

Conditional densities, mass functions, and expectations

Jason Swanson

April 22, 2007

1 Discrete random variables

Suppose that X is a discrete random variable with range $\{x_1, x_2, x_3, \dots\}$, and that Y is also a discrete random variable with range $\{y_1, y_2, y_3, \dots\}$. Their **joint probability mass function** is

$$p(x_i, y_j) = P(X = x_i, Y = y_j).$$

In order for $p(x_i, y_j)$ to be a valid joint probability mass function, it is enough that $p(x_i, y_j) \geq 0$ for all x_i and y_j , and that

$$\sum_i \sum_j p(x_i, y_j) = 1.$$

The **marginal mass function of X** is

$$p_X(x_i) = P(X = x_i) = \sum_j p(x_i, y_j),$$

and the **marginal mass function of Y** is

$$p_Y(y_j) = P(Y = y_j) = \sum_i p(x_i, y_j).$$

The **conditional mass function of X given Y** is

$$p_{X|Y}(x_i|y_j) = P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p(x_i, y_j)}{p_Y(y_j)}.$$

Similarly, the **conditional mass function of Y given X** is

$$p_{Y|X}(y_j, x_i) = P(Y = y_j|X = x_i) = \frac{P(Y = y_j, X = x_i)}{P(X = x_i)} = \frac{p(x_i, y_j)}{p_X(x_i)}.$$

If X and Y are independent, then $p_{X|Y}(x_i|y_j) = p_X(x_i)$ and $p_{Y|X}(y_j|x_i) = p_Y(y_j)$.

2 Continuous random variables

Suppose X and Y are jointly continuous random variables. This means they have a **joint density** $f(x, y)$, so that

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy$$

for all¹ subsets A of the xy -plane. In order for $f(x, y)$ to be a valid joint density function, it is enough that $f(x, y) \geq 0$ for all x and y , and that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

The **marginal densities** of X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx,$$

and the **marginal distributions** are

$$F_X(a) = P(X \leq a) = \int_{-\infty}^a f_X(x) dx, \quad F_Y(b) = P(Y \leq b) = \int_{-\infty}^b f_Y(y) dy.$$

Sometimes we are given that $Y = y$, where y is some constant, and we would like to know the conditional probability that $X \leq a$. However,

$$P(X \leq a | Y = y) \neq \frac{P(X \leq a, Y = y)}{P(Y = y)}.$$

The reason these are not equal is because the right-hand side is undefined. It is undefined because $P(Y = y) = 0$. We must therefore do something else if we want to define $P(X \leq a | Y = y)$.

To accomplish this, we define the **conditional density of X given Y** as

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

This is defined whenever $f_Y(y) > 0$. If $f_Y(y) = 0$, then the conditional density $f_{X|Y}(x|y)$ is undefined. We then define

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y) dx$$

for all subsets A of the real line. The **conditional distribution of X given Y** is then

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

¹As mentioned in the text, this is not really true for *all* subsets. It is, however, true for all subsets that we care about in this course. We will therefore ignore this technical detail in this and all similar situations.

Similarly,

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \quad (2.1)$$

is the **conditional density of Y given X** . This is defined whenever $f_X(x) > 0$. If $f_X(x) = 0$, then the conditional density $f_{Y|X}(y|x)$ is undefined. We then define

$$P(Y \in A|X = x) = \int_A f_{Y|X}(y|x) dy$$

for all subsets A of the real line. The **conditional distribution of Y given X** is

$$F_{Y|X}(b|x) = P(Y \leq b|X = x) = \int_{-\infty}^b f_{Y|X}(y|x) dy.$$

If X and Y are independent, then $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$.

Example 2.1 *The random variables X and Y are jointly continuous. The marginal density of X is*

$$f_X(x) = \begin{cases} \frac{1}{6}x^3e^{-x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

The conditional density of Y given X is

$$f_{Y|X}(y|x) = \begin{cases} \frac{3(x-y)^2}{8x^3} & \text{if } |y| < x, \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

Find $P(X > 3|Y = 2)$.

