6 Jointly continuous random variables

Again, we deviate from the order in the book for this chapter, so the subsections in this chapter do not correspond to those in the text.

6.1 Joint density functions

Recall that X is continuous if there is a function $f(x)$ (the density) such that

$$\mathbf{P}(X \le t) = \int_{-\infty}^{t} f_X(x)\,dx$$

We generalize this to two random variables.

Definition 1. *Two random variables X and Y are jointly continuous if there is a function $f_{X,Y}(x,y)$ on $\mathbb{R}^2$, called the joint probability density function, such that*

$$\mathbf{P}(X \le s, Y \le t) = \int\int_{x\le s, y\le t} f_{X,Y}(x,y)\,dxdy$$

The integral is over $\{(x,y) : x \le s, y \le t\}$. We can also write the integral as

$$\begin{aligned}\mathbf{P}(X \le s, Y \le t) &= \int_{-\infty}^{s} \left(\int_{-\infty}^{t} f_{X,Y}(x,y)\,dy \right) dx \\ &= \int_{-\infty}^{t} \left(\int_{-\infty}^{s} f_{X,Y}(x,y)\,dx \right) dy\end{aligned}$$

In order for a function $f(x,y)$ to be a joint density it must satisfy

$$\begin{aligned} f(x,y) &\ge 0 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= 1\end{aligned}$$

Just as with one random variable, the joint density function contains all the information about the underlying probability measure if we only look at the random variables X and Y. In particular, we can compute the probability of any event defined in terms of X and Y just using $f(x,y)$.

Here are some events defined in terms of X and Y :
$\{X \le Y\}$, $\{X^2 + Y^2 \le 1\}$, and $\{1 \le X \le 4, Y \ge 0\}$. They can all be written in the form $\{(X,Y) \in A\}$ for some subset A of $\mathbb{R}^2$.

Proposition 1. *For $A \subset \mathbb{R}^2$,*

$$\mathbf{P}((X,Y) \in A) = \int\int_A f(x,y)\,dxdy$$

End of lecture - Mon, Oct 16

The two-dimensional integral is over the subset A of $\mathbb{R}^2$. Typically, when we want to actually compute this integral we have to write it as an iterated integral. It is a good idea to draw a picture of A to help do this.

A rigorous proof of this theorem is beyond the scope of this course. In particular we should note that there are issues involving σ-fields and constraints on A. We will just derive it for the following rectangle : $P(a < X \leq b, c < Y \leq d)$.

MORE !!!!!!!!!

Definition: Let $A \subset \mathbb{R}^2$. We say X and Y are uniformly distributed on A if

$$f(x) = \begin{cases} \frac{1}{c}, & \text{if } (x,y) \in A \\ 0, & \text{otherwise} \end{cases}$$

where c is the area of A.

Example: Let X, Y be uniform on $[0,1] \times [0,2]$. Find $\mathbf{P}(X + Y \leq 1)$.

Example: Let X, Y have density

$$f(x,y) = \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2))$$

Compute $\mathbf{P}(X \leq Y)$ and $\mathbf{P}(X^2 + Y^2 \leq 1)$.

Example: Now suppose X, Y have density

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

Compute $\mathbf{P}(X + Y \leq t)$.

What does the pdf mean? In the case of a single discrete RV, the pmf has a very concrete meaning. $f(x)$ is the probability that $X = x$. If X is a single continuous random variable, then

$$\mathbf{P}(x \leq X \leq x + \delta) = \int_x^{x+\delta} f(u)\, du \approx \delta f(x)$$

If X, Y are jointly continuous, than

$$\mathbf{P}(x \leq X \leq x + \delta, y \leq Y \leq y + \delta) \approx \delta^2 f(x, y)$$

6.2 Independence and marginal distributions

Suppose we know the joint density $f_{X,Y}(x, y)$ of X and Y. How do we find their individual densities $f_X(x)$, $f_Y(y)$. These are called *marginal densities.* The cdf of X is

$$\begin{aligned} F_X(x) &= \mathbf{P}(X \leq x) = \mathbf{P}(-\infty < X \leq x, -\infty < Y < \infty) \\ &= \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{X,Y}(u, y)\, dy \right] du \end{aligned}$$

Differentiate this with respect to x and we get

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

In words, we get the marginal density of X by integrating y from $-\infty$ to ∞ in the joint density.

Proposition 2. *If X and Y are jointly continuous with joint density $f_{X,Y}(x, y)$, then the marginal densities are given by*

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

We will define independence of two contiunous random variables differently than the book. The two definitions are equivalent.

