6.1 Joint, marginal, and conditional distributions

Unit 6: Joint Distributions and Conditional Expectation

Adapted from Blitzstein-Hwang Chapters 7 and 9.

So far we have been focusing on the distribution of one random variable at a time, but very often we care about the relationship between multiple r.v.s in the same experiment. To give just a few examples:

- *Surveys*: When conducting a survey, we may ask multiple questions to each respondent in order to determine the relationship between, say, opinions on social issues and opinions on economic issues.
- *Medicine*: To evaluate the effectiveness of a treatment, we may take multiple measurements per patient; an ensemble of blood pressure, heart rate, and cholesterol readings can be more informative than any of these measurements considered separately.
- *Genetics*: To study the relationships between various genetic markers and a particular disease, if we only looked separately at distributions for each genetic marker, we could fail to learn about whether an *interaction* between markers is related to the disease.
- *Time series*: To study how something evolves over time, we can often make a series of measurements over time, and then study the series jointly. There are many applications of such series, such as global temperatures, stock prices, or national unemployment rates. The series of measurements considered jointly can help us deduce trends for the purpose of forecasting future measurements.

This unit considers *joint distributions*, also called *multivariate distributions*, which describe how multiple r.v.s interact with each other. We introduce multivariate analogs of the CDF, PMF, and PDF in order to provide a complete specification of the relationship between multiple r.v.s. After this groundwork is in place, we'll study a couple of famous named multivariate distributions, generalizing the Binomial and Normal distributions to higher dimensions.

The three key concepts for this section are *joint*, *marginal*, and *conditional* distributions. Recall that the distribution of a single r.v. X provides complete information about the probability of X falling into any subset of the real line. Analogously, the *joint* distribution of two r.v.s X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane. The *marginal* distribution of X is the individual distribution of X, ignoring the value of Y, and the *conditional* distribution of X given $Y = y$ is the updated distribution for X after observing $Y = y$. We'll look at these concepts in the discrete case first, then extend them to the continuous case.

Discrete

The most general description of the joint distribution of two r.v.s is the *joint CDF*, which applies to discrete and continuous r.v.s alike.

Definition 6.1.1 (Joint CDF).

The *joint CDF* of r.v.s X and Y is the function $F_{X,Y}$ given by

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y).$$

The joint CDF of n r.v.s is defined analogously.

Unfortunately, the joint CDF of discrete r.v.s is not a well-behaved function; as in the univariate case, it consists of jumps and flat regions. For this reason, with discrete r.v.s we usually work with the *joint PMF*, which also determines the joint distribution and is much easier to visualize.

Definition 6.1.2 (Joint PMF).

The *joint PMF* of discrete r.v.s X and Y is the function $p_{X,Y}$ given by

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

The joint PMF of n discrete r.v.s is defined analogously.

Just as univariate PMFs must be nonnegative and sum to 1, we require valid joint PMFs to be nonnegative and sum to 1, where the sum is taken over all possible values of X and Y:

$$\sum_x \sum_y P(X = x, Y = y) = 1.$$

Figure 6.1.3 shows a sketch of what the joint PMF of two discrete r.v.s could look like. The height of a vertical bar at (x, y) represents the probability $P(X = x, Y = y)$. For the joint PMF to be valid, the total height of the vertical bars must be 1.

