8 Continuous Random Variables: Joint Distributions

In the previous chapter, we studied individual continuous random variables. We now move on to discussing multiple random variables, which may or may not be independent of each other. Just as in Chapter 3 we used a joint probability mass function (p.m.f.), we now introduce the continuous counterpart, the joint probability density function (joint p.d.f.). We will use the joint p.d.f. to answer questions about the expected value of one random variable, given some information about the other random variable.

8.1 Joint Densities

When dealing with multiple continuous random variables, we can define a joint p.d.f. which is similar to the joint p.m.f. in Definition 3.4.

Definition 8.1 *The* **joint probability density function** *between continuous random variables X and Y is a non-negative function $f_{X,Y}(x,y)$, where*

$$\int_c^d \int_a^b f_{X,Y}(x,y)dxdy = \mathbf{P}\{a \leq X \leq b \ \ \& \ \ c \leq Y \leq d\}$$

and where

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1.$$

Definition 8.1 is illustrated in Figure 8.1.

Example 8.2 (Height and weight of two-year-olds)

Let's say that two-year-olds range in weight from 15 pounds to 35 pounds and range in height from 25 inches to 40 inches. Let W be a r.v. denoting the weight of a two-year-old, and H be a r.v. denoting the height. Let $f_{W,H}(w,h)$ denote the joint density function of weight and height.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

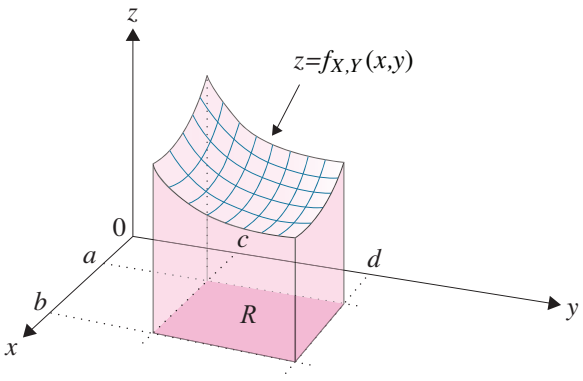


Figure 8.1 *Volume under the curve shows* $\mathbf{P}\{a \le X \le b \ \& \ c \le Y \le d\}$.

Question: What is the fraction of two-year-olds whose weight exceeds 30 pounds, but whose height is less than 30 inches?

Answer:

$$\int_{h=-\infty}^{h=30} \int_{w=30}^{w=\infty} f_{W,H}(w,h)dwdh = \int_{h=25}^{h=30} \int_{w=30}^{w=35} f_{W,H}(w,h)dwdh.$$

These are equivalent because the joint density function is only non-zero in the range where $15 \le w \le 35$ and $25 \le h \le 40$.

We can also integrate the joint p.d.f. over just one variable to get a marginal p.d.f.

Definition 8.3 *The* **marginal densities**, $f_X(x)$ *and* $f_Y(y)$, *are defined as:*

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

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

Note that $f_X(x)$ and $f_Y(y)$ are densities and not probabilities.

Question: If $f_{W,H}(w,h)$ is the joint p.d.f of weight and height in two-year-olds, what is the fraction of two-year-olds whose height is exactly 30 inches?

Answer: The event of having height exactly 30 inches is a zero-probability event, so the answer is zero. We could write

$$\int_{w=-\infty}^{w=\infty} f_{W,H}(w,30)dw = f_H(30), \qquad \text{by Definition 8.3,}$$

but, again, this is a density and hence has zero probability.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

Finally, as in Definition 3.5, we can define independence for continuous random variables.

Definition 8.4 *We say that continuous random variables X and Y are* **independent**, *written $X \perp Y$, if*

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall x, y.$$

Example 8.5 (Joint p.d.f.)

Let

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

Note that $f_{X,Y}(x,y)$ is a proper density in that $\int_0^1 \int_0^1 f_{X,Y}(x,y)dxdy = 1$.

Question: (a) What is $\mathbf{E}[X]$? (b) Is $X \perp Y$?