Definition 2. *Let X, Y be jointly continuous random variables with joint density $f_{X,Y}(x,y)$ and marginal densities $f_X(x)$, $f_Y(y)$. We say they are independent if*

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

If we know the joint density of X and Y, then we can use the definition to see if they are independent. But the definition is often used in a different way. If we know the marginal densities of X and Y **and** we know that they are independent, then we can use the definition to find their joint density.

Example: If X and Y are independent random variables and each has the standard normal distribution, what is their joint density?

$$f(x,y) = \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2 + y^2))$$

Example: Suppose that X and Y have a joint density that is uniform on the disc centered at the origin with radius 1. Are they independent?

Example: If X and Y have a joint density that is uniform on the square $[a,b] \times [c,d]$, then they are independent.

Example: Suppose that X and Y have joint density

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

Are X and Y independent?

Example: Suppose that X and Y are independent. X is uniform on $[0,1]$ and Y has the Cauchy density.
(a) Find their joint density.
(b) Compute $\mathbf{P}(0 \leq X \leq 1/2, 0 \leq Y \leq 1)$
(c) Compute $\mathbf{P}(Y \geq X)$.

6.3 Expected value

If X and Y are jointly continuously random variables, then the mean of X is still given by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

If we write the marginal $f_X(x)$ in terms of the joint density, then this becomes

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

Now suppose we have a function $g(x,y)$ from $\mathbb{R}^2$ to $\mathbb{R}$. Then we can define a new random variable by $Z = g(X,Y)$. In a later section we will see how to compute the density of Z from the joint density of X and Y. We could then compute the mean of Z using the density of Z. Just as in the discrete case there is a shortcut.

Theorem 1. *Let X, Y be jointly continuous random variables with joint density $f(x,y)$. Let $g(x,y) : \mathbb{R}^2 \to \mathbb{R}$. Define a new random variable by $Z = g(X,Y)$. Then*

$$\mathbf{E}[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dxdy$$

provided

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)|\, f(x,y)\, dxdy < \infty$$

An important special case is the following

Corollary 1. *If X and Y are jointly continuous random variables and a, b are real numbers, then*

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

Example: X and Y have joint density

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

Let $Z = X + Y$. Find the mean and variance of Z.

We now consider independence and expectation.

Theorem 2. *If X and Y are independent and jointly continuous, then*

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

Proof. Since they are independent, $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. So

$$\begin{aligned}
\mathbf{E}[XY] &= \int\int xy\, f_X(x)\, f_Y(y)\, dxdy \\
&= \left[\int x\, f_X(x)\, dx\right] \left[\int y\, f_Y(y)\, dy\right] = \mathbf{E}[X]\mathbf{E}[Y]
\end{aligned}$$

$\square$

6.4 Function of two random variables

Suppose X and Y are jointly continuous random variables. Let $g(x,y)$ be a function from $\mathbb{R}^2$ to $\mathbb{R}$. We define a new random variable by $Z = g(X,Y)$. Recall that we have already seen how to compute the expected value of Z. In this section we will see how to compute the density of Z. The general strategy is the same as when we considered functions of one random variable: we first compute the cumulative distribution function.

Example: Let X and Y be independent random variables, each of which is uniformly distributed on $[0,1]$. Let $Z = XY$. First note that the range of Z is $[0,1]$.

$$F_Z(z) = \mathbf{P}(Z \le z) = \int\int_A 1\, dxdy$$

Where A is the region

$$A = \{(x,y) : 0 \le x \le 1, 0 \le y \le 1, xy \le z\}$$

PICTURE

$$\begin{aligned}
F_Z(z) &= z + \int_z^1 \left[\int_0^{z/x} 1\, dy\right] dx \\
&= z + \int_z^1 \left[\int_0^{z/x} 1\, dy\right] dx \\
&= z + \int_z^1 \frac{z}{x}\, dx \\
&= z + z \ln x|_z^1 = z - z\ln z
\end{aligned}$$

This is the cdf of Z. So we differentiate to get the density.

$$\frac{d}{dz}F_Z(z) = \frac{d}{dz}z - z\ln z = 1 - \ln z - z\frac{1}{z} = -\ln z$$

$$f_Z(z) = \begin{cases} -\ln z, & \text{if } 0 \leq z \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Example: Let X and Y be independent random variables, each of which is exponential with parameter λ. Let $Z = X + Y$. Find the density of Z.

Should get gamma with same λ and $w = 2$.

This is special case of a much more general result. The sum of gamma(λ, w_1) and gamma(λ, w_2) is gamma$(\lambda, w_1 + w_2)$. We could try to show this as we did the previous example. But it is much easier to use moment generating functions which we will introduce in the next section.