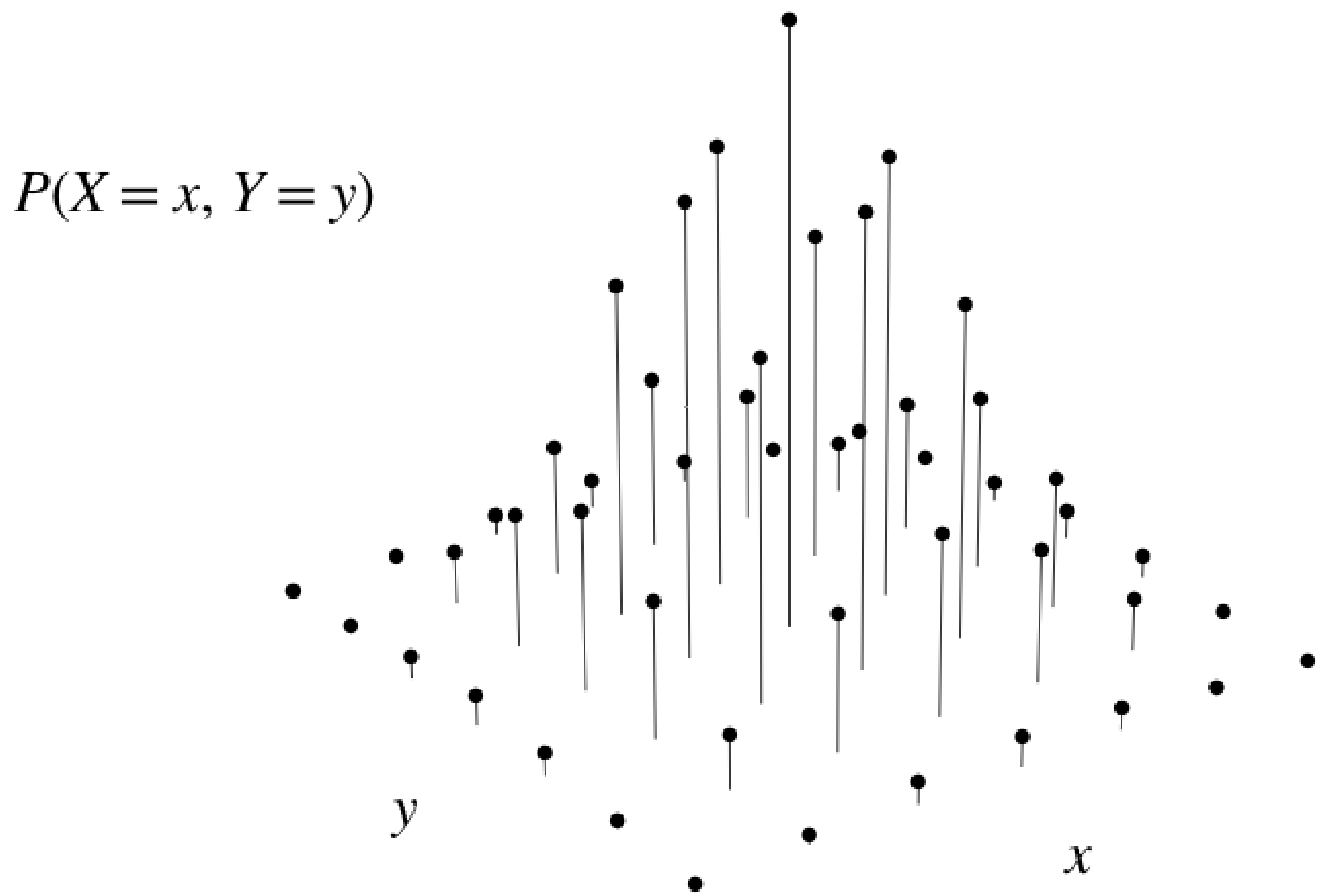

Figure 6.1.3: Joint PMF of discrete r.v.s X and Y.

View Larger Image
Image Description

From the joint distribution of X and Y, we can get the distribution of X alone by summing over the possible values of Y. This gives us the familiar PMF of X that we have seen in previous chapters. In the context of joint distributions, we will call it the *marginal* or unconditional distribution of X, to make it clear that we are referring to the distribution of X alone, without regard for the value of Y.

Definition 6.1.4 (Marginal PMF).

For discrete r.v.s X and Y, the *marginal PMF* of X is

$$P(X = x) = \sum_y P(X = x, Y = y).$$

> The marginal PMF of X *is* the PMF of X, viewing X individually rather than jointly with Y. The above equation follows from the axioms of probability (we are summing over disjoint cases). The operation of summing over the possible values of Y in order to convert the joint PMF into the marginal PMF of X is known as *marginalizing out* Y.

Similarly, the marginal PMF of Y is obtained by summing over all possible values of X. So given the joint PMF, we can marginalize out Y to get the PMF of X, or marginalize out X to get the PMF of Y. But if we only know the marginal PMFs of X and Y, there is no way to recover the joint PMF without further assumptions.

Now suppose that we observe the value of X and want to update our distribution of Y to reflect this information. Instead of using the marginal PMF $P(Y = y)$, which does not take into account any information about X, we should use a PMF that conditions on the event $X = x$, where x is the value we observed for X. This naturally leads us to consider *conditional PMFs*.

> **Definition 6.1.5 (Conditional PMF).**
>
> For discrete r.v.s X and Y, the *conditional PMF* of Y given $X = x$ is
>
> $$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}.$$
>
> This is viewed as a function of y for fixed x.

Figure 6.1.6 illustrates the definition of conditional PMF. To condition on the event $X = x$, we first take the joint PMF and focus in on the vertical bars where X takes on the value x; in the figure, these are shown in bold. All of the other vertical bars are irrelevant because they are inconsistent with the knowledge that $X = x$ occurred. Since the total height of the bold bars is the marginal probability $P(X = x)$, we then *renormalize* the conditional PMF by dividing by $P(X = x)$; this ensures that the conditional PMF will sum to 1. Therefore conditional PMFs *are* PMFs, just as conditional probabilities are probabilities. Notice that there is a different conditional PMF of Y for every possible value of X; Figure 6.1.6 highlights just one of these conditional PMFs.