Answer:

(a) To derive $\mathbf{E}[X]$, we first derive $f_X(x)$. We do this using Definition 8.3.

$$f_X(x) = \int_{y=-\infty}^{y=\infty} f_{X,Y}(x,y)dy = \int_{y=0}^{y=1} (x+y)dy = x + \frac{1}{2}$$

$$\mathbf{E}[X] = \int_{x=-\infty}^{x=\infty} f_X(x) \cdot x dx = \int_{x=0}^{1} \left(x + \frac{1}{2}\right) \cdot x dx = \frac{7}{12}.$$

(b) We will show that X and Y are not independent, using Definition 8.4:

$$f_X(x) = \int_{y=-\infty}^{y=\infty} f_{X,Y}(x,y)dy = x + \frac{1}{2} \qquad \text{for } 0 \le x \le 1$$

$$f_Y(y) = \int_{x=-\infty}^{x=\infty} f_{X,Y}(x,y)dx = y + \frac{1}{2} \qquad \text{for } 0 \le y \le 1.$$

Hence, clearly,

$$f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y).$$

Example 8.6 (Joint p.d.f. for independent random variables)

Question: What is an example of a joint p.d.f. where X and Y are independent?

Answer: Let

$$f_{X,Y}(x,y) = \begin{cases} 4xy & \text{if } 0 \le x,y \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

Again, this is a proper p.d.f., since it integrates to 1. Furthermore:

$$f_X(x) = \int_{y=0}^{1} 4xy dy = 2x \qquad \text{for } 0 \le x \le 1$$

$$f_Y(y) = \int_{x=0}^{1} 4xy dx = 2y \qquad \text{for } 0 \le y \le 1.$$

Hence,

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

as desired.

Example 8.7 (Which Exponential happens first?)

Suppose that the time until server 1 crashes is denoted by $X \sim \text{Exp}(\lambda)$ and the time until server 2 crashes is denoted by $Y \sim \text{Exp}(\mu)$. We want to know the probability that server 1 crashes before server 2 crashes. Assume that $X \perp Y$.

The goal is thus $\mathbf{P}\{X < Y\}$. We will show how to do this by integrating the joint density function between X and Y:

$$\begin{aligned}
\mathbf{P}\{X < Y\} &= \int_{x=0}^{\infty} \int_{y=x}^{\infty} f_{X,Y}(x,y) dy dx \\
&= \int_{x=0}^{\infty} \int_{y=x}^{\infty} f_X(x) \cdot f_Y(y) dy dx \\
&= \int_{x=0}^{\infty} \lambda e^{-\lambda x} \cdot \int_{y=x}^{\infty} \mu e^{-\mu y} dy dx \\
&= \int_{x=0}^{\infty} \lambda e^{-\lambda x} \cdot e^{-\mu x} dx \\
&= \lambda \int_{x=0}^{\infty} e^{-(\lambda+\mu)x} dx \\
&= \frac{\lambda}{\lambda+\mu}.
\end{aligned}$$

Question: Where did we use the fact that $X \perp Y$?

Answer: We used independence in splitting the joint p.d.f. in the second line.

8.2 Probability Involving Multiple Random Variables

We can use the joint p.d.f. to derive expectations involving multiple random variables, via Definition 8.8.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

Definition 8.8 *Let X and Y be continuous random variables with joint p.d.f. $f_{X,Y}(x,y)$. Then, for any function $g(X,Y)$, we have*

$$\mathbf{E}\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\cdot f_{X,Y}(x,y)dxdy.$$

We can also use the joint p.d.f. to define the conditional p.d.f. involving two continuous random variables, thus extending Definition 7.15.

Definition 8.9 (Conditional p.d.f. and Bayes' Law: two random variables) *Given two continuous random variables, X and Y, we define the* **conditional p.d.f.** *of r.v. X given event $Y = y$ as:*

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

The first equality in Definition 8.9 is just the definition of a conditional p.d.f., where now we're conditioning on a zero-probability event, $Y = y$. The second equality is a reapplication of the first equality, but this time with X and Y interchanged. The result is a Bayes' Law, akin to that in Definition 7.15.

Observe that the conditional p.d.f. is still a proper p.d.f. in the sense that:

$$\int_x f_{X|Y=y}(x)dx = 1.$$

Recall the Law of Total Probability for continuous random variables from Theorem 7.13, which we have repeated below in Theorem 8.10 for easy reference.

Theorem 8.10 (Law of Total Probability: Continuous) *Given any event A and any continuous r.v., Y, we can compute $\mathbf{P}\{A\}$ by conditioning on the value of Y as follows:*

$$\mathbf{P}\{A\} = \int_{-\infty}^{\infty} f_Y(y\cap A)dy = \int_{-\infty}^{\infty} \mathbf{P}\{A \mid Y = y\}\, f_Y(y)dy.$$

Here, $f_Y(y\cap A)$ denotes the density of the intersection of the event A with $Y = y$.

