xy^2}}{\pi}.$$

Example 4.12
Consider joint continuous random variables X, Y with join PDF $f_{X,Y}(x,y) = \begin{cases} e^{-x} & 0 \le y \le x \le \infty \\ 0 & \text{otherwise} \end{cases}$.

Compute the marginal PDFs for X, Y, and compute the probability that $X + Y \le c$ for a constant $c \ge 0$.

It is useful to visualize the region $R_{X,Y}$ where the joint PDF is positive. This is an infinite triangle in the x-y plane, with origin at (0,0), bounded by the x-axis and the line $x = y$. Using this to compute limits, we see that, for a fixed x, the range of possible values of y is from 0 to x; for a fixed y, the range of possible x is from y to ∞. Thus,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^x e^{-x}dy = xe^{-x} & x \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \begin{cases} \int_y^\infty e^{-x}dx = e^{-y} & y \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

To compute the probability that $X + Y \le c$, for $c \ge 0$, visualize the area in the x-y plane where $(x,y) \le c$ and $(x,y) \in R_{X,Y}$: This is a triangular area, where $y \in [0, c/2]$, and $x \in [y, c-y]$. Denote this area as B. This helps set the limits for the integrals to compute the probability as:

$$\begin{aligned}\mathbb{P}[\{X+Y \le c\}] &= \iint_B f_{X,Y}(x,y)dxdy = \int_0^{c/2} (\int_y^{c-y} e^{-x}dx)dy \\ &= \int_0^{c/2} (e^{-y} - e^{y-c})dy = 1 - e^{-c/2} - e^{-c/2} + e^{-c} = (1-e^{-c/2})^2.\end{aligned}$$

This last computation is useful if we wanted to define a derived random variable $Z = X + Y$. We have just computed

$$F_Z(z) = \mathbb{P}[\{Z \le z\}] = \mathbb{P}[\{X+Y \le z\}] = \begin{cases} (1-e^{-z/2})^2 & z \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

From this, we can compute the PDF of Z as

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \begin{cases} (1-e^{-z/2})e^{-z/2} & z \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is a useful technique for computing the PDFs of derived random variables: Compute the CDF first, then differentiate to get the PDF.

4.4.3 Conditional PDF

We want to extend the concept of conditional probability to jointly continuous random variables. Let X, Y be jointly continuous random variables, and define the set $A = \{(X,Y) \in B\}$ for some $B \subset R_{X,Y}$. Conditioned on observing that A has occurred, we define the conditional CDF of X, Y given A using the definition of conditional probability for events, as

$$F_{X,Y|A}(x,y|A) = \begin{cases} \frac{\mathbb{P}[\{X \le x, Y \le y\} \cap A]}{\mathbb{P}[A]}, & \mathbb{P}[A] > 0 \\ \text{undefined} & \mathbb{P}[A] = 0\ . \end{cases}$$
$$= \begin{cases} \frac{\iint\limits_{(-\infty,x]\times(-\infty,y]\cap B} f_{X,Y}(x,y)dxdy}{\iint\limits_{B} f_{X,Y}(x,y)dxdy} & \mathbb{P}[A] > 0 \\ \text{undefined} & \mathbb{P}[A] = 0\ . \end{cases}$$

That is, we restrict the probability to values $(x, y) \in B$, and rescale the probability so that it satisfies the normalization properties. From this, we can obtain the conditional density as

$$f_{X,Y|A}(x,y|A) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y|A}(x,y|A).$$

This yields the result:

$$f_{X,Y|A}(x,y|A) = \begin{cases} \frac{f(x,y)}{\mathbb{P}[A]}, & (x,y) \in A, \mathbb{P}[A] > 0 \\ 0, & (x,y) \notin A, \mathbb{P}[A] > 0 \\ \text{undefined} & \mathbb{P}[A] = 0\ . \end{cases}$$

which has the same interpretation we saw previously: we restrict the range of the conditional density to values in the observed set A, and we rescale the conditional density to satisfy the normalization property.

We are also interested in the conditional probability of X given observations of values of Y. Consider first observing the event $A = \{Y \le y\}$. From the definition of conditional probability for events,

$$F_{X|A}(x|A) = \frac{\mathbb{P}[\{X \le x\} \cap A]}{\mathbb{P}[A]} = \frac{\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x',y')dx'dy'}{F_Y(y)}.$$

for all y such that $F_Y(y) > 0$. From this conditional CDF, we compute the conditional PDF of X given A, as

$$f_{X|A}(x|A) = \frac{d}{dx}F_{X|A}(x|A) = \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')dy'}{F_Y(y)}.$$

Example 4.13
Let X, Y be jointly continuous random variables with joint PDF $f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le x \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$. Let $A = \{Y \le 0.5\}$. Compute the conditional density of X given that A is observed.

Note that $R_{X,Y}$, the range where the joint PDF is positive, is a triangle formed by the lines $x = 0, x = y, y = 1$, which helps us identify the limits of integration. This is shown in Figure 4.12. Proceeding as above,

$$\mathbb{P}[A] = F_Y(0.5) = \int_0^{0}.5(\int_0^{y} 2dx)dy = 0.25 \ \ \text{(2 times the area of orange triangle)}$$
$$f_{X|A}(x|A) = \frac{\int_{-\infty}^{0.5} f_{X,Y}(x,y')dy'}{F_Y(0.5)} = \begin{cases} \frac{\int_x^{0.5} 2dy}{0.25} = 4 - 2x & x \in (0, 0.5) \\ 0 & \text{otherwise.} \end{cases}$$

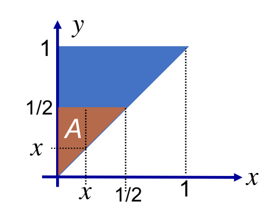


Figure 4.12: Range $R_{X,Y}$.

What if we observe the event $A = \{Y = y\}$? In this case, $\mathbb{P}[A] = 0$, so we cannot apply the definitions of conditional probability for events. We will use a limiting argument to define a conditional PDF of X given observation of the event $\{Y = y\}$, as follows.