Solution. What we want to compute is

$$P(X > 3|Y = 2) = \int_3^{\infty} f_{X|Y}(x|2) dx.$$

We therefore need to compute

$$f_{X|Y}(x|2) = \frac{f(x, 2)}{f_Y(2)}.$$

For the numerator, we use equation (2.1) to get

$$f(x, y) = f_{Y|X}(y|x)f_X(x) = \begin{cases} \frac{1}{16}(x-y)^2e^{-x} & \text{if } x > 0 \text{ and } |y| < x, \\ 0 & \text{otherwise.} \end{cases}$$

So

$$f(x, 2) = \begin{cases} \frac{1}{16}(x-2)^2e^{-x} & \text{if } x > 2, \\ 0 & \text{otherwise.} \end{cases}$$

For the denominator, we have

$$f_Y(2) = \int_{-\infty}^{\infty} f(x, 2) dx = \frac{1}{16} \int_2^{\infty} (x-2)^2 e^{-x} dx = \frac{1}{16} \int_0^{\infty} u^2 e^{-u-2} du = \frac{1}{8e^2}.$$

Putting these together gives

$$f_{X|Y}(x|2) = \frac{f(x, 2)}{f_Y(2)} = \begin{cases} \frac{e^2}{2} (x-2)^2 e^{-x} & \text{if } x > 2, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we have

$$P(X > 3 | Y = 2) = \int_3^{\infty} f_{X|Y}(x|2) dx = \frac{e^2}{2} \int_3^{\infty} (x-2)^2 e^{-x} dx = \frac{e^2}{2} \int_1^{\infty} u^2 e^{-u-2} du = \frac{5}{2e},$$

which was what we wanted to compute. \square

3 Discrete and continuous random variables

Suppose that X is a discrete random variable with range $\{x_1, x_2, x_3, \dots\}$, and that Y is a continuous random variable. If there exists a function $f(x_i, y)$ such that

$$P(X = x_i, Y \in A) = \int_A f(x_i, y) dy$$

for all subsets A of the real line, then we will call f the **joint density/mass function** of X and Y . In order for $f(x_i, y)$ to be a valid joint density/mass function, it is enough that $f(x_i, y) \geq 0$ for all x_i and y , and that

$$\sum_i \int_{-\infty}^{\infty} f(x_i, y) dy = 1.$$

It is worth noting that, in this case, we can sum and integrate in any order. In other words,

$$\sum_i \left(\int_{-\infty}^{\infty} f(x_i, y) dy \right) = \int_{-\infty}^{\infty} \left(\sum_i f(x_i, y) \right) dy.$$

The **marginal mass function** of X is

$$p_X(x_i) = P(X = x_i) = \int_{-\infty}^{\infty} f(x_i, y) dy,$$

and the **marginal density** of Y is

$$f_Y(y) = \sum_i f(x_i, y).$$

If X and Y are independent, then $f(x_i, y) = p_X(x_i)f_Y(y)$. The **marginal distribution of Y** is

$$F_Y(a) = P(Y \leq a) = \int_{-\infty}^a f_Y(y) dy.$$

If we are given that $X = x_i$ and we want to compute the conditional probability that $Y \leq a$, then we can do this in the ordinary fashion, since $P(X = x_i) > 0$. We define the **conditional density of Y given X** by

$$f_{Y|X}(y|x_i) = \frac{f(x_i, y)}{p_X(x_i)}.$$

The **conditional distribution of Y given X** is then

$$F_{Y|X}(a|x_i) = P(Y \leq a|X = x_i) = \frac{P(Y \leq a, X = x_i)}{P(X = x_i)} = \int_{-\infty}^a f_{Y|X}(y|x_i) dy.$$

However, if we are given that $Y = y$, where y is some constant, then

$$P(X = x_i|Y = y) \neq \frac{P(X = x_i, Y = y)}{P(Y = y)}.$$

As before, the right-hand side is undefined since $P(Y = y) = 0$. In this case, we define

$$P(X = x_i|Y = y) = \frac{f(x_i, y)}{f_Y(y)}.$$

Using this definition, the **conditional mass function of X** is

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

If X and Y are independent, then $f_{Y|X}(y|x_i) = f_Y(y)$ and $p_{X|Y}(x_i|y) = p_X(x_i)$.