Analogously to Theorem 8.10, we can express a density of one r.v. by conditioning on another r.v., as shown in Theorem 8.11.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

Theorem 8.11 (Law of Total Probability: Multiple Random Variables)
Let X and Y be continuous random variables. Then:

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

As a starting example, let's revisit Example 8.7 and show how it can be solved more simply by conditioning.

Example 8.12 (Which Exponential happens first – revisited)

Suppose that the time until server 1 crashes is denoted by $X \sim \text{Exp}(\lambda)$ and the time until server 2 crashes is denoted by $Y \sim \text{Exp}(\mu)$. We want to know the probability that server 1 crashes before server 2 crashes. Assume that $X \perp Y$.

The goal is thus $\mathbf{P}\{X < Y\}$. This time, we derive the quantity by conditioning on the value of X, as follows:

$$\begin{aligned}
\mathbf{P}\{X < Y\} &= \int_0^\infty \mathbf{P}\{X < Y \mid X = x\} \cdot f_X(x)dx \\
&= \int_0^\infty \mathbf{P}\{Y > x \mid X = x\} \cdot \lambda e^{-\lambda x}dx \\
&= \int_0^\infty \mathbf{P}\{Y > x\} \cdot \lambda e^{-\lambda x}dx \\
&= \int_0^\infty e^{-\mu x} \cdot \lambda e^{-\lambda x}dx \\
&= \lambda \int_0^\infty e^{-(\lambda+\mu)x}dx = \frac{\lambda}{\lambda+\mu}.
\end{aligned}$$

Question: Where did we use the fact that $X \perp Y$?

Answer: We used independence to claim that $\mathbf{P}\{Y > x \mid X = x\} = \mathbf{P}\{Y > x\}$; here we assumed that the fact that $X = x$ has no effect on the value of Y.

Now let's consider a more involved example.

Example 8.13 (Relationship between hand-in time and grade)

[Parts of this problem are borrowed from [52]] As a professor, I'm curious about whether there's a relationship between the time when a student turns in their homework and the grade that the student receives on the homework.Let T denote the amount of time prior to the deadline that the homework is submitted. I have noticed that no one ever submits the homework earlier than two days before the

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

homework is due, so $0 \leq T \leq 2$. Let G denote the grade that the homework receives, viewed as a percentage, meaning $0 \leq G \leq 1$. Both G and T are continuous random variables. Suppose their joint p.d.f. is given by

$$f_{G,T}(g,t) = \frac{9}{10}tg^2 + \frac{1}{5}, \qquad \text{where } 0 \leq g \leq 1 \text{ and } 0 \leq t \leq 2.$$

Question:

(a) What is the probability that a randomly selected student gets a grade above 50% on the homework?
(b) What is the probability that a student gets a grade above 50%, given that the student submitted less than a day before the deadline?

Answer: It's easiest to start this problem by determining the marginal density function $f_G(g)$. We will determine $f_T(t)$ as well, for future use:

$$f_G(g) = \int_{t=0}^{t=2} f_{G,T}(g,t)dt = \int_{t=0}^{t=2} \left(\frac{9}{10}tg^2 + \frac{1}{5}\right) dt = \frac{9}{5} \cdot g^2 + \frac{2}{5} \quad (8.1)$$

$$f_T(t) = \int_{g=0}^{g=1} f_{G,T}(g,t)dg = \int_{g=0}^{g=1} \left(\frac{9}{10}tg^2 + \frac{1}{5}\right) dg = \frac{3}{10} \cdot t + \frac{1}{5}. \quad (8.2)$$

To understand the probability that a randomly selected student gets a grade above 50% on the homework, we want $\mathbf{P}\left\{G > \frac{1}{2}\right\}$. We can directly use $f_G(g)$ to get this as follows:

$$\mathbf{P}\left\{G > \frac{1}{2}\right\} = \int_{g=\frac{1}{2}}^{g=1} f_G(g)dg = \int_{g=\frac{1}{2}}^{g=1} \left(\frac{9}{5} \cdot g^2 + \frac{2}{5}\right) dg = \frac{29}{40} = 0.725.$$