Example: Let (X, Y) be uniformly distributed on the triangle with vertices at $(0,0), (1,0), (0,1)$. Let $Z = X + Y$. Find the pdf of Z.

One of the most important examples of a function of two random variables is $Z = X + Y$. In this case

$$\begin{aligned} F_Z(z) &= \mathbf{P}(Z \leq z) = \mathbf{P}(X + Y \leq z) \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f(x,y)\,dy \right] dx \end{aligned}$$

To get the density of Z we need to differentiate this with respect to Z. The only z dependence is in the upper limit of the inside integral.

$$\begin{aligned} f_Z(z) = \frac{d}{dz}F_Z(z) &= \int_{-\infty}^{\infty} \left[\frac{d}{dz} \int_{-\infty}^{z-x} f(x,y)\,dy \right] dx \\ &= \int_{-\infty}^{\infty} f(x, z-x) dx \end{aligned}$$

If X and Y are independent, then this becomes

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

This is known as a convolution. We can use this formula to find the density of the sum of two independent random variables. But in some cases it is easier to do this using generating functions which we study in the next section.

Example: Let X and Y be independent random variables each of which has the standard normal distribution. Find the density of $Z = X + Y$.

We need to compute the convolution

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\frac{1}{2}x^2 - \frac{1}{2}(z-x)^2)\,dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-x^2 - \frac{1}{2}z^2 + xz)\,dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-(x-z/2)^2 - \frac{1}{4}z^2)\,dx \\ &= e^{-z^2/4}\,\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-(x-z/2)^2)\,dx \end{aligned}$$

Now the substitution $u = x - z/2$ shows

$$\int_{-\infty}^{\infty} \exp(-(x-z/2)^2)\,dx = \int_{-\infty}^{\infty} \exp(-u^2)\,du$$

This is a constant - it does not depend on z. So $f_Z(z) = ce^{-z^2/4}$. Another simple substitution allows one to evaluate the constant, but there is no need. We can already see that Z has a normal distribution with mean zero and variance 2. The constant is whatever is needed to normalize the distribution.

6.5 Moment generating functions

This will be very similar to what we did in the discrete case.

Definition 3. *For a continuous random variable X, the moment generating function (mgf) of X is*

$$M_X(t) = \mathbf{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx}\,f_X(x)\,dx$$

Example: Compute it for the gamma distribution and find

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^w$$

A special case of the gamma distribution is the exponential distribution - you just take $w = 1$. So we see that for the exponential $M(t) = \frac{\lambda}{\lambda - t}$.

Proposition 3. *(1) Let X be a continuous random variable with mgf $M_X(t)$. Then*

$$\mathbf{E}[X^k] = \frac{d^k}{dt^k} M_X(t)|_{t=0}$$

(2) If X and Y are independent continuous random variables then

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

(3) If the mgf of X is $M_X(t)$ and we let $Y = aX + b$, then

$$M_Y(t) = e^{tb} M_X(at)$$

Proof. For (1)

$$\begin{aligned}
\frac{d^k}{dt^k} M_X(t)|_{t=0} &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} f_X(x)\, e^{tx}|_{t=0}\, dx \\
&= \int_{-\infty}^{\infty} f_X(x)\, \frac{d^k}{dt^k} e^{tx}|_{t=0}\, dx \\
&= \int_{-\infty}^{\infty} f_X(x)\, x^k\, e^{tx}|_{t=0}\, dx \\
&= \int_{-\infty}^{\infty} f_X(x)\, x^k\, dx = \mathbf{E}[X^k]
\end{aligned}$$

If X and Y are independent, then

$$\begin{aligned}
M_{X+Y}(t) &= \mathbf{E}[\exp(t(X+Y))] = \mathbf{E}[\exp(tX)\exp(tY)] \\
&= \mathbf{E}[\exp(tX)]\mathbf{E}[\exp(tY)] = M_X(t) M_Y(t)
\end{aligned}$$

This calculation assumes that since X and Y are independent, then $\exp(tX)$ and $\exp(tY)$ are independent random variables. We have not shown this.

Part (3) is just

$$M_Y(t) = \mathbf{E}[e^{tY}] = \mathbf{E}[e^{t(aX+b)}] = e^{tb}\mathbf{E}[e^{taX}] = e^{tb} M_X(at)$$

$\square$

As an application of part (3) we have

Example: Let X have the exponential distribution with parameter λ. Let $Y = X/\lambda$. Use mgf's to show Y has the exponential distribution with parameter 1.