Example 3.1 Suppose we generate a random variable Y which is exponentially distributed with parameter 1. This gives us some value $Y = y$. We then generate a Poisson random variable X with parameter y . Find $P(X \geq 1, Y \leq 2)$. Also find $P(X = i)$ for $i \geq 0$.

Solution. According to the problem statement, we have

$$f_Y(y) = \begin{cases} e^{-y} & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$p_{X|Y}(i|y) = P(X = i|Y = y) = e^{-y} \frac{y^i}{i!}.$$

In general,

$$p_{X|Y}(x_i|y) = \frac{f(x_i, y)}{f_Y(y)}.$$

Hence, in this example,

$$f(i, y) = p_{X|Y}(i|y)f_Y(y) = \begin{cases} e^{-2y} \frac{y^i}{i!} & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$P(X \geq 1, Y \leq 2) = \sum_{i=1}^{\infty} P(X = i, Y \leq 2) = \sum_{i=1}^{\infty} \int_{-\infty}^2 f(i, y) dy = \sum_{i=1}^{\infty} \int_0^2 e^{-2y} \frac{y^i}{i!} dy.$$

In this case, it may be easier to move the summation sign to the inside, so that

$$\begin{aligned} P(X \geq 1, Y \leq 2) &= \int_0^2 \left(\sum_{i=1}^{\infty} e^{-2y} \frac{y^i}{i!} \right) dy = \int_0^2 \left(e^{-2y} \sum_{i=1}^{\infty} \frac{y^i}{i!} \right) dy \\ &= \int_0^2 e^{-2y} \left(\left(\sum_{i=0}^{\infty} \frac{y^i}{i!} \right) - 1 \right) dy \\ &= \int_0^2 e^{-2y} (e^y - 1) dy = \frac{1}{2} - e^{-2} + \frac{1}{2}e^{-4}. \end{aligned}$$

Another way to compute this is the following:

$$\begin{aligned} P(X \geq 1, Y \leq 2) &= P(Y \leq 2) - P(X = 0, Y \leq 2) \\ &= \int_{-\infty}^2 f_Y(y) dy - \int_{-\infty}^2 f(0, y) dy \\ &= \int_0^2 e^{-y} dy - \int_0^2 e^{-2y} dy = \frac{1}{2} - e^{-2} + \frac{1}{2}e^{-4}. \end{aligned}$$

For general $i \geq 0$,

$$P(X = i) = p_X(i) = \int_{-\infty}^{\infty} f(i, y) dy = \frac{1}{i!} \int_0^{\infty} y^i e^{-2y} dy = \frac{1}{i!2^{i+1}} \int_0^{\infty} u^i e^{-u} du.$$

By the definition of the gamma function, $\Gamma(i+1) = \int_0^{\infty} u^i e^{-u} du$. Since $\Gamma(i+1) = i!$, this gives $P(X = i) = 2^{-i-1}$. \square

Example 3.2 We generate a random variable U which is uniformly distributed on $(0, 1)$. This gives us a value $U = p$. We then create a biased coin with probability of heads p . We flip this coin 3 times. Let N denote the number of heads flipped. Compute $P(U < 0.5 | N = 2)$.

Solution. The problem tells us that

$$f_U(p) = \begin{cases} 1 & \text{if } 0 < p < 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$p_{N|U}(i|p) = P(N = i|U = p) = \binom{3}{i} p^i (1-p)^{3-i} \quad \text{if } 0 \leq i \leq 3.$$

This means the joint density/mass function is

$$f(i, p) = p_{N|U}(i|p)f_U(p) = \begin{cases} \binom{3}{i} p^i (1-p)^{3-i} & \text{if } 0 < p < 1 \text{ and } 0 \leq i \leq 3, \\ 0 & \text{otherwise.} \end{cases}$$

What we want to compute is

$$P(U < 0.5|N = 2) = \frac{P(U < 0.5, N = 2)}{P(N = 2)}.$$

Note that this is an ordinary conditional probability, since $P(N = 2) > 0$. For the numerator, we have

$$P(U < 0.5, N = 2) = \int_{-\infty}^{0.5} f(2, p) dp = \int_0^{0.5} 3p^2(1-p) dp = \frac{5}{64}.$$