To understand the probability that a student gets a grade above 50%, given that the student submitted less than a day before the deadline, we want $\mathbf{P}\left\{G > \frac{1}{2} \,\middle|\, T < 1\right\}$:

$$\begin{aligned}
\mathbf{P}\left\{G > \frac{1}{2} \,\middle|\, T < 1\right\} &= \frac{\mathbf{P}\{G > 0.5 \;\&\; T < 1\}}{\mathbf{P}\{T < 1\}} \\
&= \frac{\int_{g=0.5}^{g=1} \int_{t=0}^{t=1} f_{G,T}(g,t)dtdg}{\int_{t=0}^{t=1} f_T(t)dt} \\
&= \frac{\int_{g=0.5}^{g=1} \int_{t=0}^{t=1} \left(\frac{9}{10}tg^2 + \frac{1}{5}\right) dtdg}{\int_{t=0}^{t=1} \left(\frac{3}{10} \cdot t + \frac{1}{5}\right) dt} \\
&= \frac{0.23125}{0.35} = 0.66.
\end{aligned}$$

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

8.3 Pop Quiz

Density functions can be tricky. Below we quickly summarize what we've learned in the form of a pop quiz. Throughout, assume that X and Y are continuous random variables with joint density function

$$f_{X,Y}(x,y), \qquad \text{where } -\infty < x, y < \infty.$$

Question: What are the marginal densities $f_X(x)$ and $f_Y(y)$?

Answer:

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

Question: What is the conditional density $f_{X|Y=y}(x)$? How about $f_{Y|X=x}(y)$?

Answer:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{8.3}$$

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{8.4}$$

Question: How can we write $f_{X|Y=y}(x)$ in terms of $f_{Y|X=x}(y)$?

Answer: If we substitute $f_{X,Y}(x,y)$ from (8.4) into (8.3), we get:

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

Question: How do we write $\mathbf{P}\{X < a \mid Y = y\}$?

Answer: This is just a question of summing up the conditional density $f_{X|Y=y}(x)$ over all values of x where $x < a$:

$$\mathbf{P}\{X < a \mid Y = y\} = \int_{x=-\infty}^{x=a} f_{X|Y=y}(x)dx.$$

Question: How do we write $f_{Y|Y<a}(y)$ in terms of $f_Y(y)$?

Answer: Intuitively, we're just conditioning on the event that $Y < a$, which narrows the range of values, so the conditional density gets scaled up by a

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

constant factor. Here are all the steps. Let A denote the event that $Y < a$. Then:

$$\begin{aligned} f_{Y|Y<a}(y) &= f_{Y|A}(y) \\ &= \frac{f_Y(y \cap A)}{\mathbf{P}\{A\}} \\ &= \frac{f_Y(y \cap Y < a)}{\mathbf{P}\{Y < a\}} \\ &= \begin{cases} \frac{f_Y(y)}{\mathbf{P}\{Y<a\}} & \text{if } y < a \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Question: How do we write $f_{Y|X<a}(y)$ in terms of $f_{X,Y}(x, y)$?

Answer: In the case of $f_{Y|X<a}(y)$, we define A to be the event that $X < a$. Now we are conditioning on an event A that doesn't involve Y. Because of this, we can't simply scale up the density function, and we must instead return to the joint density. Then the steps are as follows:

$$\begin{aligned} f_{Y|X<a}(y) &= f_{Y|A}(y) \\ &= \frac{f_Y(y \cap A)}{\mathbf{P}\{A\}} \\ &= \frac{f_Y(y \cap X < a)}{\mathbf{P}\{X < a\}} \\ &= \frac{\int_{x=-\infty}^{x=a} f_{X,Y}(x, y)dx}{\mathbf{P}\{X < a\}}. \end{aligned}$$

8.4 Conditional Expectation for Multiple Random Variables

We now move on to expectation. We will extend the definitions from Section 7.5 on conditional expectation to multiple random variables. As before, the key to defining conditional expectation is to use a conditional p.d.f.

Definition 8.14 *Given continuous random variables X and Y, we define:*

$$\mathbf{E}[X \mid Y = y] = \int_x x \cdot f_{X|Y=y}(x)dx.$$

A typical situation where Definition 8.14 might come up is in computing the expected weight of two-year-olds if their height is 30 inches. Another way in which Definition 8.14 is useful is that it allows us to simplify computations of expectation by conditioning, as in Theorem 8.15.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

Theorem 8.15 *We can derive* $\mathbf{E}[X]$ *by conditioning on the value of a continuous r.v.* Y *as follows:*

$$\mathbf{E}[X] = \int_y \mathbf{E}[X \mid Y = y] \cdot f_Y(y)dy.$$

Theorem 8.15 is the direct continuous counterpart to Theorem 4.22. The proof of Theorem 8.15 follows the same lines as that of Theorem 4.22, except that we use Definition 8.14 in place of Definition 4.18.