Define the event $B = \{Y \in (y, y + \Delta]\}$ for some $\Delta > 0$. Then, $\mathbb{P}[B] = F_Y(y + \Delta) - F_Y(y)$. Assume we select y so that $\mathbb{P}[B] > 0$; that is, we select y in the interior of R_Y. Then, we define the conditional CDF and PDF of X as:

$$F_{X|B}(x|B) = \frac{\mathbb{P}[\{X \leq x\} \cap B]}{\mathbb{P}[B]} = \frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta} f_{X,Y}(x', y')dx'dy'}{F_Y(y+\Delta) - F_Y(y)}.$$

From this conditional CDF,we get the density by differentiation:

$$f_{X|B}(x|B) = \frac{d}{dx} F_{X|B}(x|B) = \frac{\int_{y}^{y+\Delta} f_{X,Y}(x, y')dy'}{\int_{y}^{y+\Delta} f_Y(y')dy'}.$$

If we take limits as $\Delta \to 0$ in the above expression, both the numerator and denominator go to zero. However, using L'Hopital's rule, we can compute the limit as:

$$\lim_{\Delta \to 0} \frac{\int_{y}^{y+\Delta} f_{X,Y}(x, y')dy'}{\int_{y}^{y+\Delta} f_Y(y')dy'} = \lim_{\Delta \to 0} \frac{\frac{d}{d\Delta} \int_{y}^{y+\Delta} f_{X,Y}(x, y')dy'}{\frac{d}{d\Delta} \int_{y}^{y+\Delta} f_Y(y')dy'} = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

as long as $f_Y(y) > 0$. This allows us to define the conditional PDF of X when $Y = y$ as this limit:

Definition 4.3
Given two jointly continuous random variables X, Y, the **conditional PDF** of X given that $Y = y$ is given by

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) > 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Similarly, the conditional PDF of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & f_X(x) > 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

The conditional PDF of X given $Y = y$ is a probability density for the continuous random variable X, and thus satisfies the following basic properties of probability densities:

- **Non-negativity:** $f_{X|Y}(x|y) \geq 0$ and $f_{Y|X}(y|x) \geq 0$ for all x and y where $f_X(x) > 0, f_Y(y) > 0$.
- **Normalization:** $\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$ for any y such that $f_Y(y) > 0$, and $\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dy = 1$ for any x such that $f_X(x) > 0$.
- **Additivity:** For any event $B \subset R_X$, the probability that X takes values in B given $Y = y$ is

  $$\mathbb{P}[\{X \in B\}|\{Y = y\}] = \int_B f_{X|Y}(x|y)\, dy.$$

  For any event $C \subset R_Y$, the probability that Y takes values in C given $X = x$ is

  $$\mathbb{P}[\{Y \in C\}|\{X = x\}] = \int_C f_{Y|X}(y|x)\, dx.$$

The techniques we developed for conditional probabilities also apply to conditional PDFs:

- **Multiplication Rule:** $f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$.
- **Law of Total Probability:**

  $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$$
  $$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, f_X(x)\, dx$$

- **Bayes' Rule:**

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$
$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}.$$

Example 4.14
Consider two jointly continuous random variables X, Y, with joint PDF given by

$$f_{X,Y}(x,y) = \begin{cases} 6(1-x-y) & 0 \leq x \leq 1-y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

This is the same joint PDF considered in Example 4.10, illustrated in Figures 4.9 and 4.10. Let $\alpha \in (0,1)$. Compute $f_{X|Y}(x|\alpha)$.

Note that in Example 4.10, we computed the marginal PDF of Y as $f_Y(y) = \begin{cases} 3(1-y)^2 & y \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$

From the definition of conditional PDF, we have

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0, \\ \text{undefined} & \text{elsewhere.} \end{cases}$$

We need to be careful to account for limits in substituting in the numerator. Note that, if $Y = \alpha$, then $f_{X,Y}(x,\alpha) = 0$ if $x > 1-\alpha$. Thus,

$$f_{X|Y}(x|\alpha) = \begin{cases} \frac{6(1-x-\alpha)}{3(1-\alpha)^2} = 2\frac{(1-x-\alpha)}{(1-\alpha)^2} & \text{if } \alpha \in (0,1),\ x \in (0,1-\alpha), \\ 0 & \text{if } \alpha \in (0,1),\ x \notin (0,1-\alpha), \\ \text{undefined} & \alpha \notin (0,1). \end{cases}$$

4.5 Conditional Probability and Expectation

Given two discrete random variables X, Y, we have defined the conditional probability mass function $P_{X|Y}(x|y)$ as the probability that $X = x$ given that we have observed the event $Y = y$. Using this conditional PMF, we define the **conditional expected value of a function** $g(X)$ **given** $Y = y$ as

$$\mathbb{E}[g(X)|Y=y] = \sum_{x \in R_X} g(x)P_{X|Y}(x|y).$$

Note that this expected value is a function of y, as we are averaging $g(X)$ over the conditional PMF of X given $Y = y$. Denote $h(y) = \mathbb{E}[g(X)|Y=y]$. Then, we can compute the expected value of $h(Y)$ over the PMF of Y, as

$$\mathbb{E}[h(Y)] = \sum_{y \in R_Y} h(y)P_Y(y).$$

Let's combine the last two equations, to get:

$$\begin{aligned}
\mathbb{E}[h(Y)] = \mathbb{E}[\mathbb{E}[g(X)|Y]] &= \sum_{y \in R_Y} \mathbb{E}[g(X)|Y=y]P_Y(y) \\
&= \sum_{y \in R_Y} \sum_{x \in R_X} g(x)P_{X|Y}(x|y)P_Y(y) \\
&= \sum_{y \in R_Y} \sum_{x \in R_X} g(x)P_{X,Y}(x,y) \\
&= \mathbb{E}[g(X)]
\end{aligned}$$

This last result is known as the smoothing property of conditional expectations. Basically, $\mathbb{E}\big[\mathbb{E}[g(X)|Y]\big] = \mathbb{E}[g(X)]$. In particular, this is true for the function $g(X) = X$, so that $\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[X]$.

We can obtain a similar result for jointly continuous random variables X, Y. Given a function $g(x)$ we compute the conditional expected value of $g(X)$ given $Y = y$ using the conditional PDF of X given $Y = y$ as

$$\mathbb{E}[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx.$$

Note that this will be a function of y, which we denote as $h(y)$. Proceeding as before,

$$\begin{aligned}
\mathbb{E}[h(Y)] = \mathbb{E}[\mathbb{E}[g(X)|Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[g(X)|Y = y] f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx) f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) f_Y(y)\, dx\, dy. \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy. = \mathbb{E}[g(X)]
\end{aligned}$$

which shows the smoothing property of conditional expectations also holds for jointly continuous random variables.

Example 4.15
Let X be a continuous random variable, uniformly selected in $(0, 1)$. Hence, $f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$

Given that $X = x$, select Y to be a uniform random variable on $[0, x]$. That is,

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x} & y \in [0, x] \\ .0 & \text{otherwise.} \end{cases}$$

Combining these two densities, we have

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \frac{1}{x} & 0 \le y \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}[X|Y = y]$, and $\text{Var}[X|Y = y]$.

We first compute the marginal density of y, given by

$$f_Y(y) = \int_y^1 \frac{1}{x}\, dx = -\ln(y).$$

Note that this integrates to 1 for $y \in [0, 1]$, as a density should. To compute the conditional density of X given $Y = y$, we need to compute the conditional density $f_{X|Y}(x|y)$, which we do using Bayes' Rule and the Law of Total Probability as

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{\frac{1}{x}}{f_Y(y)},\ 0 \le y \le x \le 1 \\
&= \frac{\frac{1}{x}}{-\ln(y)} = \frac{-1}{x\ln(y)},\ 0 \le y \le x \le 1
\end{aligned}$$

Note how the limits of integration were evaluated for computing $f_Y(y)$, as we know that $x \in [y, 1]$. Using this conditional density, we get

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx = \int_y^1 x\frac{-1}{x\ln(y)} dx = \frac{y-1}{\ln(y)},\ y \in (0, 1)$$

Let's now compute $\mathbb{E}[X]$ as:

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|Y = y]\big] = \int_0^1 \frac{y-1}{\ln(y)} f_Y(y)\, dy = \int_0^1 \frac{y-1}{\ln(y)}(-\ln(y))\, dy = \int_0^1 (1-y)\, dy = \frac{1}{2}.$$

which is exactly what it should be, as X was a uniform random variable on $[0,1]$.