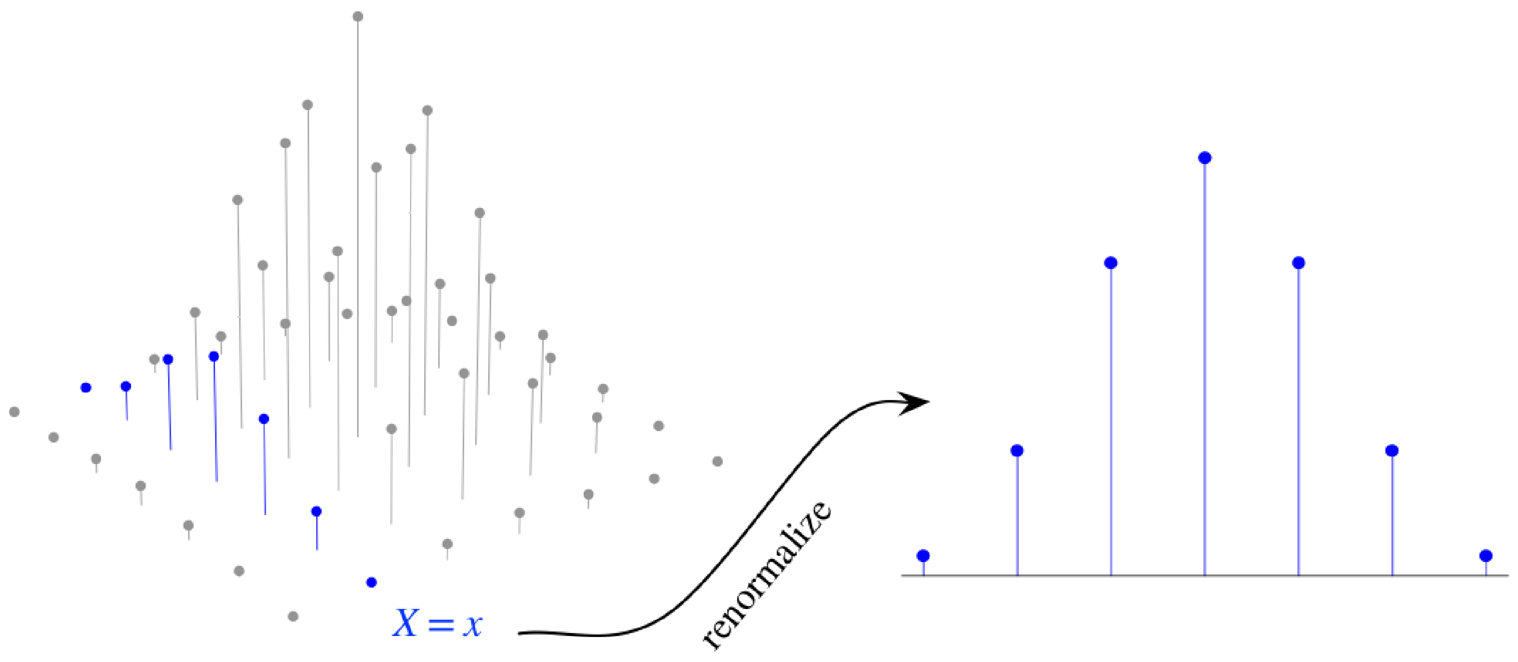

Figure 6.1.6: Conditional PMF of Y given $X = x$. The conditional PMF $P(Y = y|X = x)$ is obtained by renormalizing the column of the joint PMF that is compatible with the event $X = x$.

View Larger Image
Image Description

We can also relate the conditional distribution of Y given $X = x$ to that of X given $Y = y$, using Bayes' rule:

$$P(Y = y|X = x) = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}.$$

And using the law of total probability, we have another way of getting the marginal PMF: the marginal PMF of X is a weighted average of conditional PMFs $P(X = x|Y = y)$, where the weights are the probabilities $P(Y = y)$:

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y).$$

Armed with an understanding of joint, marginal, and conditional distributions, we can revisit the definition of independence.

Definition 6.1.7 (Independence of discrete r.v.s).

Random variables X and Y are *independent* if for all x and y,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

If X and Y are discrete, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all x and y, and it is also equivalent to the condition

$$P(Y = y|X = x) = P(Y = y)$$

for all y and all x such that $P(X = x) > 0$.

Using the terminology from this chapter, the definition says that for independent r.v.s, the *joint CDF* factors into the product of the *marginal CDFs*, or that the *joint PMF* factors into the product of the *marginal PMFs*. Remember that in general, the marginal distributions do *not* determine the joint distribution: this is the entire reason why we wanted to study joint distributions in the first place! But in the special case of independence, the marginal distributions are all we need in order to specify the joint distribution; we can get the joint PMF by multiplying the marginal PMFs.

Another way of looking at independence is that *all the conditional PMFs* are the same as the *marginal PMF*. In other words, starting with the marginal PMF of Y, no updating is necessary when we condition on $X = x$, regardless of what x is. There is no event purely involving X that influences our distribution of Y, and vice versa.

Example 6.1.8 (Chicken-egg).

Suppose a chicken lays a random number of eggs, N, where $N \sim \text{Pois}(\lambda)$. Each egg independently hatches with probability p and fails to hatch with probability $q = 1 - p$. Let X be the number of eggs that hatch and Y the number that do not hatch, so $X + Y = N$. What is the joint PMF of X and Y?