Let's now return to Example 8.13, this time from the perspective of expectation.

Example 8.16 (Relationship between hand-in time and grade, continued)

Let T denote the number of days prior to the deadline that the homework is submitted. No one ever submits the homework earlier than two days before the homework is due, so $0 \leq T \leq 2$. Let G denote the grade that the homework receives, viewed as a percentage, meaning $0 \leq G \leq 1$. Both G and T are continuous random variables. Their joint p.d.f. is given by

$$f_{G,T}(g,t) = \frac{9}{10}tg^2 + \frac{1}{5}.$$

Question: A random student submits at $T = 0$, that is, exactly when the homework is due. What is the student's expected grade?

Answer:

$$\begin{aligned}
\mathbf{E}[G \mid T = 0] &= \int_{g=0}^{1} g \cdot f_{G|T=0}(g)dg && \text{by Definition 8.14} \\
&= \int_{g=0}^{1} g \frac{f_{G,T}(g,0)}{f_T(0)}dg && \text{by Definition 8.9} \\
&= \int_{g=0}^{1} g \frac{\frac{1}{5}}{\frac{1}{5}}dg \\
&= \frac{1}{2}. && (8.5)
\end{aligned}$$

Question: Who has a higher expected grade: a student who submits exactly when the homework is due, or a student who submits more than 1 day early?

Answer: To answer this, we must compare $\mathbf{E}[G \mid T = 0]$ from (8.5) with $\mathbf{E}[G \mid 1 < T < 2]$.

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

We derive $\mathbf{E}[G \mid 1 < T < 2]$ in the same way as we derived $\mathbf{E}[G \mid T = 0]$:

$$\begin{aligned}
\mathbf{E}[G \mid 1 < T < 2] &= \int_{g=0}^{1} g f_{G|1<T<2}(g)dg \qquad \text{by Definition 7.17} \\
&= \int_{g=0}^{1} g \frac{f_G(g \cap (1 < T < 2))}{\mathbf{P}\{1 < T < 2\}} dg \qquad \text{by Definition 7.15} \\
&= \int_{g=0}^{1} g \frac{\int_{t=1}^{t=2} f_{G,T}(g,t)dt}{\int_{t=1}^{t=2} f_T(t)dt} dg \\
&= \int_{g=0}^{1} g \frac{\int_{t=1}^{t=2} \left(\frac{9}{10}tg^2 + \frac{1}{5}\right) dt}{\int_{t=1}^{t=2} \left(\frac{3}{10}t + \frac{1}{5}\right) dt} dg \qquad f_T(t) \text{ is from (8.2)} \\
&= \int_{g=0}^{1} g \frac{\frac{1}{20}\left(27g^2 + 4\right)}{0.65} dg \\
&= 0.673.
\end{aligned}$$

So the expected grade is higher for those who turn in their homework more than a day early, as compared with those who turn in their homework exactly on time. This makes sense!

8.5 Linearity and Other Properties

In this chapter and the prior one on continuous random variables, we have not bothered to repeat all the prior results that we saw for discrete random variables, such as Linearity of Expectation (Theorem 4.10), Linearity of Variance (Theorem 5.8), and Expectation of a Product (Theorem 4.8). However, all of these results extend to continuous random variables as well. The proofs are straightforward and are deferred to the exercises.

8.6 Exercises

8.1 **Linearity of expectation for continuous random variables**
Let X and Y be continuous random variables. Prove that

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

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

8.2 **Product of continuous random variables**
Let X and Y be continuous random variables, where $X \perp Y$. Prove that

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

8.3 **Two Uniforms**
Let $X \sim \text{Uniform}(0,1)$ and $Y \sim \text{Uniform}(0,2)$ be independent random variables. What is $\mathbf{P}\{X \leq Y\}$?
(a) Solve this via the joint p.d.f. of X and Y.
(b) Solve this by conditioning on X.