Example: In the homework you show that the mgf for the normal density is

$$M_X(t) = \exp(\mu t) M_Z(\sigma t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

Proposition 4. *(a) If $X_1, X_2, \cdots, X_n$ are independent and each is normal with mean μ_i and variance σ_i^2, then $Y = X_1 + X_2 + \cdots + X_n$ has a normal distribution with mean μ and variance σ^2 given by*

$$\begin{aligned} \mu &= \sum_{i=1}^{n} \mu_i, \\ \sigma^2 &= \sum_{i=1}^{n} \sigma_i^2 \end{aligned}$$

(b) If $X_1, X_2, \cdots, X_n$ are independent and each is exponential with parameter λ, then $Y = X_1 + X_2 + \cdots + X_n$ has a gamma distribution with parameters $\lambda = \lambda$ and $w = n$.
(c) If $X_1, X_2, \cdots, X_n$ are independent and each is gamma with parameters λ, w_i, then $Y = X_1 + X_2 + \cdots + X_n$ has a gamma distibution with parameters λ and $w = w_1 + \cdots + w_n$.

We will prove the theorem by proving statements about generating functions. For example, for part (a) what we will really prove is that the moment generating function of Y is that of a normal with the stated parameters. To complete the proof we need to know that if two random variables have the same moment generating functions then they have the same densities. This is a theorem but it is a hard theorem and it requires some technical assumptions on the random variables. We will ignore these subtleties and just assume that if two RV's have the same mgf, then they have the same density.

Proof. We prove all three parts by simply computing the mgf's involved. $\square$

6.6 Cumulative distribution functions and more independence

Recall that for a discrete random variable X we have a probability mass function $f_X(x)$ which is just $f_X(x) = \mathbf{P}(X = x)$. And for a continuous random variable X we have a probability density function $f_X(x)$. It is a density in the sense that if $\epsilon > 0$ is small, then $\mathbf{P}(x \leq X \leq x + \epsilon) \approx f(x)\epsilon$.

For both types of random variables we have a cumulative distribution function and its definition is the same for all types of RV's.

Definition 4. *Let X, Y be random variables (discrete or continuous). Their joint (cumulative) distribution function is*

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

If X and Y are jointly continuous then we can compute the joint cdf from their joint pdf:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(u,v)\, dv \right] du$$

If we know the joint cdf, then we can compute the joint pdf by taking partial derivatives of the above :

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

The joint cdf has properties similar to the cdf for a single RV.

Proposition 5. *Let $F(x,y)$ be the joint cdf of two continuous random variables. Then $F(x,y)$ is a continuous function on $\mathbb{R}^2$ and*

$$\lim_{x,y \to -\infty} F(x,y) = 0, \qquad \lim_{x,y \to \infty} F(x,y) = 1,$$

$$F(x_1,y) \leq F(x_2,y) \ if \ x_1 \leq x_2, \quad F(x,y_1) \leq F(x,y_2) \ if \ y_1 \leq y_2$$

$$\lim_{x \to \infty} F(x,y) = F_Y(y) \qquad \lim_{y \to \infty} F(x,y) = F_X(x)$$

We will use the joint cdf to prove more results about independent of RV's.

Theorem 3. *If X and Y are jointly continuous random variables then they are independent if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.*

The theorem is true for discrete random variables as well.

Proof. $\square$

Example: Suppose that the joint cdf of X and Y is

$$F(x,y) = \begin{cases} \frac{1}{2}(1-e^{-2x})(y+1) & \text{if } x \geq 0, -1 \leq y \leq 1 \\ (1-e^{-2x}) & \text{if } x \geq 0, y \geq 1 \\ 0 & \text{if } x \geq 0, y < -1 \\ 0 & \text{if } x < 0 \end{cases}$$

Show that X and Y are independent and find their joint density.

Theorem 4. *If X and Y are independent jointly continuous random variables and g and h are functions from $\mathbb{R}$ to $\mathbb{R}$ then $g(X)$ and $h(Y)$ are independent random variables.*

We will only prove a special case of the theorem. In the homework you prove two other special cases.

Proof. We prove the theorem for g and h that are increasing. We also assume they are differentiable. Let $W = g(X)$, $Z = h(Y)$. By the previous theorem we can show that W and Z are independent by showing that $F_{W,Z}(w,z) = F_W(w)F_Z(z)$. We have

$$F_{W,Z}(w,z) = \mathbf{P}(g(X) \leq w, h(Y) \leq z)$$

Because g and h are increasing, the event $\{g(X) \leq w, h(Y) \leq z\}$ is the same as the event $\{X \leq g^{-1}(w), Y \leq h^{-1}(z)\}$. So

$$\begin{aligned} F_{W,Z}(w,z) &= \mathbf{P}(X \leq g^{-1}(w), Y \leq h^{-1}(z)) \\ &= F_{X,Y}(g^{-1}(w), h^{-1}(z)) = F_X(g^{-1}(w))F_Y(h^{-1}(z)) \end{aligned}$$

where the last equality comes from the previous theorem and the independence of X and Y. The individual cdfs of W and Z are

$$\begin{aligned} F_W(w) &= \mathbf{P}(X \leq g^{-1}(w)) = F_X(g^{-1}(w)) \\ F_Z(z) &= \mathbf{P}(Y \leq h^{-1}(z)) = F_Y(h^{-1}(z)) \end{aligned}$$