Solution

We seek the joint PMF $P(X = i, Y = j)$ for nonnegative integers i and j. Conditional on the total number of eggs N, the eggs are independent Bernoulli trials with probability of success p, so by the story of the Binomial, the conditional distributions of X and Y are $X|N = n \sim \text{Bin}(n, p)$ and $Y|N = n \sim \text{Bin}(n, q)$. Since our lives would be easier if only we knew the total number of eggs, let's use wishful thinking: condition on N and apply the law of total probability. This gives

$$P(X = i, Y = j) = \sum_{n=0}^{\infty} P(X = i, Y = j|N = n)P(N = n).$$

The sum is over all possible values of n, holding i and j fixed. But unless $n = i + j$, it is impossible for X to equal i and Y to equal j. For example, the only way there can be 5 hatched eggs and 6 unhatched eggs is if there are 11 eggs in total. So

$$P(X = i, Y = j|N = n) = 0$$

unless $n = i + j$, which means all other terms in the sum can be dropped:

$$P(X = i, Y = j) = P(X = i, Y = j|N = i + j)P(N = i + j).$$

Conditional on $N = i + j$, the events $X = i$ and $Y = j$ are exactly the same event, so keeping both is redundant. We'll keep $X = i$; the rest is a matter of plugging in the Binomial PMF to get $P(X = i|N = i + j)$ and the Poisson PMF to get $P(N = i + j)$. Thus,

$$\begin{aligned} P(X=i,Y=j) &= P(X=i|N=i+j)P(N=i+j) \\ &= \binom{i+j}{i} p^i q^j \cdot \frac{e^{-\lambda}\lambda^{i+j}}{(i+j)!} \\ &= \frac{e^{-\lambda p}(\lambda p)^i}{i!} \cdot \frac{e^{-\lambda q}(\lambda q)^j}{j!}. \end{aligned}$$

The joint PMF factors into the product of the $\mathrm{Pois}(\lambda p)$ PMF (as a function of i) and the $\mathrm{Pois}(\lambda q)$ PMF (as a function of j). This tells us two elegant facts: (1) X and Y are independent, since their joint PMF is the product of their marginal PMFs, and (2) $X \sim \mathrm{Pois}(\lambda p)$ and $Y \sim \mathrm{Pois}(\lambda q)$.

At first it may seem deeply counterintuitive that X is independent of Y. Doesn't knowing that a lot of eggs hatched mean that there are probably not so many that didn't hatch? For a *fixed* number of eggs, this independence would be impossible: knowing the number of hatched eggs would perfectly determine the number of unhatched eggs. But in this example, the number of eggs is *random*, following a Poisson distribution, and this happens to be the right kind of randomness to make X and Y unconditionally independent (this is a special property of the Poisson).

Continuous

Once we have a handle on discrete joint distributions, it isn't much harder to consider continuous joint distributions. We simply make the now-familiar substitutions of integrals for sums and PDFs for PMFs, remembering that the probability of any individual point is now 0.

Formally, in order for X and Y to have a continuous joint distribution, we require that the joint CDF

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

be differentiable with respect to x and y. The partial derivative with respect to x and y is called the *joint PDF*. The joint PDF determines the joint distribution, as does the joint CDF.

Definition 6.1.9 (Joint PDF).

If X and Y are continuous with joint CDF $F_{X,Y}$, their *joint PDF* is the derivative of the joint CDF with respect to x and y:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

We require valid joint PDFs to be nonnegative and integrate to 1:

$$f_{X,Y}(x,y) \ge 0, \text{ and } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1.$$

In the univariate case, the PDF was the function we integrated to get the probability of an interval. Similarly, the joint PDF of two r.v.s is the function we integrate to get the probability of a two-dimensional region. For example,

$$P(X < 3,\ 1 < Y < 4) = \int_1^4 \int_{-\infty}^3 f_{X,Y}(x,y)dxdy.$$

For a general set $A \subseteq \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint\limits_A f_{X,Y}(x,y)dxdy.$$

Figure 6.1.10 shows a sketch of what a joint PDF of two r.v.s could look like. As usual with continuous r.v.s, we need to keep in mind that the height of the surface $f_{X,Y}(x,y)$ at a single point does *not* represent a probability. The probability of any specific point in the plane is 0; furthermore, now that we've gone up a dimension, the probability of any line or curve in the plane is also 0. The only way we can get nonzero probability is by integrating over a region of *positive area* in the xy-plane.

When we integrate the joint PDF over an area A, what we are calculating is the volume under the surface of the joint PDF and above A. Thus, probability is represented by *volume under the joint PDF*. The total volume under a valid joint PDF is 1.

