4 Multivariate Distributions

4.1 Introduction

Experimenters will often measure more than one quantity, and are often interested in the distribution of all observed quantities. As as example, a naturalist measures the height and weight of chimpanzees. They might be interested in the distribution of height-weight pairs. Since the distribution involves two quantities, we call it a *bivariate distribution*. One question might be whether or not these two quantities are independent. (We suspect they are not, since a reasonable assumption is that taller chimpanzees tend to weigh more.)

Another important application of multivariate distributions is statistical sampling. Suppose $Y_1, Y_2, \ldots, Y_n$ are n successive trials of an experiment. Statisticians are interested in the distribution of $(Y_1, Y_2, \ldots, Y_n)$, and can use information about this distribution to infer characteristics of the experiment or the population from which the experiment sampled.

In this section, we will primarily be interested in bivariate distributions, the probability distribution of two random variables. As before, we start with the discrete case and then consider the continuous case.

4.2 Distribution of Two Discrete Random Variables

First, let's define the joint probability distribution for a pair of discrete random variables.

Joint probability distribution, discrete case

Let Y_1 and Y_2 be two discrete random variables. Then the *joint distribution* of Y_1 and Y_2 is given by the function of two variables:

$$p(y_1, y_2) = \mathbb{P}(Y_1 = y_1, Y_2 = y_2) \qquad \text{for all possible pairs } (y_1, y_2)$$

Sometimes this is called the *joint probability mass function* (joint pmf). Note that $\mathbb{P}(Y_1 = y_1, Y_2 = y_2)$ means $\mathbb{P}(Y_1 = y_1 \cap Y_2 = y_2)$. This is standard notation for expressing the joint probability of two random variables.

We have already encountered one example of a bivariate distribution. Recall the distribution of the rolls of two standard, six-sided dice which we discussed several times in the section on discrete random variables. Let X_1 be the roll of the first die and X_2 the roll of the second die. Then since we have a discrete uniform distribution, the joint distribution of X_1 and X_2 is given by:

$$p(x_1, x_2) = \frac{1}{36} \qquad x_1, x_2 = 1, 2, 3, 4, 5, 6$$

Just as in the case for a single discrete random variable, for a joint distribution of two discrete random variable, all the possible probabilities are nonnegative and they sum to 1.

Let Y_1 and Y_2 be discrete random variables with joint probability distribution $p(y_1, y_2)$. Then

$$0 \leq p(y_1, y_2) \leq 1 \text{ for all } y_1, y_2$$
$$\sum_{\text{all } (y_1, y_2} p(y_1, y_2) = 1$$

where the sum is taken over all possible pairs (y_1, y_2).

Just as in the case of a single discrete random variable, we can construct a valid joint probability distribution of two discrete random variables by assigning probabilities that add up to 1. Let Y_1 be a discrete random variable with m possible output values, and Y_2 a discrete random variable with n possible output values. Then there are mn possible joint outputs of the pair of random variables. Each of the outputs is an ordered pair of the form (y_1, y_2). If we make an mxn table and assign probabilities to each of the mn possible joint outputs so they add up to 1, we have constructed a joint probability distribution for the two discrete random variables Y_1 and Y_2.

Let's consider an example.

Example. Imagine we surveyed Brown undergraduates and asked them two questions.

1. Do you have an exam this week?

2. How many cups of coffee have you drunk today?

Let X_1 be the discrete random variable with values {`yes`, `no`} indicating whether or not a student has an exam this week. Let X_2 be the number of cups of coffee a student has drunk today. For simplicity, we will let X_2 take only the values {0, 1, 2} (Whether or not this is a realistic simplification is beyond the scope of this course!)

We can display the joint probability distribution for the pair (X_1, X_2) in a 2 x 3 table. There are 6 possible values for the pair (x_1, x_2). We can choose any probabilities for the six pairs as long as they sum to 1. One possible choice is shown in the table below.

		X_2		
		0	1	2
X_1	`yes`	2/20	3/20	3/20
	`no`	6/20	4/20	2/20

You can verify that the six probabilities do indeed sum to 1.

4.2.1 Marginal distribution

Consider again a joint distribution (Y_1, Y_2) of two discrete random variables with pmf $p(y_1, y_2)$. (You can think of the exam-coffee example above). Y_1 and Y_2 are themselves discrete random variables. What are their distributions?

Suppose we wish to find the distribution for Y_1 by itself. We call this the *marginal distribution* of Y_1. Essentially what we want to do is take Y_2 out of the picture entirely. How do we do that? All we have to do is sum over all the possible values of Y_2! The probability that $Y_1 = y_1$ is the sum of $p(y_1, y_2)$ over all possible values y_2 that Y_2 can take; this is the marginal distribution of Y_1, and is written $p_1(y_1)$. Similarly, we can sum over all possible values Y_1 to get $p_2(y_2)$, the marginal distribution of Y_2. This is summarized below.