For the denominator, we have

$$P(N = 2) = p_N(2) = \int_{-\infty}^{\infty} f(2, p) dp = \int_0^1 3p^2(1-p) dp = \frac{1}{4}.$$

Therefore, $P(U < 0.5|N = 2) = (5/64)/(1/4) = 5/16$. □

4 Conditional expectations

If X is a discrete random variable, and Y is either discrete or continuous, then the **conditional expectation of X given $Y = y$** is

$$E[X|Y = y] = \sum_i x_i P(X = x_i|Y = y) = \sum_i x_i p_{X|Y}(x_i|y).$$

If X is a continuous random variable, and Y is either discrete or continuous, then

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

It is important to remember that $E[X|Y = y]$ is a function of y .

The **conditional expectation of X given Y** (which is denoted by $E[X|Y]$) is the random variable we get when we plug Y into the function $E[X|Y = y]$. In other words, if

$$g(y) = E[X|Y = y],$$

then

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

Since it is so easy to be confused by this, let me re-emphasize: $E[X|Y = y]$ is a function of y , but $E[X|Y]$ is a random variable. We obtain $E[X|Y]$ by plugging Y into the function $E[X|Y = y]$.

Similarly, the **conditional expectation of Y given $X = x$** is

$$E[Y|X = x] = \sum_j y_j P(Y = y_j|X = x) = \sum_j y_j p_{Y|X}(y_j|x)$$

when Y is discrete, and

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

when Y is continuous. In either case, $E[Y|X = x]$ is a function of x . The **conditional expectation of Y given X** (which is denoted by $E[Y|X]$) is the random variable we get when we plug X into that function. In other words, if

$$h(x) = E[Y|X = x],$$

then

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

To re-emphasize, $E[Y|X = x]$ is a function of x , but $E[Y|X]$ is a random variable. We obtain $E[Y|X]$ by plugging X into the function $E[Y|X = x]$.

Example 4.1 In Example 2.1, find $E[Y|X = x]$ and $E[Y|X]$.

Solution. We want to first compute

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

In Example 2.1, $f_{Y|X}(y|x)$ is given by equation (2.3). Remember that $f_{Y|X}(y|x)$ is only defined when $f_X(x) > 0$. Looking at equation (2.2), we see that $f_{Y|X}(y|x)$ is undefined for $x \leq 0$. This means that $E[Y|X = x]$ is also undefined for $x \leq 0$. In our computations, therefore, we may assume that $x > 0$. We now compute, using equation (2.3),

$$\begin{aligned} E[Y|X = x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \int_{-x}^x \frac{3y(x-y)^2}{8x^3} dy \\ &= \frac{3}{8x^3} \int_{-x}^x (x^2y - 2xy^2 + y^3) dy \\ &= \frac{3}{8x^3} \left(\frac{1}{2}x^2y^2 - \frac{2}{3}xy^3 + \frac{1}{4}y^4 \right) \Big|_{y=-x}^{y=x} = -\frac{x}{2}. \end{aligned}$$

So

$$E[Y|X = x] = \begin{cases} -\frac{x}{2} & \text{if } x > 0, \\ \text{undefined} & \text{if } x \leq 0. \end{cases}$$

This, of course, is a function of x , as it should be. In order to get $E[Y|X]$, we must plug X into this function. You might think there is a problem with this. After all, how can we plug X into a function which might not even be defined? Well, remember that $f_X(x) = 0$ for all $x \leq 0$. This means that $P(X > 0) = 1$. In other words, even though X will take on a random value, it will always be a random positive value, and the function $E[Y|X = x]$ is perfectly well-defined when we plug in positive numbers. So, in the end,

$$E[Y|X] = -\frac{X}{2}.$$

This, of course, is a random variable, as it should be. □

Since $E[X|Y]$ and $E[Y|X]$ are random variables, we can take their expectations. As an example, suppose that X is continuous. To compute $E[X|Y]$, we first compute

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

We then obtain $E[X|Y]$ by plugging Y into this function. If we give this function a name, say $g(y) = E[X|Y = y]$, then $E[X|Y] = g(Y)$. Since $g(Y)$ is a random variable, we can take its expectation. We then get

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

By the definition of the conditional density of X given Y , this gives

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \left(\frac{f(x, y)}{f_Y(y)} \right) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dy dx \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx = E[X]. \end{aligned}$$

This fact holds in general and is an identity of fundamental importance.