To compute the conditional variance of X given $Y=y$, we compute first $\mathbb{E}[X^2|Y=y]$:

$$\begin{aligned}\mathbb{E}[X^2|Y=y] &= \int_0^1 x^2 \frac{-1}{x\ln(y)}\,dx = \frac{-1}{\ln(y)}\int_0^1 x\,dx \\ &= \frac{-1}{2\ln(y)} \\ \mathsf{Var}[X|Y=y] &= \mathbb{E}[X^2|Y=y] - \Big(\mathbb{E}[X|Y=y]\Big)^2 = \frac{-1}{2\ln(y)} - \frac{(1-y)^2}{(\ln(y))^2}\end{aligned}$$

4.6 Independence of Pairs of Random Variables

In a probability space $(\Omega,\mathcal{E},\mathbb{P})$, two events $A,B\in\mathcal{E}$ are called independent if $\mathbb{P}[A\cap B]=\mathbb{P}[A]\mathbb{P}[B]$. For pairs of random variables X, Y, the concept of independence is stronger: we want events of the type $A=\{X\in C\subset R_X\}$ and $B=\{Y\in D\subset R_Y\}$ to be independent for any choice of $C\subset R_X, D\subset R_Y$. Fortunately, there is a simple way to check for independence without checking all such pairs of events. If the sets $C=(-\infty,x]$, and $D=(\infty,y]$, then $\mathbb{P}[\{X\in C\}\cap\{Y\in D\}]=F_{X,Y}(x,y)$. Furthermore, $\mathbb{P}[\{X\in C\}]=F_X(x)$ and $\mathbb{P}[\{Y\in D\}]=F_Y(y)$. Thus, independence requires that $F_{X,Y}(x,y)=F_X(x)F_Y(y)$ for all $(x,y)\in\Re^2$. It turns out that this condition is also sufficient to guarantee that $\mathbb{P}[\{X\in C\}\cap\{Y\in D\}]=\mathbb{P}[\{X\in X\}]\mathbb{P}[\{Y\in D\}]$ for any sets C and D defined by unions and intersections of intervals (Borel sets), because all those probabilities can be computed from the joint and marginal CDFs.

Definition 4.4
A pair of random variables X and Y are **independent** if and only if $F_{X,Y}(x,y)=F_X(x)F_Y(y)$.

For pairs of discrete random variables, the above condition leads to a characterization of independence in terms of the probability mass functions, as follows:

Lemma 4.1
A pair of discrete random variables X,Y are independent if and only if $P_{X,Y}(x,y)=P_X(x)P_Y(y)$.

Proof: To show the if part, assume $P_{X,Y}(x,y)=P(x)P(y)$. Note this means $R_{X,Y}=R_X\times R_Y$, because $P_{X,Y}(x,y)>0$ implies both $P_X(x)$ and $P_Y(y)$ are positive. Then,

$$\begin{aligned}F_{X,Y}(x,y) &= \sum_{\substack{(x_i,y_j)\in R_{X,Y}\\ x_i\le x,y_j\le y}} P(x_i)P(y_j) = \sum_{\substack{x_i\in R_X\\ x_i\le x}}\sum_{\substack{y_j\in R_Y\\ y_j\le y}} P(x_i)P(y_j) \\ &= \sum_{\substack{x_i\in R_X\\ x_i\le x}} P(x_i)\sum_{\substack{y_j\in R_Y\\ y_j\le y}} P(y_j) = F_X(x)F_Y(y)\end{aligned}$$

and hence the random variables X,Y are independent.

To show the only if part, assume X,Y are independent, so $F_{X,Y}(x,y)=F_X(x)F_Y(y)$. Again, this implies $R_{X,y}=R_X\times R_y$, because $F_{X,Y}(x,y)$ must change values everywhere $F_X(x)$ changes value and $F_Y(y)$ changes value. Let (x,y) be a point in $R_{X,Y}$. Since R_X,R_Y are discrete sets, there is an $\epsilon>0$ such that $F_X(x)-F_X(x-\epsilon)=P_X(x)$ and $F_Y(y)-F(y-\epsilon)=P_Y(y)$. We want to compute $\mathbb{P}[\{X\in(x-\epsilon,x]\}\cap\{Y\in(y-\epsilon,x]\}=P_{X,Y}(x,y)$. In Example 4.1 we showed that

$$\mathbb{P}[\{X\in(x-\epsilon,x]\}\cap\{Y\in(y-\epsilon,x]\}]=F_{X,Y}(x,y)-F_{X,Y}(x-\epsilon,y)-F_{X,Y}(x,y-\epsilon)+F_{X,Y}(x-\epsilon,y-\epsilon)$$

Since X, Y are independent, this means

$$\begin{aligned}
P_{X,Y}(x,y) &= F_X(x)F_Y(y) - F_X(x-\epsilon)F_Y(y) - F_X(x)F_Y(y-\epsilon) + F_X(x-\epsilon)F_Y(y-\epsilon) \\
&= (F_X(x) - F_X(x-\epsilon))F_Y(y) - (F_X(x) - F_X(x-\epsilon))F_Y(y-\epsilon) \\
&= P_X(x)F_Y(y) - P(x)F_Y(y-\epsilon) = P_X(x)(F_Y(y) - F_Y(y-\epsilon)) \\
&= P_X(x)P_Y(y)
\end{aligned}$$

For pairs of continuous random variables, we have a similar equivalent condition in terms of probability density functions:

Lemma 4.2
A pair of jointly continuous random variables X, Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

proof: The if direction is easy to prove, because

$$\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x',y')dy'dx' = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x',y')dy'dx' = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(x')f_Y(y')dy'dx' \\
&= (\int_{-\infty}^{x} f_X(x')dx')(\int_{-\infty}^{y} f_Y(y)dy')) = F_X(x)F_Y(y)
\end{aligned}$$

and hence, X and Y are independent.

To show the only if direction, let X, Y be independent. Then, $F_{X,Y}(x,y) = F_X(x)F_y(y)$. Then,

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_X(x)F_y(y) \\
&= (\frac{\partial}{\partial x}F_X(x))(\frac{\partial}{\partial y}F_Y(y)) = f_X(x)f_Y(y),
\end{aligned}$$

establishing the result.

Independence is one of the most important properties used in modeling experiments with multiple random variables. By assuming independence, we can describe the two-dimensional joint PDF as a product of two one-dimensional PDFs.