8.4 **Quality of service**
A company pays a fine if the time to process a request exceeds 7 seconds. Processing a request consists of two tasks: (a) retrieving the file, which takes some time X that is Exponentially distributed with mean 5; and (b) processing the file, which takes some time Y that is independent of X and is distributed Uniform$(1,3)$. Given that the mean time to process a request is clearly 7 seconds, the company views the fine as unfair, because it will have to pay the fine on half its requests. Is this right? What is the actual fraction of time that the fine will have to be paid?

8.5 **Meeting up**
Eric and Timmy have agreed to meet between 2 and 3 pm to work on homework. They are rather busy and are not quite sure when they can arrive, so assume that each of their arrival times is independent and uniformly distributed over the hour. Each agrees to wait 15 minutes for the other, after which he will leave. What is the probability that Eric and Timmy will be able to meet?

8.6 **Practice with joint random variables**
Let X and Y be continuous random variables with the following joint p.d.f.

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

(a) Start by drawing the region R where the joint p.d.f. is non-zero. This will help you determine the limits of integration for the remaining parts.
(b) What is $f_X(x)$? State the region over which this p.d.f. is non-zero.
(c) What is $f_Y(y)$? State the region over which this p.d.f. is non-zero.
(d) What is $f_{Y|X=x}(y)$, where $x > 0$? State the region over which this p.d.f. is non-zero.
(e) What is $\mathbf{E}[Y \mid X = x]$, where $x > 0$?

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

8.7 **Distance between darts**
We are given a line segment, $[0, 1]$. Kristy and Timmy each independently throw a dart uniformly at random within the line segment. What is the expected distance between Kristy's and Timmy's darts?

8.8 **Comparison of two darts**
We are given a line segment, $[0, 1]$. Two darts are each independently thrown uniformly at random within the line segment. What is the probability that the value of one dart is at least three times the value of the other?

8.9 **Sum of independent random variables**
The *convolution* of two functions $f(\cdot)$ and $g(\cdot)$ is defined as

$$f \circ g(z) = \int_{-\infty}^{\infty} f(z-x)g(x)dx.$$

Let X and Y be two independent continuous random variables. Define a new r.v. $Z = X + Y$. In this problem, you will show that the p.d.f. of Z is the convolution of the probability density functions of X and Y. Follow these steps:
(a) Show that $F_Z(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(t-x)dxdt$.
(b) Differentiate $F_Z(z)$ to obtain $f_Z(z)$.
[Hint: You will need to invoke the Fundamental Theorem of Calculus from Section 1.3.]

8.10 **Bear problem**
[Proposed by Weina Wang] You stand at a position $X \sim \text{Exp}(1)$ on the line. Your friend stands at position $Y \sim \text{Exp}(2)$. Assume that X and Y are independent. A bear comes from the left. The bear will eat the first person it comes to; however, if the distance between you and your friend is < 1, then the bear will eat both of you. What is the probability that you get eaten?

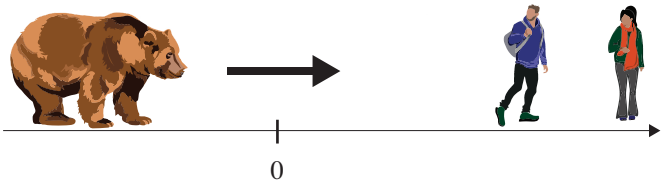


Figure 8.2 *Figure for Exercise 8.10.*

8.11 **Bayes of our lives**
The number of seasons in a television series is $N \sim \text{Geometric}(P)$. After each season, there is a fixed probability, P, that the series is canceled. However, the parameter P depends on the popularity of the series, so

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

we don't know what it is in general. For a new series, we assume that $P \sim \text{Uniform}(0, 1)$. A television series has been running for 37 seasons (and renewed for more). Derive the expected value of P, given this information, that is, derive $\mathbf{E}[P \mid N > 37]$.

8.12 **Density of choice**

Suppose that X and Y are continuous random variables and let

$$Z = \begin{cases} X & \text{w/prob } p \\ Y & \text{w/prob } 1-p \end{cases}.$$

(a) Derive the p.d.f. of Z in terms of the probability density functions of X and Y.

(b) The Double Exponential distribution is defined via the random variable W, where $W = ST$ and $T \sim \text{Exp}(\lambda)$ and S is a discrete r.v. with equal probability of being 1 or -1. Use part (a) to derive the p.d.f. of W.