Theorem 4.2 $E[E[X|Y]] = E[X]$.

It is useful to remember what this theorem says in the separate cases that Y is discrete or Y is continuous. If Y is continuous, then it says that

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y) dy.$$

If Y is discrete, then it says that

$$E[X] = \sum_j E[X|Y = y_j] p_Y(y_j).$$

Example 4.3 In Example 3.2, compute $E[N]$.

Solution. We can compute this by conditioning on U . We have

$$E[N] = E[E[N|U]] = \int_{-\infty}^{\infty} E[N|U = p] f_U(p) dp = \int_0^1 E[N|U = p] dp$$

If we are given $U = p$, then N is binomially distributed with parameters 3 and p . So $E[N|U = p] = 3p$. This gives $E[N] = \int_0^1 3p dp = 3/2$. \square

Let us pause momentarily and regain our perspective. The purpose of this part of the course is to do conditioning. We learned early on how to condition on an event B . To compute the probability of A given B , we simply compute $P(A|B) = P(A \cap B)/P(B)$. From this, we see that whenever we want to condition on an event B , we are going to end up dividing by $P(B)$. If $P(B) = 0$, then we have a problem.

This is exactly the problem we are facing in this part of the course. Right now, we are trying to condition on the event $\{Y = y\}$. If Y is a continuous random variable, then $P(Y = y) = 0$, and this is what has forced us to define and work with all of the objects in these notes.

So far, we have succeeded in defining $P(X \in A|Y = y)$ whenever A is a subset of the real line. But what if we want to compute $P(F|Y = y)$ where F is some arbitrary event? How can we define this?

In order to proceed, we must recall the notation for indicator functions. If F is an event, then the indicator function of F is the random variable 1_F defined by

$$1_F = \begin{cases} 1 & \text{if } F \text{ occurs,} \\ 0 & \text{if } F \text{ does not occur.} \end{cases}$$

Also remember that $P(F) = E[1_F]$.

Motivated by this, we define

$$P(F|Y = y) = E[1_F|Y = y], \quad P(F|Y) = E[1_F|Y].$$

Notice that $P(F|Y = y)$ is a function of y . We obtain $P(F|Y)$ by plugging Y into that function. So $P(F|Y)$ is a random variable. Using this in conjunction with the previous theorem, we get that

$$P(F) = E[1_F] = E[E[1_F|Y]] = E[P(F|Y)].$$

This is a very useful result and it is worth remembering what this says in the separate cases that Y is discrete or Y is continuous. If Y is continuous, then it says that

$$P(F) = \int_{-\infty}^{\infty} P(F|Y = y) f_Y(y) dy.$$

If Y is discrete, then it says that

$$P(F) = \sum_j P(F|Y = y_j) p_Y(p_j).$$

In the case that Y is discrete, we already learned this fact. In the case that Y is continuous, however, this is something new.

Example 4.4 *Joe's TV Shop sells expensive electronic equipment. Some customers come to Joe's, browse for a while, and then leave. Some customers (Joe hopes) actually buy something before they leave. It is known that older customers are more likely to buy something than younger customers. In fact, it is known that when a customer who is x years old enters Joe's TV shop, they will buy something with probability $x/(20 + x)$. Let X denote the age of the next customer to enter Joe's TV Shop. Assume that $X \sim \text{Unif}(20, 40)$. What is the probability that this customer buys something?*

Solution. Let F denote the event that the next customer buys something. We are given that

$$P(F|X = x) = \frac{x}{20 + x}.$$

Therefore,

$$\begin{aligned} P(F) &= \int_{-\infty}^{\infty} P(F|X = x) f_X(x) dx \\ &= \frac{1}{20} \int_{20}^{40} \frac{x}{20 + x} dx \\ &= \frac{1}{20} \int_{20}^{40} \left(1 - \frac{20}{20 + x} \right) dx \\ &= 1 - \log(20 + x) \Big|_{x=20}^{x=40} = 1 - \log(3/2). \end{aligned}$$

So $P(F) \approx 0.594535$. □

Many times we are considering independent random variables X and Y . In this case, $E[X|Y] = E[X]$ and $E[Y|X] = E[Y]$. This is an important and often useful fact to know. An even more useful fact is the following.