Independence between a pair of random variables has implications on the conditional probability. For a pair of discrete random variables X, Y, we know that the conditional probability mass function of X given observations that $Y = y$ satisfies the following relationship: $P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y)$.. If X, Y are independent, then $P_{X,Y}(x,y) = P_X(x)P_Y(y)$. This means that, for independent X, Y, the conditional probability mass function is equal to the marginal, unconditional probability mass function:

$$P_{X|Y}(x|y) = P_X(x) \text{ for all } y \in R_Y \ .$$

A similar result applies to jointly continuous random variables X, Y that are independent. For jointly continuous X, Y, we know $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$. If X, Y are independent, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Thus, for jointly continuous, independent X, Y, we have

$$f_{X|Y}(x|y) = f_X(x) \text{ for all } y \in R_Y \ .$$

Independence between pairs of random variables is often a property that is assumed. To prove that a pair of random variables are independent, one would have to verify the factorization property $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ or $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all values (x,y). In some cases, we can recognize that X, Y are dependent simply by looking at the range sets $R_X, R_Y, R_{X,Y}$. Specifically, if X, Y are independent random variables, then $R_{X,Y} = R_X \times R_Y$. For discrete random variables, this is because $P_{X,Y}(x,y) > 0$ only if both $P_X(x), P_Y(y) > 0$. Thus, to recognize two discrete random variables X, Y are dependent, we simply need to find a pair (x, y) where $P_{X,Y}(x, y) = 0$, but $P_X(x) > 0, P_Y(y) > 0$. We can recognize this by finding a zero entry in the table representation of the joint PMF, where neither the entire row nor the entire column containing that entry is zero.

For jointly continuous, independent random variables X, Y, $R_{X,Y} = R_X \times R_Y$ follows because $f_{X,Y}(x, y) > 0$ only if both $f_X(x), f_Y(y) > 0$. Thus, to recognize that two random variables are dependent, we simply need to find a pair $(x, y) \in \Re^2$ where $f_{X,Y}(x, y) = 0$, but $f_X(x) > 0, f_Y(y) > 0$. We can recognize this by finding a point $(x', y') \in \Re^2$ where $f_{X,Y}(x', y') = 0$, but either the line $x = x'$ or the line $y = y'$ intersect $R_{X,Y}$. Thus, for jointly continuous, independent random variables, the range $R_{X,Y}$ must be of rectangular type with boundaries parallel to the edges.

Note that showing $R_{X,Y} = R_X \times R_Y$ is insufficient to show independence of X, Y. It is a necessary condition, so if it is not satisfied, then the random variables are not independent.

Example 4.16
Consider two discrete random variables X, Y with the joint PMF function used in examples 4.3 and 4.4, which is shown in the table below. We can quickly see that X, Y are not independent, as $P_{X,Y}(4, 4) = 0$, but the row corresponding to

$Y \backslash X$	1	2	3	4
1	0	$\frac{1}{20}$	0	0
2	0	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	0

$Y = 4$ and the column corresponding to $X = 4$ are not identically zero. We could have picked several other zero entries to verify that X, Y are not independent.

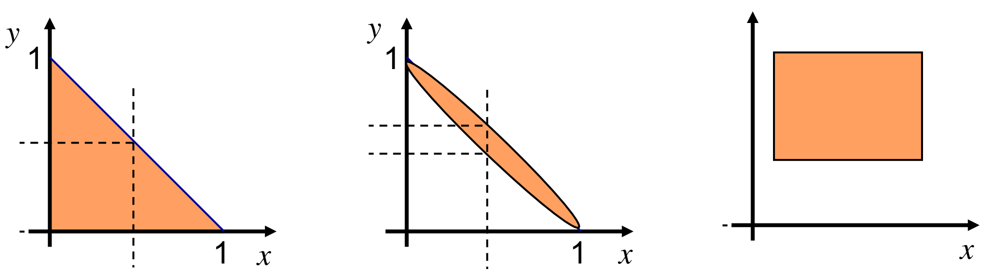


Figure 4.13: Figure for example 4.17.

Example 4.17
Assume X, Y are jointly continuous random variables with range $R_{X,Y}$ as one of the three ranges depicted in Figure 4.13. For which one of the three ranges can X, Y be independent random variables?

Consider the range on the left. We can select a point $(x, y) \notin R_{X,Y}$, such as $(0.6, 0.6)$, where $f_{X,Y}(x, y) = 0$. However, $f_X(0.6) > 0$, and $f_Y(0.6) > 0$, because the line $x = 0.6$ intersects the range $R_{X,Y}$ with a non-zero length, and the line $y = 0.6$ intersects the range $R_{X,Y}$ also with a non-zero length. Therefore, X and Y cannot be independent random variables.

Consider next the range in the center. Again, we can select a point $(x, y) = (0.3, 0.3) \notin R_{X,Y}$ so that the vertical and horizontal lines through this point intersect $R_{X,Y}$ with non-zero length. This implies that $f_{X,Y}(0.3, 0.3) = 0$ while $f_X(0.3) > 0, f_Y(0.3) > 0$, so X, Y cannot be independent.

On the other hand, if the range $R_{X,Y}$ is as depicted in the figure on the right, then we cannot find a point $(x', y') \notin R_{X,Y}$ where both the vertical line $x = x'$ and the horizontal line $y = y'$ have positive length intersection with $R_{X,Y}$. In this case, it is possible that X, Y are independent. To show independence, we need to verify that, for all $(x, y) \in R_{X,Y}$, we have $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

4.7 Expected Value of a Function of Two Random Variables

In the previous section, we have developed the concept of joint probability mass functions, and joint probability density functions, to characterize the properties of pairs of random variables X, Y on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Consider now a function $g : \Re^2 \to \Re$. This function defines a new random variable $W = g(X, Y)$. We can compute the **expected value**, or mean of W, using the joint PMF or the joint PDF of X, Y, as follows:

$$\text{Discrete:} \quad \mathbb{E}[W] = \sum_{x \in R_X} \sum_{y \in R_Y} g(x,y) P_{X,Y}(x,y)$$
$$\text{Continuous:} \quad \mathbb{E}[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

In either case, note that the expectation operation is a **linear operation:** For any functions $g_1(x, y), g_2(x, y)$, and constants a_1, a_2,

$$\mathbb{E}\big[a_1 g_1(X,Y) + a_2 g_2(X,Y)\big] = a_1 \mathbb{E}[g_1(X,Y)] + + a_2 \mathbb{E}[g_n(X,Y)],$$

This is because the expectation operation is based on summation and integration, both of which are linear operations. That is, for discrete random variables X, Y,

$$\begin{aligned}
\mathbb{E}\big[a_1 g_1(X,Y) + a_2 g_2(X,Y)\big] &= \sum_{x \in R_X} \sum_{y \in R_Y} \big(a_1 g_1(x,y) + a_2 g_2(x,y)\big) P_{X,Y}(x,y) \\
&= \sum_{x \in R_X} \sum_{y \in R_Y} a_1 g_1(x,y) P_{X,Y}(x,y) + \sum_{x \in R_X} \sum_{y \in R_Y} a_2 g_2(x,y) P_{X,Y}(x,y) \\
&= a_1 \sum_{x \in R_X} \sum_{y \in R_Y} g_1(x,y) P_{X,Y}(x,y) + a_2 \sum_{x \in R_X} \sum_{y \in R_Y} g_2(x,y) P_{X,Y}(x,y) \\
&= a_1 \mathbb{E}\big[g_1(X,Y)\big] + a_2 \mathbb{E}\big[g_2(X,Y)\big]
\end{aligned}$$

A similar argument shows the result for jointly continuous random variables using integrals instead of sums.