8.13 **When the parameters of a distribution are random variables**

There are many situations where the parameters of a distribution are themselves random variables. For example, let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Uniform}(0, X)$. (a) What is $\mathbf{E}[Y]$? (b) What is $\mathbf{Var}(Y)$?

8.14 **Smallest interval**

A dart is thrown uniformly at random at the unit interval $[0, 1]$. The dart splits the interval into two segments, one to its right and one to its left. What is the expected length of the smaller segment?

8.15 **Smallest interval with two darts**

Two independent darts are thrown uniformly at random at the unit interval $[0, 1]$. The two darts naturally split the interval into three segments. Let S be the length of the smallest segment. What is $\mathbf{E}[S]$? [Hint: There are several ways to solve this problem. A good way to start is to derive the tail of S, and then integrate the tail to get $\mathbf{E}[S]$, as in Exercise 7.9. To get the tail, it may help to draw a 2D picture of where each of the darts is allowed to fall.]

8.16 **Different views on conditional expectation**

[Proposed by Misha Ivkov] Let $X \sim \text{Uniform}(0, 1)$ and $Y \sim \text{Uniform}(0, 1)$. Our goal is to understand

$$\mathbf{E}[X \mid X + Y = 1.5].$$

(a) Dong makes the realization that $X + Y = 1.5$ implies that $X = 1.5 - Y$. He then reasons that

$$\mathbf{E}[X \mid X + Y = 1.5] = \mathbf{E}[1.5 - Y].$$

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

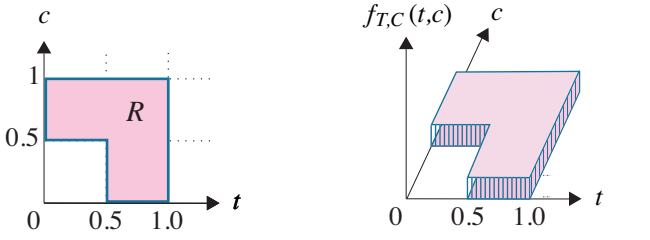
What's the result via Dong's approach? Is Dong right? Why or why not?

(b) Lisa suggests that one should first compute the conditional density function, $f_{X|X+Y=1.5}(x)$, using Definition 8.14 and then use that to get $\mathbf{E}[X \mid X+Y=1.5]$. Follow Lisa's approach to derive $\mathbf{E}[X \mid X+Y=1.5]$.

(c) Misha believes that pictures are the only way to prove things. Draw a 2D plot that allows you to understand $\mathbf{E}[X \mid X+Y=1.5]$.

8.17 **Hiring for tech**

At a popular tech company, candidates are rated on two axes: technical skills (T) and communication skills (C). The values of T and C can in theory be any real number from 0 (worst) to 1 (best). In practice, however, it never happens that a candidate gets a rating of less than 0.5 in both categories (that's just too harsh), so candidate scores actually fall within region R in Figure 8.3(a). Assume that candidate scores are uniformly distributed over region R, as shown in Figure 8.3(b).



(a) Region R (b) Joint density function $f_{T,C}(t,c)$

Figure 8.3 *For Exercise 8.17.*

(a) What is the joint density function $f_{T,C}(t,c)$?

(b) What is the marginal p.d.f. of T, that is, $f_T(t)$?

(c) What is $\mathbf{E}[C \mid T < 0.75]$? (Write out the full conditional density and then integrate it appropriately.)

8.18 **On the probability of a triangle**

Suppose we have an interval of length 1. We throw two darts at the interval independently and uniformly at random. The two darts divide our interval into three segments. We want to know the probability that the three segments form a triangle.

(a) Describe the criterion we need in order to achieve our goal.

(b) If the first dart lands at $x \in [0, \frac{1}{2}]$, what's the probability that the resulting segments give us a triangle?

(c) What is the probability that the three segments form a triangle?

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

8.19 **Relating laptop quality to lifetime**

[Proposed by Weina Wang] You have a laptop whose quality is represented by $Q \sim \text{Uniform}(1, 2)$, with a larger number representing higher quality. Laptops with higher quality have higher expected lifetimes. Let X be the lifetime of the laptop in years (assume this ranges from 0 to ∞). We are told that, given that $Q = q$, the lifetime of the laptop is $X \sim \text{Exp}\left(\frac{1}{q}\right)$ for $1 \leq q \leq 2$.

(a) Assume $1 \leq q \leq 2$. What is $f_{X|Q=q}(x)$? What is $\mathbf{E}[X \mid Q = q]$?