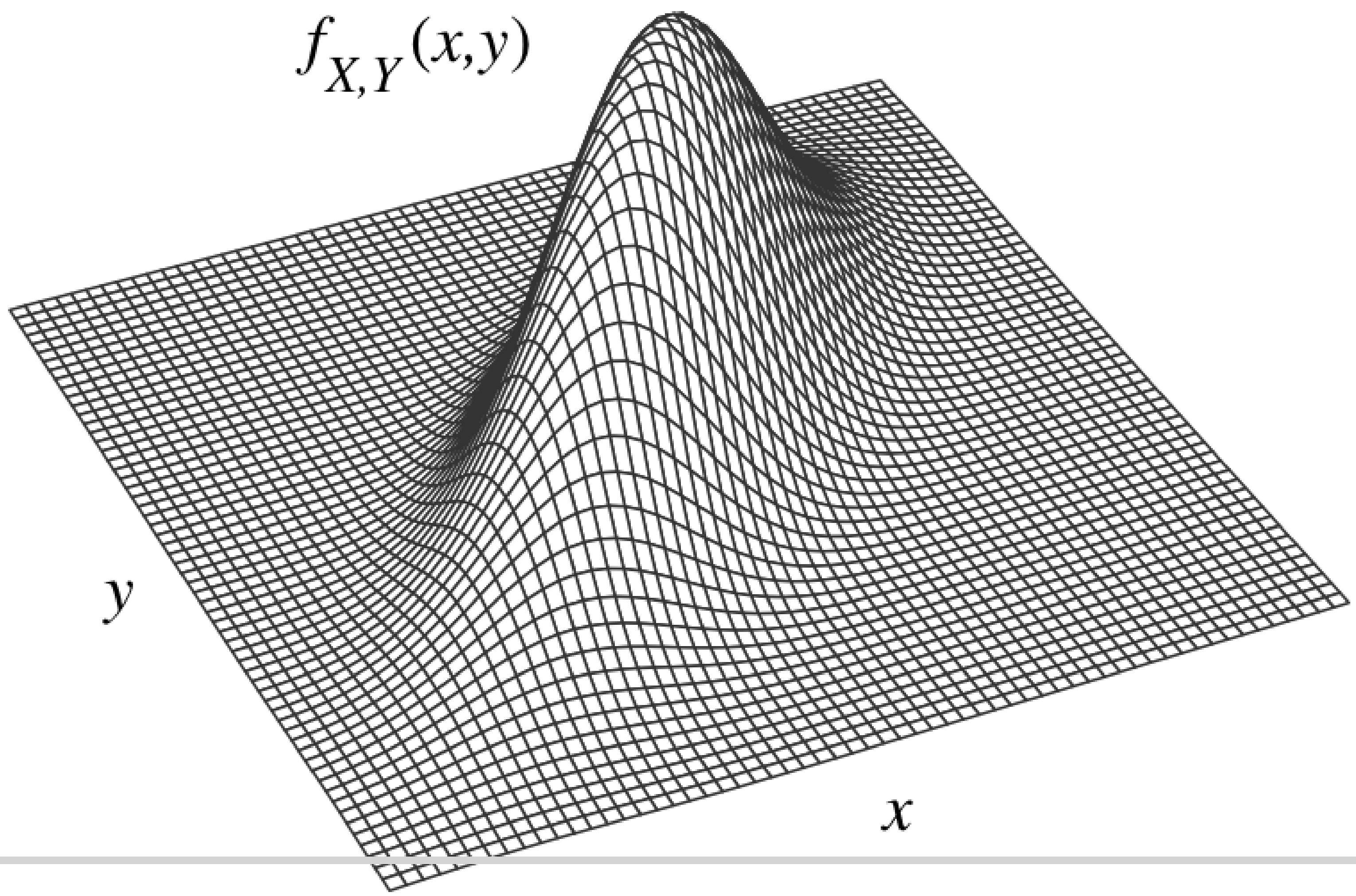

Figure 6.1.10: Joint PDF of continuous r.v.s X and Y.

View Larger Image
Image Description

In the discrete case, we get the marginal PMF of X by summing over all possible values of Y in the joint PMF. In the continuous case, we get the *marginal PDF* of X by integrating over all possible values of Y in the joint PDF.

Definition 6.1.11 (Marginal PDF).

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the *marginal PDF* of X is

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

This *is* the PDF of X, viewing X individually rather than jointly with Y.

To simplify notation, we have mainly been looking at the joint distribution of two r.v.s rather than n r.v.s, but marginalization works analogously with any number of variables. For example, if we have the joint PDF of X, Y, Z, W but want the joint PDF of X, W, we just have to integrate over all possible values of Y and Z:

$$f_{X,W}(x,w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z,W}(x,y,z,w)dydz.$$

Conceptually this is very easy---just integrate over the unwanted variables to get the joint PDF of the wanted variables---but computing the integral may or may not be difficult. Returning to the case of the joint distribution of two r.v.s X and Y, let's consider how to update our distribution for Y after observing the value of X, using the *conditional PDF*.

Definition 6.1.12 (Conditional PDF).

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the *conditional PDF* of Y given $X = x$ is

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

This is considered as a function of y for fixed x.