So we have shown $F_{W,Z}(w,z) = F_W(w)F_Z(z)$. $\square$

Corollary 2. *If X and Y are independent jointly continuous random variables and g and h are functions from $\mathbb{R}$ to $\mathbb{R}$ then*

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$$

Recall that for any two random variables X and Y, we have $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$. If they are independent we also have

Theorem 5. *If X and Y are independent and jointly continuous, then*

$$var(X+Y) = var(X) + var(Y)$$

Proof. $\square$

6.7 More than two jointly continuous RV's

6.8 Change of variables

Suppose we have two random variables X and Y and we know their joint density. We have two functions $g : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$, and we define two new random variables by $U = g(X, Y)$, $W = h(X, Y)$. Can we find the joint density of U and W? In principle we can do this by computing their joint cdf and then taking partial derivatives. In practice this can be a mess. There is a another way involving Jacobians which we will study in this section. But we start by illustrating the cdf approach with an example.

Example Let X and Y be independent standard normal RV's. Let $U = X + Y$ and $W = X - Y$. Find the joint density of U and W. After a lot of computation you should find that U and W are independent and each is a normal RV with mean zero and variance 2.

There is a another way to compute the joint density of W, Y that we will now study. First we return to the case of a function of a single random variable. Support that X is a continuous random variable and we know it's density. g is a function from $\mathbb{R}$ to $\mathbb{R}$ and we define a new random variable $Y = g(Z)$. We want to find the density of Y. Our previous approach was to compute the cdf first. Now suppose that g is strictly increasing on the range of X. Then we have the following formula.

Proposition 6. *If X is a continuous random variable whose range is D and $f : D \to \mathbb{R}$ is strictly increasing and differentiable, then*

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$$

Proof.

$$\mathbf{P}(Y \leq y) = \mathbf{P}(g(X) \leq y) = \mathbf{P}(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx$$

Now differentiate both sides with respect to y to finish the proof. ∎

We review some multivariate calculus. Let D and S be open subsets of $\mathbb{R}^2$. Let $T(x,y)$ be a map from D to S that is 1-1 and onto. (So it has an inverse.) We also assume it is differentiable. For each point in D, $T(x,y)$ is in $\mathbb{R}^2$. So we can write T as $T(x,y) = (u(x,y), w(x,y))$ We have an integral

$$\int\int_D f(x,y)\,dxdy$$

that we want to rewrite as an integral over S with respect to u and w. This is like doing a substitution in a one-dimensional integral. In that case you have $dx = \frac{dx}{du} du$ The analog of dx/du here is the Jacobian

$$J(u,w) = det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial w} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial w}$$

We then have

$$\int\int_D f(x,y)\,dxdy = \int\int_S f(T^{-1}(u,w))\,|J(u,w)|\,dudw$$

Often $f(T^{-1}(u,w))$ is simply written as $f(u,w)$. In practice you write f, which is originally a function of x and y as a function of u and w.

If A is a subset of D, then we have

$$\int\int_A f(x,y)\,dxdy = \int\int_{T(A)} f(T^{-1}(u,w))\,|J(u,w)|\,dudw$$

We now state what this results says about joint pdf's.

Proposition 7. *Let $T(x,y)$ be a 1-1, onto map from D to S. Let X, Y be random variables such that range of (X,Y) is D, and let $f_{X,Y}(x,y)$ be their joint density. Define two new random variables by $(U,W) = T(X,Y)$. Then the range of (U,W) is S and their joint pdf on this range is*

$$f_{U,W}(u,w) = f(T^{-1}(u,w))\,|J(u,w)|$$

where the Jacobian $J(u,w)$ is defined above.

Example - Polar coordinates Let X and Y be independent standard normal random variables. Define new random variables R, Θ by

$$x = r\cos(\theta), \quad y = r\sin(\theta)$$

Find the joint density of R, Θ.