Marginal distribution, discrete random variables

Let Y_1 and Y_2 be discrete random variables with joint probability distribution $p(y_1, y_2)$. Then the marginal distribution of Y_1 is given by:

$$p_1(y_1) = \sum_{\text{all } y_2} p(y_1, y_2)$$

and the marginal distribution of Y_2 is given by.

$$p_2(y_2) = \sum_{\text{all } y_1} p(y_1, y_2)$$

In both cases, we just sum the joint distribution over all the possibilities of the other random variable.

Let's return to our example above.

Example. In the exam-coffee example above, compute the marginal distributions for X_1 and X_2.

To find the marginal distributions for each variable, we sum over all the possibilities of the other variable. If the joint distribution is presented in a two-dimensional table, this is easy. To find the marginal distribution of X_2, we sum the values in each column. The bottom row, which we will label "total", is the marginal distribution of X_2. Similarly, we can find the marginal distribution for X_1 by summing each row. The rightmost column, also labeled "total", is the marginal distribution for X_2. In fact, the marginal distribution is called "marginal" because its values lie in the margins of the joint distribution table.

		X_2			
		0	1	2	total
X_1	`yes`	2/20	3/20	3/20	8/20
	`no`	6/20	4/20	2/20	12/20
	total	8/20	7/20	5/20	

You can check that the two marginal distributions sum to 1 and are thus valid probability distributions for discrete random variables.

4.2.2 Conditional distribution

Suppose again we have a joint distribution (Y_1, Y_2) of two discrete random variables with joint pmf $p(y_1, y_2)$. Another question we might ask is what is the distribution of Y_1 given that $Y_2 = y_2$. In other words, what is the conditional distribution of Y_1 given that Y_2 takes a specific value.

Let's look once more a the exam-coffee example to see how we can do this.

Example. In the exam-coffee example above,what is the distribution of the number of cups of coffee drunk today (X_2) given that a student has a midterm this week ($X_1 = \texttt{yes}$)?

To do this, we look at the first row of the table, which corresponds to $X_1 = \texttt{yes}$. This is not a valid probability mass function, because the elements do not sum to 1. But we can fix that! All we have to do is divide by the marginal probability $p_1(\texttt{yes}) = \mathbb{P}(X_1 = \texttt{yes})$, which is conveniently located just to the right in the "total" column. If we do that, we get the conditional probability for X_2 given $X_1 = \texttt{yes}$, which we can write as $p(y_2|\texttt{yes})$ or $p(y_2|Y_1 = \texttt{yes})$:

y	$p(y\|\text{midterm})$
0	2/8
1	3/8
2	3/8

Now that we have seen an example, we will give the formal definition of the conditional distribution of two discrete random variables.

Conditional distribution, discrete random variables

Let Y_1 and Y_2 be discrete random variables with joint probability distribution $p(y_1, y_2)$. Let $p_2(y_2)$ be the marginal distribution of Y_2. Then the conditional distribution of Y_1 given $Y_2 = y_2$ is:

$$p(y_1|y_2) = \mathbb{P}(Y_1 = y_1|Y_2 = y_2) = \frac{\mathbb{P}(Y_1 = y_1, Y_2 = y_2)}{\mathbb{P}(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

where $p_2(y_2) > 0$. In words, the conditional distribution is the joint distribution divided by the marginal distribution. We can similarly define the conditional distribution of Y_2 given $Y_1 = y_1$.

4.2.3 Independence

The final question to settle is independence. Roughly speaking, two random variables are independent of if the probabilities of each one are not affected by the value of the other one. The following will serve as our definition for independence of two discrete random variables.

Independence of discrete random variables

Let Y_1 and Y_2 be discrete random variables with joint probability distribution $p(y_1, y_2)$. Let $p_1(y_1)$ and $p_2(y_2)$ be the marginal distributions of Y_1 and Y_2. Then Y_1 and Y_2 are independent if

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \qquad \text{for all } y_1, y_2$$

In other words, two random variables are independent if their joint distribution is the product of the two marginal distributions.

In the exam-coffee example above, using just about any pair of y_1 and y_2, we can show that Y_1 and Y_2 and not independent. Did we really expect them to be independent?

4.3 Distribution of Two Discrete Continuous Variables

We will essentially repeat the same discussion for a pair of continuous random variables. Since working with continuous random variables requires integration, this will require integration in two dimensions, i.e. multivariable calculus. Since it is likely that many of you have not taken multivariable calculus, all multivariable techniques will be taught as they are needed.