Notation 6.1.13

The subscripts that we place on all the f's are just to remind us that we have three different functions on our plate. We could just as well write $g(y|x) = f(x, y)/h(x)$, where f is the joint PDF, h is the marginal PDF of X, and g is the conditional PDF of Y given $X = x$, but that makes it more difficult to remember which letter stands for which function. Figure 6.1.14 illustrates the definition of conditional PDF. We take a vertical slice of the joint PDF corresponding to the observed value of X; since the total area under this slice is $f_X(x)$, we then divide by $f_X(x)$ to ensure that the conditional PDF will have an area of 1. Therefore conditional PDFs satisfy the properties of a valid PDF.

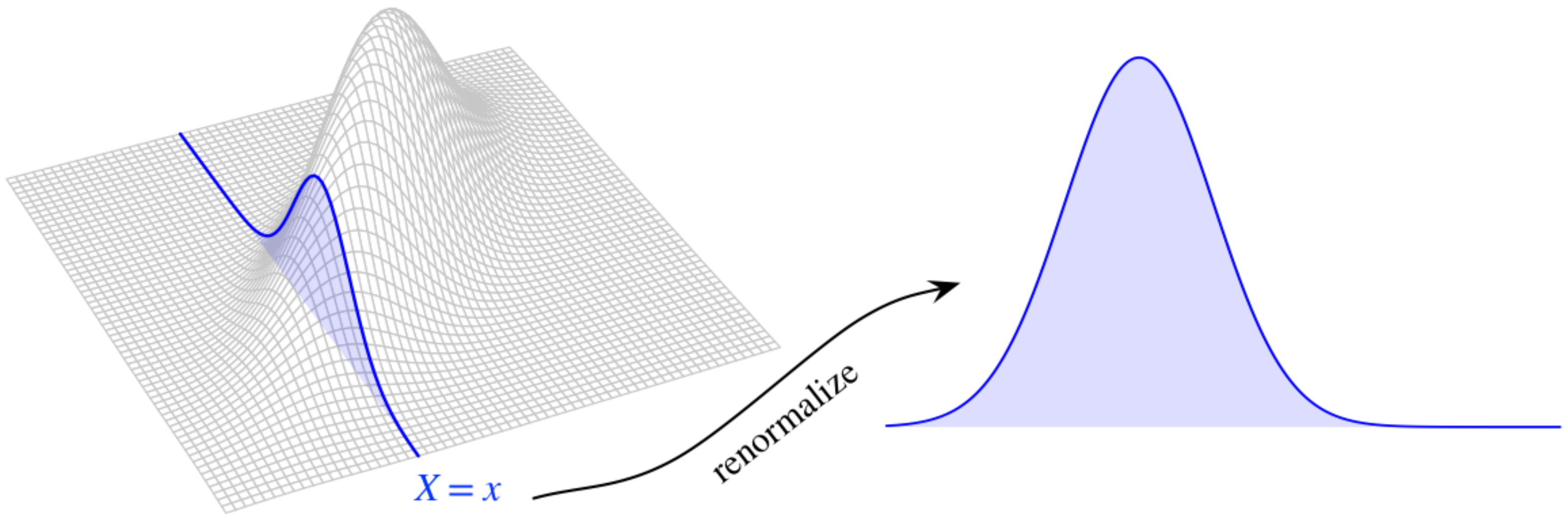

Figure 6.1.14: Conditional PDF of Y given $X = x$. The conditional PDF $f_{Y|X}(y|x)$ is obtained by renormalizing the slice of the joint PDF at the fixed value x.

View Larger Image
Image Description

Note that we can recover the joint PDF $f_{X,Y}$ if we have the conditional PDF $f_{Y|X}$ and the corresponding marginal f_X:

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

Similarly, we can recover the joint PDF if we have $f_{X|Y}$ and f_Y:

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

This allows us to develop continuous analogs of Bayes' rule and LOTP. These formulas still hold in the continuous case, replacing probabilities with probability density functions.

Theorem 6.1.15 (Continuous form of Bayes' rule and LOTP).

For continuous r.v.s X and Y,

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

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

Proof

By definition of conditional PDFs, we have

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

The continuous version of Bayes' rule follows immediately from dividing by $f_X(x)$. The continuous version of LOTP follows immediately from integrating with respect to y:

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

We now have versions of Bayes' rule and LOTP for two discrete r.v.s and for two continuous r.v.s. Better yet, there are also versions when we have one discrete r.v. and one continuous r.v. After understanding of the discrete versions, it is easy to remember and use the other versions since they are analogous, replacing probabilities by PDFs when appropriate. Here are the four versions of Bayes' rule, summarized in a table.

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x) = \frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x) = \frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x) = \frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x) = \frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

And here are the four versions of LOTP, summarized in a table.

	Y **discrete**	Y **continuous**
X **discrete**	$P(X=x) = \sum_y P(X=x\|Y=y)P(Y=y)$	$P(X=x) = \int_{-\infty}^{\infty} P(X=x\|Y=y) f_Y(y) dy$
X **continuous**	$f_X(x) = \sum_y f_X(x\|Y=y)P(Y=y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X\|Y}(x\|y) f_Y(y) dy$

Finally, let's discuss the definition of independence for continuous r.v.s; then we'll turn to concrete examples. As in the discrete case, we can view independence of continuous r.v.s in two ways. One is that the joint CDF factors into the product of the marginal CDFs, or the joint PDF factors into the product of the marginal PDFs. The other is that the conditional PDF of Y given $X=x$ is the same as the marginal PDF of Y, so conditioning on X provides no information about Y.