Suppose we have a function $h(x, y)$ and we want to compute $E[h(X, Y)|Y = y]$. Let us define $U = h(X, Y)$ and $V = Y$. Then what we are interested in is $E[U|V = y]$. Before we begin, let us establish some notation. The joint density of X and Y is $f(x, y) = f_X(x)f_Y(y)$, since X and Y are independent. Let $g(u, v)$ be the joint density of U and V . In general, U and V are not independent. Now, we want to compute

$$E[U|V = y] = \int_{-\infty}^{\infty} u g_{U|V}(u|y) du.$$

We know that

$$g_{U|V}(u|v) = \frac{g(u, v)}{g_V(v)}.$$

This means that

$$E[U|V = y] = \int_{-\infty}^{\infty} u \frac{g(u, y)}{g_V(y)} du.$$

Since $V = Y$, they must have the same marginal densities. In other words, $g_V(y) = f_Y(y)$. So we can write

$$E[U|V = y] = \int_{-\infty}^{\infty} u \frac{g(u, y)}{f_Y(y)} du. \quad (4.1)$$

To finish off the calculation, we must compute the joint density $g(u, v)$. We see that (U, V) is the image of (X, Y) under the transformation

$$\begin{aligned} u &= h(x, y), \\ v &= y. \end{aligned}$$

Let us assume that this transformation has an inverse,

$$\begin{aligned} x &= k(u, v), \\ y &= v. \end{aligned}$$

We then get

$$g(u, v) = f(k(u, v), v) / |J(k(u, v), v)|,$$

where $J(x, y)$ is the Jacobian of the transformation. That is,

$$J(x, y) = \det \begin{pmatrix} h_x(x, y) & h_y(x, y) \\ 0 & 1 \end{pmatrix} = h_x(x, y).$$

If we assume that $h_x(x, y) > 0$, then

$$g(u, v) = f(k(u, v), v) / h_x(k(u, v), v).$$

But remember that $f(x, y) = f_X(x)f_Y(y)$. So

$$g(u, v) = f_X(k(u, v))f_Y(v) / h_x(k(u, v), v).$$

Substituting this into (4.1) gives

$$E[U|V = y] = \int_{-\infty}^{\infty} u \frac{f_X(k(u, y))f_Y(y) / h_x(k(u, y), y)}{f_Y(y)} du = \int_{-\infty}^{\infty} u \frac{f_X(k(u, y))}{h_x(k(u, y), y)} du.$$

In this integral, remember that y is a constant. If we make the substitution $u = h(x, y)$, then $du = h_x(x, y) dx$, or $dx = du / h_x(k(u, y), y)$. This means that

$$E[U|Y = y] = \int_{-\infty}^{\infty} h(x, y) f_X(x) dx = E[h(X, y)].$$

This is true in general and is an important fact.

Theorem 4.5 *If X and Y are independent, then $E[h(X, Y)|Y = y] = E[h(X, y)]$.*

It is important here that X and Y be independent. The theorem is not true otherwise.

Example 4.6 *Let X and Y be independent with $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$. Find $P(X < Y)$.*

Solution. We can compute this by conditioning on X . We have

$$\begin{aligned} P(X < Y) &= E[P(X < Y|X)] = \int_{-\infty}^{\infty} P(X < Y|X = x)f_X(x) dx \\ &= \int_0^{\infty} P(X < Y|X = x)\lambda e^{-\lambda x} dx. \end{aligned}$$

Now, using the previous theorem,

$$P(X < Y|X = x) = E[1_{\{X < Y\}}|X = x] = E[1_{\{x < Y\}}] = P(x < Y).$$

Since $Y \sim \text{Exp}(\mu)$, $P(x < Y) = e^{-\mu x}$. So

$$P(X < Y) = \int_0^{\infty} e^{-\mu x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\mu + \lambda}.$$

This makes intuitive sense, since $EX = 1/\lambda$ and $EY = 1/\mu$. If λ is very large relative to μ , then X will be very small relative to Y , so it will be very likely that $X < Y$. Conversely, if μ is much larger than λ , then Y will likely be much smaller than X . \square

This example illustrates an important consequence of Theorem 4.5. If X and Y are independent, then

$$P((X, Y) \in A|Y = y) = P((X, y) \in A)$$

for all subsets A of the xy -plane. Also,

$$P((X, Y) \in A|X = x) = P((x, Y) \in A).$$

These facts are not true in general when X and Y are dependent.