Some calculation shows the Jacobian is r. (This is the same r you saw in vector calc: $dxdy = rdrd\theta$.) And the joint density is

$$f_{R,\Theta}(r,\theta) = \begin{cases} \frac{1}{2\pi} r e^{-r^2/2} & \text{if } r \geq 0, 0 \leq \theta \leq 2\pi \\ 0, & \text{otherwise} \end{cases}$$

Note that this implies that R and Θ are independent.

Example We redo the example that we started this section with and did using the joint cdf. X and Y are independent standard normal RV's. $U = X + Y$ and $W = X - Y$.

Example Let X and Y be independent random variables. They both have an exponential distribution with $\lambda = 1$. Let

$$\begin{aligned} U &= X + Y, \\ W &= \frac{X}{X+Y} \end{aligned}$$

Find the joint density of U and W.

Let $T(x,y) = (x+y, \frac{x}{x+y})$. Then T is a bijection from $[0,\infty) \times [0,\infty)$ onto $[0,\infty) \times [0,1]$. We need to find its inverse, i.e., find x, y in terms of u, w. Multiply the two equations to get $x = uw$. Then $y = u - x = u - uw$. So $T^{-1}(u,w) = (uw, u - uw)$. And so

$$J(u,w) = det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} = det \begin{pmatrix} w & u \\ 1-w & -u \end{pmatrix} = -u$$

So

$$f_{U,W}(u,w) = \begin{cases} ue^{-u} & \text{if } u \geq 0, 0 \leq w \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Example Let X and Y be independent random variables. X has a gamma distribution with parameters $\lambda = 1$ and $w = 2$. Y has an exponential distribution with parameter $\lambda = 1$. Let $U = X + Y$ and $W = Y/X$. Find the joint pdf of U, W.

$X = U/(1+V)$, $Y = UV/(1+V)$.

Jacobian is $U/(1+V)^2$.

Bivariate normal If X and Y are independent standard normal RV's, then their joint density is proportional to $\exp(-\frac{1}{2}(x^2+y^2))$. This is a special case of a bivariate normal distribution. In the more general case they need not be independent. We find consider a special case of the bivariate normal. Let $-1 < \rho < 1$. Define

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2))$$

You can compute the marginals of this joint distribution by the usual trick of completing the square. You find that X and Y both have a standard normal distribution. Note that the stuff in the exponential is a quadratic form in x and y. A more general quadratic form would have three parameters:

$$\exp(-(Ax^2 + 2Bxy + Cy^2))$$

In order for the intergal to converge the quadratic form $Ax^2 + 2Bxy + Cy^2$ must be positive.

Now suppose we start with two independent random variables X and Y which are independent and define

$$U = aX + bY, \quad W = cX + dY$$

where a, b, c, d are real numbers. In matrix notation

$$\begin{pmatrix} U \\ W \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

What is the joint density of U and W ? The transformation T is linear and so its inverse is linear (assuming it is invertible). So the Jacobian will just be a constant. So the joint density of U, W will be of the form $\exp(-\frac{1}{2} mess)$ where $mess$ is what we get when we rewrite $x^2 + y^2$ in terms of u and w. Argue this will be of the form $Au^2 + 2Buw + Cw^2$. So we get some sort of bivariate normal.

This can be generalized to n RV's. A joint normal (or Gaussian) distribution is of the form

$$c \exp(-\frac{1}{2}(x, Mx))$$

where M is a positive definite n by n matrix and c is the normalizing constant.

Correlation coefficient

If X and Y are independent, then $\mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0$. If there are not independent, it need not be zero and it is in some sense a measure of how dependent they are.

Definition 5. *The covariance of X and Y is*

$$cov(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

The correlation coefficient is a

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

The correlation coefficient has the advantage that it is scale invariant: $\rho(aX, bY) = \rho(X, Y)$. It can be shown that for any random variables $-1 \leq \rho(X, Y) \leq 1$.

Bivariate normal - cont We return to the joint density

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2))$$

Note that $f(-x, -y) = f(x, y)$. This implies $\mathbf{E}[X] = \mathbf{E}[Y] = 0$. So $cov(X, Y) = \mathbf{E}[XY]$.

$$\begin{aligned}
\mathbf{E}[XY] &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int\int xy \exp(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)})\, dxdy \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int x \exp(-\frac{1}{2(1-\rho^2)}x^2) \left[\int y \exp(-\frac{y^2 - 2\rho xy}{2(1-\rho^2)})\, dy\right] dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int x \exp(-\frac{1}{2(1-\rho^2)}x^2) \left[\int y \exp(-\frac{(y-\rho x)^2 - \rho^2 x^2}{2(1-\rho^2)})\, dy\right] dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int x \exp(-\frac{1}{2}x^2) \left[\int (y+\rho x) \exp(-\frac{y^2}{2(1-\rho^2)})\, dy\right] dx \\
&= \rho \frac{1}{2\pi\sqrt{1-\rho^2}} \left[\int x^2 \exp(-\frac{1}{2}x^2)\, dx\right] \left[\int \exp(-\frac{y^2}{2(1-\rho^2)})\, dy\right] \\
&= \rho
\end{aligned}
$$

So the correlation coefficient is ρ. Of coure this is why we wrote the density in the form that we did.