A useful special case is when the function $g(x, y)$ is an affine function, so that $g(x, y) = ax + by + c$ for some constants a, b, c. In this case,

$$\mathbb{E}[aX + bY + c] = \mathbb{E}[aX] + \mathbb{E}[bY] + \mathbb{E}[c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Note that this is true regardless of whether X, Y are independent or not. It is strictly a consequence of the linearity of the expectation operator $\mathbb{E}[\cdot]$.

However, if X, Y were independent, and $g(x, y) = f_1(x) f_2(y)$ so that it can be written as a separable product of two functions, we have an interesting decomposition. Assume that X, Y were jointly continuous random variables with joint PDF $f_{X,Y}(x, y)$. Then,

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(x) f_2(y) f_{X,Y}(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(x) f_2(y) f_X(x) f_Y(y)\, dx\, dy \quad \text{because } X, Y \text{ are independent,} \\
&= \left(\int_{-\infty}^{\infty} f_1(x) f_X(x)\, dx \right) \left(\int_{-\infty}^{\infty} f_2(y) f_Y(y)\, dy \right) \\
&= \mathbb{E}[f_1(X)]\mathbb{E}[f_2(X)]
\end{aligned}$$

The smoothing property of conditional expectation continues to apply to functions $g(X,Y)$. We show this for jointly continuous X,Y below, as

$$\begin{aligned}\mathbb{E}[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X|Y}(x,y)f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} g(x,y)f_{X|Y}(x,y)\,dx)f_Y(y)\,dy \\ &= \int_{-\infty}^{\infty}\mathbb{E}[g(X,y)|Y=y])f_Y(y)\,dy = \mathbb{E}\big[\mathbb{E}[g(X,Y)|Y]\big]\end{aligned}$$

The above results allow us to compute the expected value of a random variable W that is derived from X,Y by a function $W = g(X,Y)$.

Example 4.18

Assume that the number of people in line at the bank when you arrive is N, where N is random, having a Poisson distribution with parameter α. The time T that it takes to serve each person ahead of you can be described by an exponential distribution with parameter λ, and is independent of N. The time to serve each person is thee same. How long do you expect to wait before someone starts to serve you?

Let W be the time you will wait. W is a function of N, T, as $W = NT$.

Since N,T are independent, $E[W] = E[T]E[N] = \frac{alpha}{\lambda}$.

4.7.1 Transformation of pairs of random variables

In some cases we want to compute the full probability mass function or probability density function of W, depending on whether W is discrete or continuous. If W is discrete with range R_W, then for each $w_i \in R_W$, we can define the inverse image of w as the set $g^{-1}(w) = A_w = \{(x,y) \in R_{X,Y} : g(x,y) = w\}$. As long as the function g is well-behaved, we compute the probability mass function of W as:

$$P_W(w) = \mathbb{P}[\{\omega : (X(\omega),Y(\omega)) \in A_w\}] = \sum_{(x,y)\in A_w} F_{X,Y}(x,y).$$

Thus, we can readily derive the probability mass function of W from the joint probability mass function of X,Y, as long as we can readily compute the inverse image $g^{-1}(w)$.

If X,Y are jointly continuous, and the map $W = g(X,Y)$ results in a continuous random variable W, the above approach is limited because the probability that W takes on a particular value is zero. In this case,we can instead compute the cumulative distribution function $F_W(w)$. Let $B_w = \{(x,y) \in R_{X,Y} : g(x,y) \le w\}$ be the region in $R_{X,Y}$ that maps into values $g(x,y) \le w$. In this case, the CDF $F_W(w)$ can be computed as

$$F_W(w) = \iint_{(x,y)\in B_w} f_{X,Y}(x,y)\,dx\,dy.$$

From the CDF, we can get the PDF of W by differentiation, as $f_W(w) = \frac{d}{dw}F_W(w)$.

The above equations were derived for general functions $g(x,y)$, and require solving for the inverse maps of a region of values B_w when W is continuous W or A_w for discrete W. This can be challenging for complicated functions $g(x,y)$. However, there are cases of functions $g(x,y)$ where these inverse maps are straightforward to compute. For instance, let $g(x,y) = ax + by + c$ be a linear function, where $a,b \ne 0$. Then, the line $ax + by + c = w$ divides the x-y plane into two half planes, one of which is B_w. In particular, let's consider $W = X + Y$.

In this case, for discrete X, Y, the set $A_w = \{(x, y) \in \Re^2 : x + y = w\} = \{(x, w - x) : x \in \Re\}$. Therefore, the PMF of W can be computed as

$$P_W(w) = \sum_{x \in \Re} P_{X,Y}(x, w - x) = \sum_{x \in R_X} P_{X,Y}(x, w - x),$$

where the second equality follows because $P_{X,Y}(x, w-x) = 0$ unless $x \in R_X$.

This operation is illustrated in Figure 4.14. To get the probability mass $P_W(w)$, we sum up all the probability masses $P_{X,Y}(x, y)$ on the line $x + y = w$.

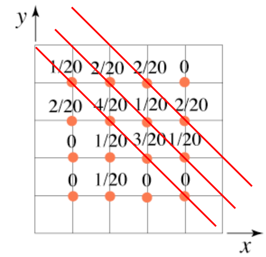


Figure 4.14: Projection to compute PMF of $X + Y$.

For jointly continuous X, Y, the set the set $B_w = \{(x, y) \in \Re^2 : x + y \leq w\} = \{(x, y) : x \in \Re, y \in (-\infty, w - x]\}$. Therefore, we compute the CDF $F_W(w)$ as

$$F_W(w) = \mathbb{P}[\{X + Y \leq w\}] = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{w-x} f_{X,Y}(x, y)\, dy\, dx.$$

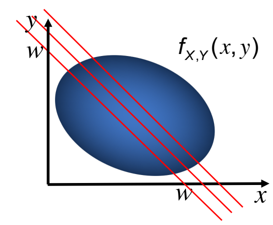


Figure 4.15: Projection to compute PDF of $X + Y$.

From this CDF, we compute the PDF of W by differentiating:

$$f_W(w) = \frac{d}{dw} \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{w-x} f_{X,Y}(x, y)\, dy\, dx = \int_{x=-\infty}^{\infty} (\frac{d}{dw} \int_{y=-\infty}^{w-x} f_{X,Y}(x, y)\, dy)\, dx = \int_{x=-\infty}^{\infty} f_{X,Y}(x, w - x)\, dx.$$

This operation is shown in Figure 4.15. In essence, one integrates the joint PDF along the line $x + y = w$. This is similar to computing a marginal distribution from a joint distribution, except we integrate along an inclined line instead of a vertical or horizontal line.