4.3.1 Joint Probability Density

Recall that in the discrete case, a probability distribution was described by a probability density function (pdf). For the joint distribution of two continuous random variable, we have a joint density function, which is the continuous analogue of the joint distribution function in the discrete case.

Joint probability density, continuous case

Let Y_1 and Y_2 be two continuous random variables. Then the *joint density* of Y_1 and Y_2 is a function of two variables $f(y_1, y_2)$ with the properties that:

1. $f(y_1, y_2) \geq 1$ for all y_1, y_2.
2. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)dy_1dy_2 = 1$$

In other words, $f(y_1, y_2)$ is nonnegative and integrates to 1.

In the bivariate case, instead of finding the probability that a single variable lies in an interval $[a, b]$, we find the probability that a pair of random variables lies within a region of the plane. To do this, we integrate the joint density over that region.

Probability of an event, continuous bivariate case

Let Y_1 and Y_2 be two continuous random variables with joint density $f(y_1, y_2)$. Let A be a region of the plane. Then

$$\mathbb{P}((Y_1, Y_2) \in A) = \int\int_A f(y_1, y_2) dy_1 dy_2$$

This notation may not be precise. Don't worry about it for now, we will do plenty of examples. Just remember the key idea: to find the probability that a pair of random variables lie in a region, integrate the joint density over that region.

The extra complication here is the double integral. Whereas a single integral is defined on a closed interval, a double integral is defined on a two-dimensional region of the plane. The key to success for any double integral problem is to draw the region of integration before doing anything else. Since this is so important, I will repeat it: always draw the region of integration!

We will learn how to handle this through a series of examples. We will keep coming back to this first example throughout this section.

Example. Let X and Y be random variables with joint distribution function $f(x, y)$ given by:

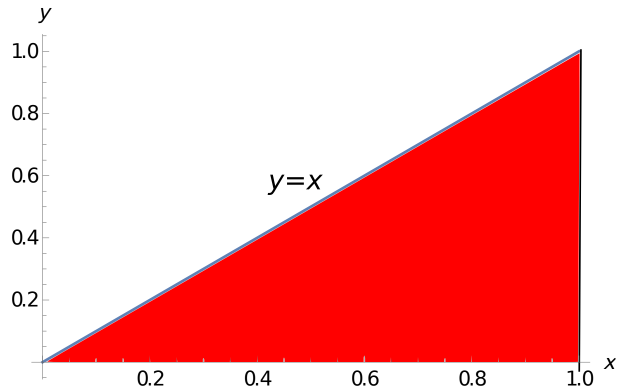
$$f(x, y) = \begin{cases} kxy & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the value of k such that $f(x, y)$ is a valid joint probability density function.

   As defined above (and similar to the one-dimensional case), for a joint probability density function to be valid, its integral must be 1 over the region of integration, i.e.

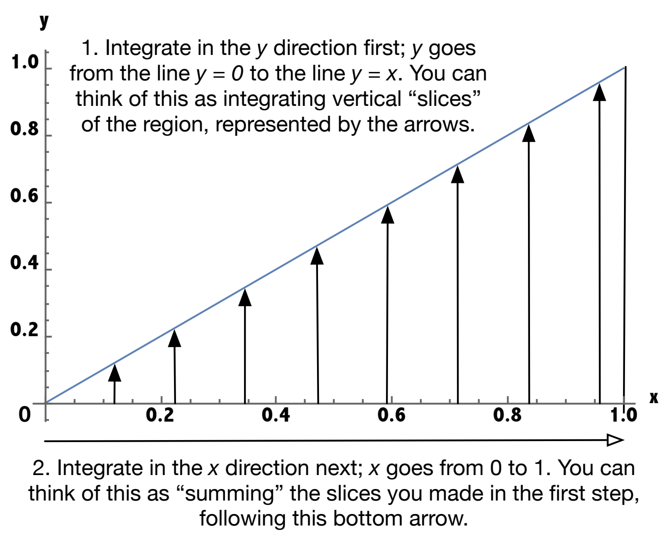
   $$\int\int f(x, y) dx\, dy = 1$$

   The bounds of the density function are the following: $y \ge 0, y \le x$, and $x \le 1$. This describes the triangular region illustrated below.



Whenever we have a double integral, we have two choices when we do our integration. We can integrate in x direction first, or we can integrate in the y direction first. Both ways give the same answer, but sometimes one is easier than the other. We will show both of them here.

Let's start by integrating in the y direction first.