Definition 6.1.16 (Independence of continuous r.v.s).

Random variables X and Y are *independent* if for all x and y,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

If X and Y are continuous with joint PDF $f_{X,Y}$, this is equivalent to the condition

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

for all x and y, and it is also equivalent to the condition

$$f_{Y|X}(y|x) = f_Y(y)$$

for all y and all x such that $f_X(x) > 0$.

A simple case of a continuous joint distribution is when the joint PDF is constant over some region in the plane. In the following example, we'll compare a joint PDF that is constant on a square to a joint PDF that is constant on a disk.

Example 6.1.17 (Uniform on a region in the plane).

Let (X,Y) be a completely random point in the square $\{(x,y) : x, y \in [0,1]\}$, in the sense that the joint PDF of X and Y is constant over the square and 0 outside of it:

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

The constant 1 is chosen so that the joint PDF will integrate to 1. This distribution is called the *Uniform distribution* on the square.

Intuitively, it makes sense that X and Y should be $\text{Unif}(0,1)$ marginally. We can check this by computing

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \int_0^1 1dy = 1,$$

and similarly for f_Y. Furthermore, X and Y are independent, since the joint PDF factors into the product of the marginal PDFs (this just reduces to $1 = 1 \cdot 1$, but it's important to note that the value of X does not constrain the possible values of Y). So the conditional distribution of Y given $X = x$ is $\text{Unif}(0,1)$, regardless of x. Now let (X,Y) be a completely random point in the unit disk $\{(x,y) : x^2 + y^2 \leq 1\}$, with joint PDF

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

Again, the constant $1/\pi$ is chosen to make the joint PDF integrate to 1; the value follows from the fact that the integral of 1 over some region in the plane is the area of that region.

Note that X and Y are *not* independent, since in general, knowing the value of X constrains the possible values of Y: larger values of $|X|$ restrict Y to be in a smaller range. To see from the definition that X and Y are not independent, note that, for example, $f_{X,Y}(0.9, 0.9) = 0$ since $(0.9, 0.9)$ is not in the unit disk, but $f_X(0.9)f_Y(0.9) \neq 0$ since 0.9 is in the supports of both X and Y.

The marginal distribution of X is now

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi}\sqrt{1-x^2}, \quad -1 \leq x \leq 1.$$

By symmetry, $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$. Note that the marginal distributions of X and Y are *not* Uniform on $[-1, 1]$; rather, X and Y are more likely to fall near 0 than near ± 1.