For the special case that X, Y are independent,

$$f_W(w) = \int_{x=-\infty}^{\infty} f_{X,Y}(x, w - x)\, dx = \int_{x=-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx,$$

which shows that the probability density of the sum of independent random variables X and Y is the convolution of their probability densities.

Example 4.19

Assume we have a pair of continuous random variables X, Y with joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 6(1 - x - y) & x \geq 0, y \geq 0, x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $Z = \max(X, Y)$. Find the probability density function $f_Z(z)$.

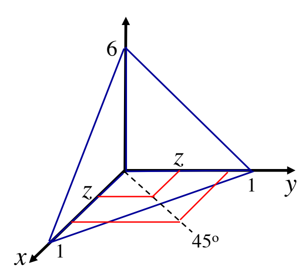


Figure 4.16: Figure for example 4.19.

The joint PDF of X, Y is illustrated in Figure 4.16, where we have drawn also contours for equal values of z, illustrated by the red squares on the x-y plane. We first compute the cumulative distribution of Z for values $z \leq 0.5$. In this range, the region of integration B_z lies entirely in $R_{X,Y}$.

Using the limits as indicated in Figure 4.16, we obtain for $0 \leq z \leq 0.5$,

$$\begin{aligned} \mathbb{P}[\{Z \leq z\}] = F_Z(z) = \mathbb{P}[\{X \leq z\} \cap \{Y \leq z\}] &= \int_0^z \int_0^z f_{X,Y}(x, y)\ dx\ dy \\ &= \int_0^z \int_0^z 6(1 - x - y)\ dx\ dy = \int_0^z 6(1 - y)z - 3z^2\ dy = 6z^2 - 6z^3, z \in [0, 0.5] \end{aligned}$$

Note that $F_Z(z) = 0$ for $z \leq 0$. Furthermore, $F_Z(z) = 1$ for $z > 1$, as the region of integration expands to include all of the range $R_{X,Y}$.

For $z \in [0.5, 1]$, examine the diagram in Figure 4.16. The region of integration B_w now expands beyond $R_{X,Y}$. It is easier to compute this as follows:

$$\begin{aligned}
\mathbb{P}[\{Z \leq z\}] = F_Z(z) &= 1 - \mathbb{P}[\{X \geq z\}] - \mathbb{P}[\{Y \geq z\}] \\
&= 1 - \int_z^1 \int_0^{1-x} 6(1-x-y)\ dy\ dx - \int_z^1 \int_0^{1-y} 6(1-x-y)\ dx\ dy \\
&= 1 - \int_z^1 3(1-x)^2\ dx - \int_z^1 3(1-y)^2\ dy \\
&= 1 - 2(1-z)^3,\ z \in [0.5, 1]
\end{aligned}$$

At $z = 0.5$, $F_Z(z) = \frac{3}{4}$, which agrees with the value computed previously, as $F_Z(z)$ is a continuous function.

The density $f_Z(z)$ is now readily obtained by differentiating, to get

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \begin{cases} 0 & z \notin [0,1], \\ 12z - 18z^2 & z \in (0, 0.5), \\ 6(1-z)^2 & .z \in (0.5, 1). \end{cases}$$

Example 4.20
Let X, Y be independent, uniform$(0,1)$ random variables, and let $Z = X + Y$. Find the PDF of Z.

Note first that the range of Z will be $R_Z = [0, 2]$, the set of values that can have probability. Using the formula provided above for the sum of random variables,

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_{X,Y}(x, z-x)\ dx = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \quad \text{by independence,}$$

We use the fact that $R_X = [0,1], R_Y = [0,1]$ to determine the limits of integration, as follows:

$$f_Z(z) = \begin{cases} 0 & z \notin [0,2], \\ \int_0^1 f_Y(z-x)\ dx = \int_0^z\ dx = z & z \in [0,1], \\ \int_0^1 f_Y(z-x)\ dx = \int_{z-1}^1\ dx = 2-z & z \in [1,2] \end{cases}$$

Example 4.21
Let X, Y be independent, exponential(λ) random variables, and let $Z = X + Y$. Find the PDF of Z.

Note first that the range of Z will be $R_Z = [0, \infty)$, the set of values that can have probability. Using the formula provided above for the sum of random variables,

$$\begin{aligned}
f_Z(z) &= \int_{x=-\infty}^{\infty} f_{X,Y}(x, z-x)\ dx = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \quad \text{by independence,} \\
&= \begin{cases} 0 & z \leq 0, \\ \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\ dx & z \geq 0 \end{cases} = \begin{cases} 0 & z \leq 0, \\ \lambda^2 z e{-}\lambda z\ dx & z \geq 0 \end{cases}
\end{aligned}$$

The sum of two independent exponential random variables with the same rate parameter defines a random variable that has an Erlang$(2, \lambda)$ distribution. If we were to sum n independent exponential random variables with the same rate parameter λ, we obtain an Erlang(n, λ) random variable.

Example 4.22
Let X, Y be independent standard Gaussian random variables, so $X, Y \sim N(0,1)$, and let $Z = aX + bY + c$ for some constants $a \neq 0, b > 0, c$. Find the PDF of Z.

We start by finding the CDF of Z, exploiting the independence of X, Y and the positivity of b, as

$$F_Z(z) = \mathbb{P}[\{aX + bY + c \leq z\}] = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\frac{z-ax-c}{b}} f_X(x) f_Y(y)\, dy\, dx.$$

Differentiating, we get

$$f_Z(z) = \int_{x=-\infty}^{\infty} \left(\frac{d}{dz} \int_{y=-\infty}^{\frac{z-ax-c}{b}} f_X(x) f_Y(y)\, dy \right) dx$$
$$= \int_{x=-\infty}^{\infty} \frac{1}{b} f_X(x) f_Y(\frac{z-ax-c}{b})\, dx$$

Substitute the Gaussian PDF formulas for X, Y to get:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-ax-c)^2}{2b^2}}\, dx$$

Let's manipulate the exponent in the integral to isolate the dependence on x as follows:

$$\begin{aligned}
-\frac{x^2}{2} - \frac{(z-ax-c)^2}{2b^2} &= -\frac{x^2}{2} - \frac{(ax)^2}{2b^2} + \frac{2ax(z-c)}{2b^2} - \frac{(z-c)^2}{2b^2} \\
&= -(1+\frac{a^2}{b^2})\frac{x^2}{2} + \frac{a(z-c)}{b^2}x - \frac{(z-c)^2}{2b^2} \\
&= \frac{(a^2+b^2)}{b^2}\left(-\frac{x^2}{2} + \frac{a(z-c)}{a^2+b^2}x - \frac{(z-c)^2}{2(a^2+b^2)} \right) \\
&= \frac{(a^2+b^2)}{b^2}\left(-\frac{(x-\frac{a(z-c)}{a^2+b^2})^2}{2} + \frac{a^2(z-c)^2}{2(a^2+b^2)^2} - \frac{(z-c)^2}{2(a^2+b^2)} \right) \\
&= -\frac{(x-\frac{a(z-c)}{a^2+b^2})^2}{2K^2} + \frac{a^2(z-c)^2}{2(a^2+b^2)b^2} - \frac{(z-c)^2}{2b^2} \text{ where } K^2 = \frac{b^2}{a^2+b^2} \\
&= -\frac{(x-\frac{a(z-c)}{a^2+b^2})^2}{2K^2} - \frac{(z-c)^2}{2(a^2+b^2)} \quad \text{ where } K^2 = \frac{b^2}{a^2+b^2}
\end{aligned}$$