(b) What is the joint p.d.f. $f_{X,Q}(x, q)$?

(c) Suppose your laptop is still working after one year. What is the expected quality of your laptop given that fact?

[Note: In your final expression, you will get some integrals that you can't compute. Here are approximations to use: $\int_1^2 te^{-\frac{1}{t}}dt \approx 0.78$ and $\int_1^2 e^{-\frac{1}{t}}dt \approx 0.5$.]

8.20 **Gambling at the casino**

[Proposed by Weina Wang] Your friend Alice is visiting Las Vegas and takes X dollars to a casino, where $X \sim \text{Uniform}(10, 20)$. At the end of the day, she brings back Y dollars. Given that she takes $X = x$ dollars, the density of Y is

$$f_{Y|X=x}(y) = \begin{cases} -\frac{y}{2x^2} + \frac{1}{x} & \text{if } 0 \leq y \leq 2x \\ 0 & \text{otherwise} \end{cases}.$$

(a) What is Alice's expected return from gambling, that is, $\mathbf{E}[Y - X]$? The following steps will help:

(i) Derive $\mathbf{E}[X]$. This is uncomplicated.

(ii) Derive $\mathbf{E}[Y \mid X = x]$.

(iii) Use (ii) to derive $\mathbf{E}[Y]$.

(iv) Derive $\mathbf{E}[Y] - \mathbf{E}[X]$.

(b) Suppose you know that Alice wins at gambling. What is the expectation of the amount of money she takes to the casino? [Hint: The problem is asking for $\mathbf{E}[X \mid Y > X]$. To get this, you will need $f_{X|Y>X}(x)$.]

8.21 **Modeling expected disk delay**

The delay to read a single byte from a hard disk consists of two components: (1) *seek time* – this is the time needed to move the disk head to the desired track; and (2) *rotation time* – this is the time needed to rotate the disk head on the track to reach the desired byte (Figure 8.4). Suppose that bytes are uniformly distributed across the disk. Let T be a r.v. denoting the time to reach a single (randomly located) byte. Your goal is to compute $\mathbf{E}[T]$.

Assume that the disk has radius r (a constant). Assume that the tracks are infinitely thin. Assume the disk head starts from wherever the last byte was read. Assume that the time to traverse the full radius r is 15 ms. At all

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.

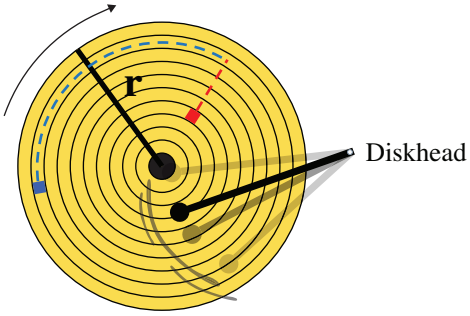


Figure 8.4 *For Exercise 8.21. A disk with radius r. Each circle represents a track. The red square (on the inner track) shows the byte most recently read. The blue square (on the outer track) shows the next byte requested. To read the blue byte, the disk head first seeks to the outer track of the blue byte and then waits for the disk to rotate to the correct byte.*

times, the disk rotates at 6,000 RPM (rotations per minute) in one direction only. Provide your final answer in ms. [Hint: Outer tracks hold more bytes than inner ones.]

8.22 **Hula hoop cutting**

You are holding a hula hoop of unit radius with your hand at its top (12 o'clock; Figure 8.5). In a moment, two points on the hoop will be selected uniformly at random, and the hoop will be cut at those points, splitting it into two arcs.

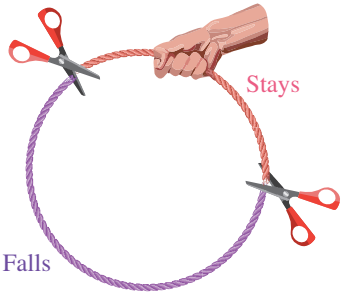


Figure 8.5 *For Exercise 8.22. The cuts create two arcs: a pink one that you're holding, and a purple one that falls.*

(a) Compute the expected value of the angular difference between the two arcs.

(b) When the cuts are made, one arc falls to the ground while the other one stays in your hand. What is the probability that you are holding the larger arc?

Mor Harchol-Balter. *Introduction to Probability for Computing,*
Cambridge University Press, 2024. Not for distribution.