6.9 Conditional density and expectation

We first review what we have done. For events A, B,

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

provided $\mathbf{P}(B) > 0$. If we define $Q(A) = \mathbf{P}(A|B)$, then Q is a new probability measure.

Let X be a discrete RV with pmf $f_X(x)$. If we know B occurs the pmf for X will be different. The *conditional pmf of X given B* is

$$f(x|B) = \mathbf{P}(X = x|B) = \frac{\mathbf{P}(X = x, B)}{\mathbf{P}(B)}$$

The conditional expectation of X given B is

$$E[X|B] = \sum_x x\, f(x|B)$$

A partition is a collection of disjoint events B_n whose union is all of the sample space Ω. The partition theorem says that for a random variable X.

$$E[X] = \sum_n \mathbf{E}[X|B_n]\mathbf{P}(B_n)$$

Most of our applications were of the following form. Let Y be another discrete RV. Define $B_n = \{Y = n\}$ where n ranges over the range of Y. Then

$$\mathbf{E}[X] = \sum_n \mathbf{E}[X|Y = n]\mathbf{P}(Y = n)$$

Now suppose X and Y are continuous random variables. We want to condition on $Y = y$. We cannot do this since $\mathbf{P}(Y = y) = 0$. How can we make sense of something like $\mathbf{P}(a \leq X \leq b | Y = y)$? We can define it by a limiting process:

$$\lim_{\epsilon \to 0} \mathbf{P}(a \leq X \leq b | y - \epsilon \leq Y \leq y + \epsilon)$$

Now let $f(x, y)$ be the joint pdf of X and Y.

$$\mathbf{P}(a \leq X \leq b | y - \epsilon \leq Y \leq y + \epsilon) = \frac{\int_a^b \left(\int_{y-\epsilon}^{y+\epsilon} f(u, w)\, dw \right) du}{\int_{-\infty}^{\infty} \left(\int_{y-\epsilon}^{y+\epsilon} f(u, w)\, dw \right) du}$$

Assuming f is continuous and ϵ is small,

$$\int_{y-\epsilon}^{y+\epsilon} f(u, w)\, dw \approx 2\epsilon f(u, y)$$

So the above just becomes

$$\frac{\int_a^b 2\epsilon f(u, y) du}{\int_{-\infty}^{\infty} 2\epsilon f(u, y) du} = \int_a^b \frac{f(u, y)}{f_Y(y)}\, du$$

This motivates the following definition:

Definition 6. *Let X, Y be jointly continuous RV's with pdf $f_{X,Y}(x, y)$. The conditional density of X given $Y = y$ is*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, if\ f_Y(y) > 0$$

When $f_Y(y) = 0$ we can just define it to be 0. We also define

$$\mathbf{P}(a \leq X \leq b | Y = y) = \int_a^b f_{X|Y}(x|y)\, dx$$

We have made the above definitions. We could have defined $f_{X|Y}$ and $\mathbf{P}(a \leq X \leq b | Y = y)$ as limits and then proved the above as theorems.

What happens if X and Y are independent? Then $f(x, y) = f_X(x) f_Y(y)$. So $f_{X|Y}(x|y) = f_X(x)$ as we would expect.

Example (X,Y) is uniformly distributed on the triangle with vertices $(0,0),(0,1)$ and $(1,0)$. Find the conditional density of X given Y.

The joint density is 2 on the triangle.

$$f_Y(y) = 2\int_0^{1-y} dx = 2(1-y), \quad 0 \le y \le 1$$

And we have

$$f_{X|Y}(x|y) = \frac{2}{2(1-y)} = \frac{1}{1-y}, \quad 0 \le x \le 1-y$$

So given $Y = y$, X is uniformly distributed on $[0, 1-y]$.

The conditional expectation is defined in the obvious way

Definition 7.

$$\mathbf{E}[X|Y=y] = \int x\, f_{X|Y}(x|y)\, dx$$

Note that $\mathbf{E}[X|Y=y]$ is a function of y. In our example, $\mathbf{E}[X|Y=y] = \frac{1}{2}(1-y)$.

Example Let X, Y be independent,each having an exponential distribution with the same λ. Let $Z = X+Y$. Find $f_{Z|X}, f_{X|Z}$, $\mathbf{E}[Z|X=x]$ and $\mathbf{E}[X|Z=z]$.