The reason for this transformation is to express the integral in terms of an integral for a Gaussian PDF. With this transformation, we have

$$f_Z(z) = \frac{1}{b}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\frac{a(z-c)}{a^2+b^2})^2}{2K^2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-c)^2}{2(a^2+b^2)}}\, dx = \frac{1}{\sqrt{2\pi(a^2+b^2)}} e^{-\frac{(z-c)^2}{2(a^2+b^2)}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi K^2}} e^{-\frac{(x-\frac{a(z-c)}{a^2+b^2})^2}{2K^2}}\, dx$$
$$= \frac{1}{\sqrt{2\pi(a^2+b^2)}} e^{-\frac{(z-c)^2}{2(a^2+b^2)}}$$

because the last integral is the integral of a Gaussian PDF with variance K^2 and a given mean, which equals 1 because of the normalization property of PDFs. Furthermore, note that $f_Z(z)$ is also a Gaussian pdf, with mean c and variance a^2+b^2 (thus $Z \sim N(c, a^2+b^2)$.) We have just shown that an affine combination of two independent Gaussians will also be a Gaussian random variable. With a similar argument, we can show that any affine combination of Gaussian random variables will be a Gaussian random variable.

Example 4.23
Let X, Y be independent continuous random variables, and let $Z = \max(X, Y)$. Find the PDF of Z.

In contrast with Example 4.19, we don't specify the pdf of the random variables, but we specify that they are independent. We first derive the CDF of Z:

$$\mathbb{P}[\{Z \le z\}] = \mathbb{P}[\{X \le z\} \cap \{Y \le z\}] = \mathbb{P}[\{X \le z\}]\mathbb{P}[\{Y \le z\}] \quad \text{by independence.}$$

Hence,

$$F_Z(z) = F_X(z)F_Y(z)$$

and the PDF of Z can be obtained as

$$f_Z(z) = \frac{d}{dz}F_Z(z) = F_X(z)f_Y(z) + F_Y(z)f_X(z).$$

Given the CDF and PDF of the random variables X, Y, we can get the CDF and PDF of Z.

To illustrate this, consider the following pair of jointly continuous random variables X, Y, with joint PDF given by

$$f_{X,Y}(x,y) = \begin{cases} (1-\frac{x}{2})(1-\frac{y}{2}) & 0 \le x, y \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Z = \max(X, Y)$. Then,

$$f_X(x) = \begin{cases} (1 - \frac{x}{2}) & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases} \qquad f_Y(y) = \begin{cases} (1 - \frac{y}{2}) & 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0, \\ x - \frac{x^2}{4} & 0 \le x \le 2, \\ 1 & x > 2 \end{cases} \qquad F_Y(Y) = \begin{cases} 0 & y < 0, \\ y - \frac{y^2}{4} & 0 \le y \le 2, \\ 1 & y > 2 \end{cases}$$

Using the above formula,

$$f_Z(z) = \begin{cases} 0 & z < 0, \\ 2(z - \frac{z^2}{4})(1 - \frac{z}{2}) & 0 \le z \le 2, \\ 0 & z > 2. \end{cases}$$

Does the same idea work for the minimum of two random variables? Let $W = \min(X, Y)$. Then,

$$\mathbb{P}[\{W > w\}] = \mathbb{P}[\{X > w\} \cap \{Y > w\}] = \mathbb{P}[\{X > w\}]\mathbb{P}[\{Y > w\}] \quad \text{by independence.}$$

Hence, $1 - F_W(w) = (1 - F_X(w))(1 - F_Y(w))$ which leads to

$$F_W(w) = 1 - (1 - F_X(w))(1 - F_Y(w)) = F_X(w) + F_Y(w) - F_X(w)F_Y(w).$$

Differentiating with respect to w yields

$$f_W(w) = \frac{d}{dw}F_W(w) = (1 - F_Y(w))f_X(w) + (1 - F_X(w))f_Y(w)$$

We conclude this chapter with two examples from a mathematics competition. Questions like these often show up as interview questions for companies like Google. We state first the word problems, and then formulate the problem using pairs of random variables. These examples are difficult, but show how the techniques of this Chapter are used to formulate and solve problems.

Example 4.24

You have a stick of length 1. You pick a point along the stick, uniformly distributed, to break it into two pieces. You take the longer of the two pieces, you pick a point uniformly along that piece, and break the long piece into two pieces. You now have three pieces. What is the expected length of the shortest of the three pieces remaining?

Let X denote the length of the shorter piece remaining after the first break. Since the first break was uniform distributed, it is straightforward to compute the PDF of X as

$$f_X(x) = \begin{cases} 2 & 0 \le x \le 0.5 \\ .0 & \text{otherwise.} \end{cases}$$

The length of the longer piece is $1 - X$. Let Y denote the length of the shortest of the two pieces that remain after breaking the longer piece. Then, Y has conditional PDF

$$f_{Y|X}(y|x) = \begin{cases} \frac{2}{1-x}, & y \in [0, \frac{1-x}{2}] \\ 0 & \text{otherwise.} \end{cases}$$

and thus is distributed uniformly in $[0, \frac{1-X}{2}]$ using a similar argument as before. The joint PDF of X, Y is defined using using the multiplication rule as:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x) = \begin{cases} \frac{4}{1-x} & 0 \le x \le 0.5, 0 \le y \le \frac{1-x}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

We already know that Y is the shortest of the two pieces from the second break, and X is the length of the shortest piece after the first break. Hence, the length of the shortest of the three pieces is $\min(X, Y)$. We have now transformed the original problem into computing the expected value of a function of two random variables, where we know the joint PDF:

$$\mathbb{E}[\min(X, Y)] = \int_0^{\frac{1}{2}} \Big(\int_0^{\frac{1-x}{2}} \min(x, y) \frac{4}{1-x}\, dy \Big)\, dx$$