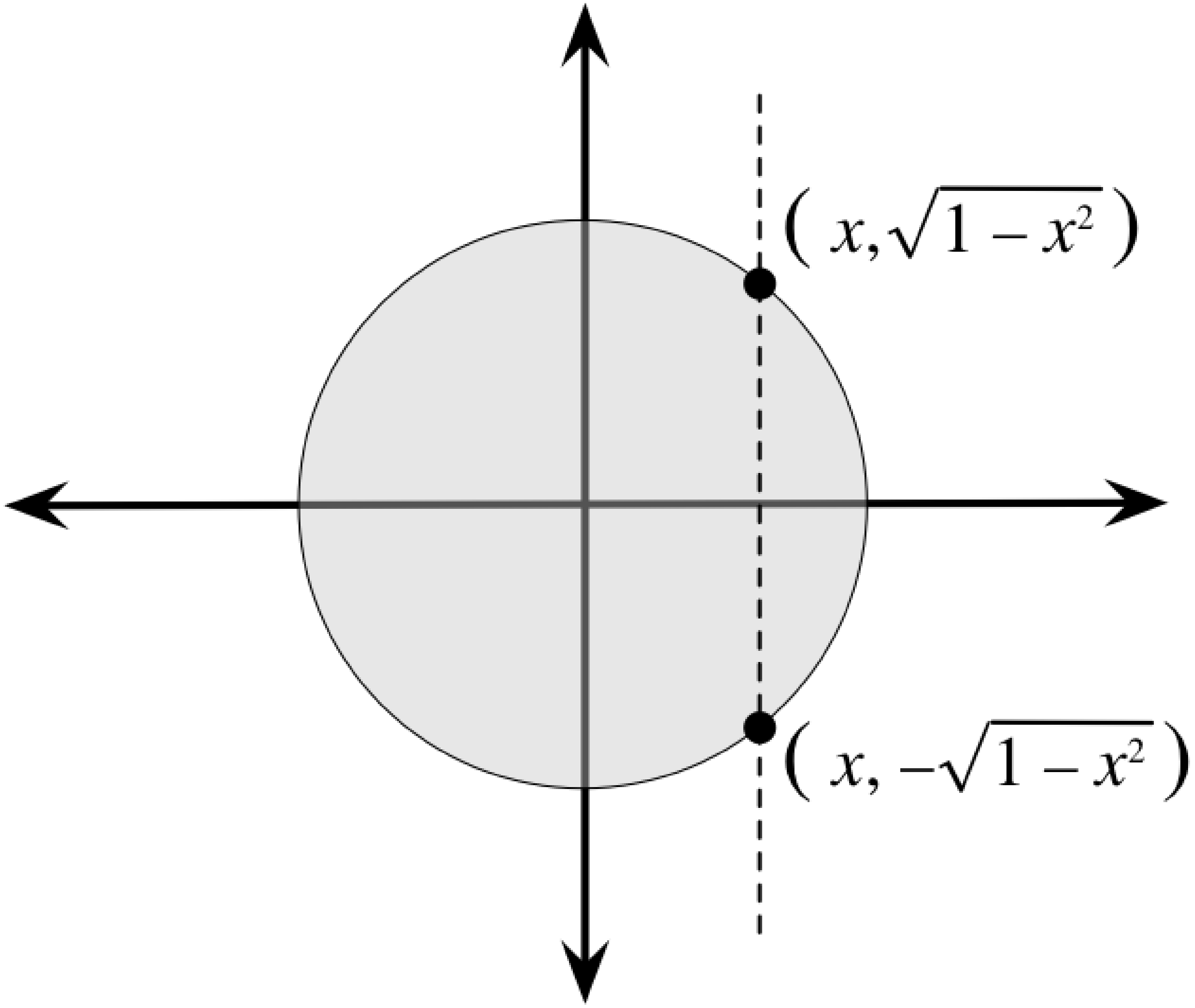

Figure 6.1.18: Bird's-eye view of the Uniform joint PDF on the unit disk. Conditional on $X = x$, Y is restricted to the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$.

View Larger Image
Image Description

Suppose we observe $X = x$. As illustrated in Figure 6.1.18, this constrains Y to lie in the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. Specifically, the conditional distribution of Y given $X = x$ is

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

for $-\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$, and 0 otherwise. This conditional PDF is constant as a function of y, which tells us that the conditional distribution of Y is Uniform on the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. The fact that the conditional PDF is not free of x confirms the fact that X and Y are not independent.

In general, for a region R in the plane, the Uniform distribution on R is defined to have joint PDF that is constant inside R and 0 outside R. As above, the constant is the reciprocal of the area of R. If R is the rectangle $\{(x,y) : a \le x \le b, c \le y \le d\}$, then X and Y will be independent; unlike for a disk, the vertical slices of a rectangle all look the same. A way to generate a random (X, Y) would then be to draw $X \sim \mathrm{Unif}(a,b)$ and $Y \sim \mathrm{Unif}(c,d)$ independently. But for any region where the value of X constrains the possible values of Y or vice versa, X and Y will *not* be independent.

As another example of working with joint PDFs, let's consider a question that comes up often when dealing with Exponentials of different rates.

Example 6.1.19 (Comparing Exponentials of different rates).

Let $T_1 \sim \mathrm{Expo}(\lambda_1)$ and $T_2 \sim \mathrm{Expo}(\lambda_2)$ be independent. Find $P(T_1 < T_2)$. For example, T_1 could be the lifetime of a refrigerator and T_2 could be the lifetime of a stove (if we are willing to assume Exponential distributions for these), and then $P(T_1 < T_2)$ is the probability that the refrigerator fails before the stove. We know from Chapter 5 that $\min(T_1, T_2) \sim \mathrm{Expo}(\lambda_1 + \lambda_2)$, which tells us about *when* the first appliance failure will occur, but we also may want to know about *which* appliance will fail first.

Solution

We just need to integrate the joint PDF of T_1 and T_2 over the appropriate region, which is all (t_1, t_2) with $t_1 > 0, t_2 > 0$, and $t_1 < t_2$. This yields

$$\begin{aligned}
P(T_1 < T_2) &= \int_0^\infty \int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2}\, dt_1\, dt_2 \\
&= \int_0^\infty \left(\int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1}\, dt_1 \right) \lambda_2 e^{-\lambda_2 t_2}\, dt_2 \\
&= \int_0^\infty (1 - e^{-\lambda_1 t_2}) \lambda_2 e^{-\lambda_2 t_2}\, dt_2 \\
&= 1 - \int_0^\infty \lambda_2 e^{-(\lambda_1 + \lambda_2) t_2}\, dt_2 \\
&= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

To get from the first line to the second, we took out $\lambda_2 e^{-\lambda_2 t_2}$ from the inner integral since it is treated as a constant when integrating with respect to t_1.

This result makes sense intuitively if we interpret λ_1 and λ_2 as rates. For example, if refrigerators have twice the failure rate of stoves, then it says that the odds are 2 to 1 in favor of the refrigerator failing first. As a simple check, note that the answer reduces to $1/2$ when $\lambda_1 = \lambda_2$ (in that case, we already knew $P(T_1 < T_2) = 1/2$ by symmetry).