Example 4.7 *Let X have density*

$$f_X(x) = \begin{cases} cx^{-1} & \text{if } 0.5 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y \sim \text{Exp}(1)$. Assume X and Y are independent. Find $E[X^Y]$.

Solution. We solve this by conditioning on X . We have

$$E[X^Y] = E[E[X^Y|X]] = \int_{-\infty}^{\infty} E[X^Y|X = x]f_X(x) dx = \int_{0.5}^1 E[x^Y]cx^{-1} dx.$$

We first compute $E[x^Y]$. We can write $x^Y = e^{(\log x)Y}$, so

$$E[x^Y] = \int_0^\infty e^{(\log x)y} e^{-y} dy = \frac{1}{1 - \log x}.$$

We then have

$$E[X^Y] = c \int_{0.5}^1 \frac{1}{1 - \log x} x^{-1} dx.$$

Making the substitution $u = 1 - \log x$ gives $du = -x^{-1} dx$. The limits of integration transform from $x = 0.5$ to $u = 1 - \log 0.5 = 1 + \log 2$ and from $x = 1$ to $u = 1 - \log 1 = 1$. So we have

$$E[X^Y] = c \int_1^{1+\log 2} \frac{1}{u} du = c \log(1 + \log 2).$$

Finally, we must compute c . Since

$$1 = \int_{0.5}^1 cx^{-1} dx = c(\log 1 - \log 0.5) = c \log 2,$$

we have $c = 1/\log 2$, which gives us

$$E[X^Y] = \frac{\log(1 + \log 2)}{\log 2}.$$

Numerically, $E[X^Y] \approx 0.759707$. □

5 Partial Summary

If X and Y are continuous, then

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y) dx,$$

where

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

If X is discrete and Y is continuous, then the joint density/mass function is the function $f(x_i, y)$ such that

$$P(X = x_i, Y \in A) = \int_A f(x_i, y) dy.$$

In this case,

$$\begin{aligned} p_X(x_i) &= \int_{-\infty}^{\infty} f(x_i, y) dy \\ f_Y(y) &= \sum_i f(x_i, y), \\ f_{Y|X}(y|x_i) &= \frac{f(x_i, y)}{p_X(x_i)}, \text{ and} \\ P(X = x_i | Y = y) &= \frac{f(x_i, y)}{f_Y(y)}. \end{aligned}$$

If X is discrete, then

$$E[X|Y = y] = \sum_i x_i P(X = x_i | Y = y).$$

If X is continuous, then

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

In either case, $E[X|Y = y]$ is a function of y . If we plug the random variable Y into this function, then we get a new random variable which we call $E[X|Y]$. This random variable satisfies $E[E[X|Y]] = E[X]$. If Y is discrete, this means that

$$E[X] = \sum_j E[X|Y = y_j] p_Y(y_j).$$

If Y is continuous, this means that

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y) dy.$$

If F is an event, then

$$P(F|Y = y) = E[1_F|Y = y],$$

and

$$P(F|Y) = E[1_F|Y].$$

In particular, this means that $P(F) = E[P(F|Y)]$. If Y is discrete, this means that

$$P(F) = \sum_j P(F|Y = y_j) p_Y(y_j).$$

If Y is continuous, this means that

$$P(F) = \int_{-\infty}^{\infty} P(F|Y = y) f_Y(y) dy.$$

All of the above is true in general. In particular, it is true even if X and Y are dependent.

If X and Y are independent, then more is true. We also have

$$\begin{aligned} E[X|Y] &= E[X], \\ E[g(X, Y)|Y = y] &= E[g(X, y)], \text{ and} \\ P((X, Y) \in A|Y = y) &= P((X, y) \in A). \end{aligned}$$

These three facts are not true in general when X and Y are dependent.