First we need to find the joint density of X and Z. We use change of variables. Let $U = X, W = X+Y$. The inverse is $x = u, y = w-u$.The Jacobian is

$$J(u,w) = det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \end{pmatrix} = det \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1$$

We have $f_{X,Y}(x,y) = \lambda^2 \exp(-\lambda(x+y))$ for $x, y \ge 0$. So

$$f_{X,Z}(x,z) = \begin{cases} \lambda^2 e^{-\lambda z}, & \text{if } 0 \le x \le z \\ 0, & \text{otherwise} \end{cases}$$

It is convenient to write the condition on x, z as $1(0 \le x \le z)$. This notation means the function is 1 if $0 \le x \le z$ is satisfied and 0 if it is not. So $f_{X,Z}(x,z) = \lambda^2 e^{-\lambda z}\, 1(0 \le x \le z)$. So we have for $x \ge 0$,

$$f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_X(x)} = \frac{\lambda^2 e^{-\lambda z}\, 1(0 \le x \le z)}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(z-x)}\, 1(0 \le x \le z)$$

Using this and the sub $u = z - x$, we find

$$\mathbf{E}[Z|X = x] = \int_x^\infty z\,\lambda e^{-\lambda(z-x)}\,dz = \int_0^\infty (u+x)\lambda e^{-\lambda u}\,du = x + \frac{1}{\lambda}$$

For the other one, we first find the marginal for Z:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^\infty f_{X,Z}(x,z)\,dx = \int_{-\infty}^\infty \lambda^2 e^{-\lambda z}\,1(0 \le x \le z)\,dx \\ &= \int_0^z \lambda^2 e^{-\lambda z}\,dx = \lambda^2 z e^{-\lambda z} \end{aligned}$$

So we have

$$f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_Z(z)} = \frac{\lambda^2 e^{-\lambda z}\,1(0 \le x \le z)}{\lambda^2 z e^{-\lambda z}} = \frac{1}{z}1(0 \le x \le z)$$

So given that $Z = z$, X is uniformly distributed on $[0, z]$. So $\mathbf{E}[X|Z = z] = z/2$.

Recall the partition theorem for discrete RV's X and Y,

$$\mathbf{E}[Y] = \sum_n \mathbf{E}[Y|X = n]\mathbf{P}(X = n)$$

For continuous random variables we have

Theorem 6. *Let X, Y be jointly continuous random variables. Then*

$$\mathbf{E}[Y] = \int \mathbf{E}[Y|X = x]\,f_X(x)\,dx$$

where the integral is over the range of x where $f_X(x) > 0$, i.e., the range of X.

Proof. Recall the definition:

$$\begin{aligned} \mathbf{E}[Y|X = x] &= \int y\,f_{Y|X}(y|x)\,dy \\ &= \int y\,\frac{f_{Y,X}(y,x)}{f_X(x)}\,dy \end{aligned}$$

So

$$\int \mathbf{E}[Y|X = x]\,f_X(x)\,dx = \int \left[\int y\,f_{X,Y}(x,y)dx\right]dy = \mathbf{E}[Y]$$

□

Recall that the partition theorem was useful when it was hard to compute the expected value of Y, but easy to compute the expected value of Y given that some other random variable is known.

There was another partition theorem for discrete random variable that gave a formula for $P(B)$ where B is an event in terms of conditional probabilities of B. Here is a special case where B and the partition both come from random variables. Let X and Y be discrete RV's. Then

$$P(a \leq Y \leq b) = \sum_n P(a \leq Y \leq b | X = n)\mathbf{P}(X = n)$$

For continuous random variables we have

Theorem 7. *Let X, Y be jointly continuous random variables. Then*

$$P(a \leq X \leq b) = \int P(a \leq X \leq b | Y = y)\, f_Y(y)\, dy$$

where

$$P(a \leq X \leq b | Y = y) = \int_a^b f_{X|Y}(x|y) dx$$

Proof. Straightforward using the above definitions. □

Example: Quality of lightbulbs varies because ... For fixed factory conditions, the lifetime of the lightbulb has an exponential distribution. We model this by assuming the parameter λ is uniformly distributed between 5×10^{-4} and 8×10^{-4}. Find the mean lifetime of a lightbulb and the pdf for its lifetime. Is it exponential?

Example: Let X, Y be independent standard normal RV's. Let $Z = X + Y$. Find $f_{Z|X}$, $f_{X|Z}$, $\mathbf{E}[Z|X = x]$ and $\mathbf{E}[X|Z = z]$.