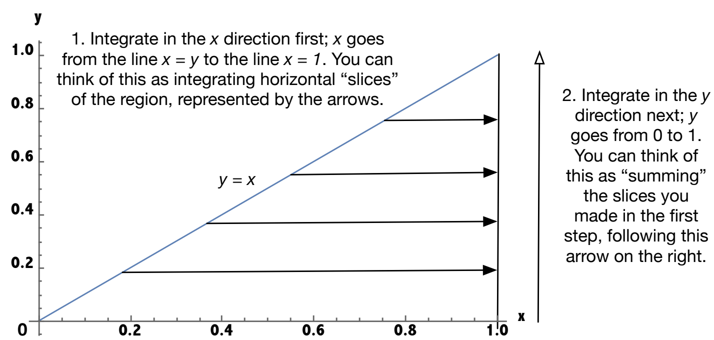


Following the directions on the picture above, we integrate y first; y goes from 0 to the diagonal line $y = x$, so those are the limits for the integral with respect to y (inner integral). Note that the upper limit is a function of x. Each integral in y is a vertical "slice" of our region. Then we integrate with respect to x. Imagining this as stacking our vertical slices side-by-side in the horizontal direction, x goes from 0 to 1, so those are the limits for the integral with respect to x (outer integral). Putting this together, we have:

$$\begin{aligned}1 &= \int_0^1 \int_0^x kxy\, dydx \\ &= k\int_0^1 x\frac{y^2}{2}\Big|_0^x dx \\ &= \frac{k}{2}\int_0^1 x^3 dx \\ &= \frac{k}{2}\frac{x^4}{4}\Big|_0^1 = \frac{k}{8}\end{aligned}$$

Multiplying by 8 gives us $k = 8$.

Let's do the integral the other way and verify to see if we get the same result. This time we integrate in the x direction first.



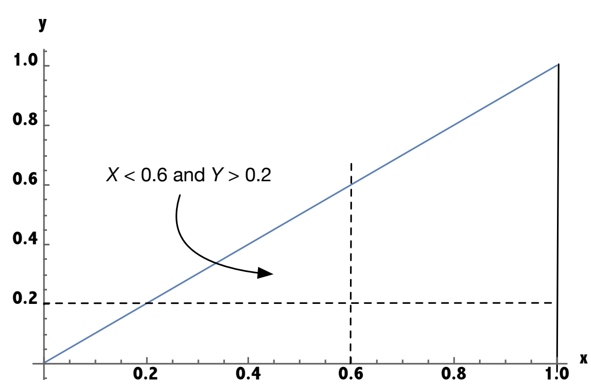
Once again following the directions on the picture above, we integrate x first; x starts at the diagonal line $y = x$ and goes to $x = 1$. The diagonal line has the equation $y = x$, which we solve for x to get $x = y$. Thus the lower limit is the function $x = y$. The upper limit is 1, so the limits of the integral with respect to x (inner integral) are y and 1. Each integral in x is a horizontal "slice" of our region. Then we integrate with respect to y. Imagining this as stacking our horizontal slices one on top of the other in the vertical direction, y goes from 0 to 1, so those are the limits for the integral with respect to y (outer integral). Putting this together, we have:

$$\begin{aligned} 1 &= \int_0^1 \int_y^1 kxy \, dxdy \\ &= k \int_0^1 y \frac{x^2}{2}\Big|_y^1 dy \\ &= \frac{k}{2} \int_0^1 y(1-y^2)dy \\ &= \frac{k}{2} \int_0^1 (y-y^3)dy \\ &= \frac{k}{2}\left(\frac{y^2}{2} - \frac{y^4}{4}\right)\Big|_0^1 = \frac{k}{8} \end{aligned}$$

We get the same answer! Which one was easier?

2. Find $\mathbb{P}(X < 0.6 \cap Y > 0.2)$.

Here we are finding the probability that the pair (X, Y) falls in a specific region of the plane. The first step (as always) is to draw the region.



From this picture, we get the limits of integration. Let's integrate in the y direction first. Recall that we found that $k = 8$above. This gives us the integral:

$$\mathbb{P}(X < 0.6 \cap Y > 0.2) = \int_{0.2}^{0.6} \int_{0.2}^{x} 8xy \, dydx$$

This integral can be evaluated like the one above. The most important thing to know is how to set the problem up with the correct limits, but for completeness we will do the computation below.

$$\begin{aligned}
\mathbb{P}(X < 0.6 \cap Y > 0.2) &= \int_{0.2}^{0.6} \int_{0.2}^{x} 8xy \, dy dx \\
&= 8 \int_{0.2}^{0.6} x \frac{y^2}{2} \Big|_{0.2}^{x} dx \\
&= 4 \int_{0.2}^{0.6} (x^3 - 0.04x) dx \\
&= 4 \left(\frac{x^4}{4} - 0.04 \frac{x^2}{2} \right) \Big|_{0.2}^{0.6} \\
&= \left(0.6^4 - 0.2^4\right) - 0.08 \left(0.6^2 - 0.2^2\right) \\
&= 0.1024
\end{aligned}$$

We could also have integrated in the x direction first.