The rest is tedious calculus that is easy to do with a computer. We have completed the probability part of the problem, and written the correct integral. Nevertheless, let's show the calculus computation. The trick is to figure out the regions where we can write explicitly the minimum of x, y. First, assume $x \geq \frac{1-x}{2}$, which is equivalent to $x \geq \frac{1}{3}$. Then, $\min(x, y) = y$ for $y \in [0, \frac{1-x}{2}]$. Next, if $x < \frac{1}{3}$, then $\min(x, y) = x$ for $y \in [x, \frac{1-x}{2}]$, and $\min(x, y) = y$ for $y \in [0, x]$. We use this to rewrite the integral as:

$$\mathbb{E}[\min(X,Y)] = \int_0^{\frac{1}{3}} \Big(\int_0^{\frac{1-x}{2}} \min(x,y)\frac{4}{1-x}\,dy\Big)\,dx + \int_{\frac{1}{3}}^{\frac{1}{2}} \Big(\int_0^{\frac{1-x}{2}} \min(x,y)\frac{4}{1-x}\,dy\Big)\,dx$$

$$\int_{\frac{1}{3}}^{\frac{1}{2}} \Big(\int_0^{\frac{1-x}{2}} \min(x,y)\frac{4}{1-x}\,dy\Big)\,dx = \int_{\frac{1}{3}}^{\frac{1}{2}} \Big(\int_0^{\frac{1-x}{2}} y\frac{4}{1-x}\,dy\Big)\,dx$$

$$= \int_{\frac{1}{3}}^{\frac{1}{2}} \frac{4(1-x)^2)}{8(1-x)}\,dx = \int_{\frac{1}{3}}^{\frac{1}{2}} \frac{(1-x))}{2}\,dx = \frac{1}{9} - \frac{1}{16}$$

$$\int_0^{\frac{1}{3}} \Big(\int_0^{\frac{1-x}{2}} \min(x,y)\frac{4}{1-x}\,dy\Big)\,dx = \int_0^{\frac{1}{3}} \Big(\int_0^{x} y\frac{4}{1-x}\,dy + \int_x^{\frac{1-x}{2}} x\frac{4}{1-x}\,dy\Big)\,dx$$

$$= \int_0^{\frac{1}{3}} \Big(\frac{2x^2}{1-x} + \frac{2x(1-3x)}{1-x}\Big)\,dx$$

$$= \int_0^{\frac{1}{3}} \frac{2x-4x^2}{1-x}\,dx \approx 0.078.$$

$$\mathbb{E}[\min(X,Y)] = 0.078 + \frac{1}{9} - \frac{1}{16} \approx 0.1266$$

Example 4.25

You have a stick of length 1. You pick two points along the stick, uniformly distributed, selected independently, You break the stick at the two points, resulting in three sticks. What is the expected length of the shortest stick?

The difference in this example from the previous example is that the points are selected independently, not sequentially. Let's propose a formulation using pairs of random variables. Let X be one of the points, and Y be the other. We know the joint PDF of X, Y, given by

$$f_{X,Y}(x,y) = \begin{cases} .1 & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

In terms of X, Y, what is the length of the shortest stick? Let $S(X, Y)$ be this length. If $X > Y$, then $S(X, Y) = \min(Y, X - Y, 1 - X)$. Let B be the event that $X > Y$. By symmetry, $\mathbb{P}[B] = \frac{1}{2}$. Then, the conditional joint PDF of X, Y is given by

$$f_{X,Y|B}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{\mathbb{P}[B]} & (x,y) \in B, \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 2 & 0 \leq y \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Our answer is the conditional expected value of $S(X, Y)$ given the event B, because either X or Y has to be the smallest, so without loss of generality, we call X the smallest. Note that this introduces a factor of 2 to the conditional density, corresponding to mapping the original probability density from the unit square to the triangle $0 \leq y \leq x \leq 1$. Hence, our answer is

$$\mathbb{E}[S(X,Y)|X > Y] = \int_0^1 \Big(\int_0^x 2\min(y, x-y, 1-x)\,dy\Big)\,dx$$

The rest is calculus...it does require breaking down the integral into regions where we can recognize which one of the terms is the minimum so we can do the integrals. A diagram will be most useful. We need to identify the regions in the triangle $0 \leq y \leq x \leq 1$ where $\min(y, x - y, 1 - x) = y$, $\min(y, x - y, 1 - x) = x - y$, and $\min(y, x - y, 1 - x) = 1 - x$ and compute the appropriate expected values in those regions.

The diagram is shown on the right. The three regions have a common point $(2/3, 1/3)$ where all three lengths are equal. Region 1 in the diagram is the region where the minimum is y, so $y < x - y, y < 1 - x$. Hence, this region is $y < x/2, y < 1 - x$. Region 2 is where the minimum is $x - y$, so $x - y < y, x - y < 1 - x$ so $y > x/2, y > 2x - 1$. Region 3 is where the minimum is $1 - x$, so $1 - x < y, 1 - x < x - y$ and therefore $y > 1 - x, y > 2x - 1$. The answer we want is

$$\mathbb{E}[S(X,Y)|X > Y] = \iint_{R_1} 2y\,dx\,dy + \iint_{R_2} 2(x-y)\,dx\,dy + \iint_{R_3} 2(1-x)\,dx\,dy$$

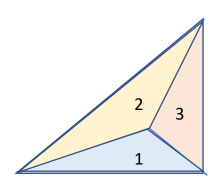


Figure 4.17: Example 4.25.

Computing each integral yields:

$$\iint_{R_1} 2y\,dx\,dy = \int_0^{\frac{1}{3}} \Big(\int_{2y}^{1-y} 2y\,dx\Big)\,dy = \int_0^{\frac{1}{3}} 2y(1-3y)\,dy = \frac{1}{9} - \frac{2}{27} = \frac{1}{27}$$

$$\begin{aligned}
\iint_{R_2} 2(x-y)\,dx\,dy &= \int_0^{\frac{2}{3}} \Big(\int_{\frac{x}{2}}^{x} 2(x-y)\,dy\Big)\,dx + \int_{\frac{2}{3}}^{1} \Big(\int_{2x-1}^{x} 2(x-y)\,dy\Big)\,dx \\
&= \int_0^{\frac{2}{3}} (x^2 - x^2 + \frac{x^2}{4})\,dx + \int_{\frac{2}{3}}^{1} (2x(1-x) - x^2 + (2x-1)^2)\,dx \\
&= \int_0^{\frac{2}{3}} \frac{x^2}{4}\,dx + \int_{\frac{2}{3}}^{1} (2x - 2x^2 - x^2 + 4x^2 - 4x + 1)\,dx \\
&= \frac{1}{12} \cdot \frac{8}{27} + \int_{\frac{2}{3}}^{1} (x^2 - 2x + 1)dx = \frac{2}{81} + \frac{1}{81} = \frac{1}{27}
\end{aligned}$$

$$\iint_{R_3} 2(1-x)\,dx\,dy = \int_{\frac{2}{3}}^{1} \Big(\int_{1-x}^{2x-1} (1-x)\,dy\Big)\,dx = \int_{\frac{2}{3}}^{1} (1-x)^2\,dx = \frac{1}{27}$$

Assembling the answer yields that the expected value of the shortest piece is $\frac{1}{9}$. Note that this is a little shorter than the answer to the previous example. The reason is that, in the previous problem, after we selected the first point, we broke the longer of the two pieces. Here, we select the second break randomly, so we can break the shorter of the two pieces, thereby resulting in shorter pieces. It is useful to check that your answers have common sense explanations.