Joint distributions
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

Today we'll look at joint random variables and joint distributions in detail.

Product distributions. If Ω_1 and Ω_2 are sample spaces, then their distributions $P : \Omega_1 \to \mathbf{R}$ and $P : \Omega_2 \to \mathbf{R}$ determine a product distribution on $P : \Omega_1 \times \Omega_2 \to \mathbf{R}$ as follows. First, if E_1 and E_2 are events in Ω_1 and Ω_2, then define $P(E_1 \times E_2)$ to be $P(E_1)P(E_2)$. That defines P on "rectangles" in $\Omega_1 \times \Omega_2$. Next, extend that to countable disjoint unions of rectangles. There's a bit of theory to show that what you get is well-defined. Extend to complements of those unions. Again, more theory. Continue by closing under countable unions and complements. A lot more theory. But it works. That theory forms part of the the topic called measure theory which is included in courses on real analysis.

When you have the product distribution on $\Omega_1 \times \Omega_2$, the random variables X and Y are independent, but there are applications where we have a distribution on $\Omega_1 \times \Omega_2$ that differs from the product distribution, and for those, X and Y won't be independent. That's the situation we'll look at now.

Definition. A *joint random variable* (X, Y) is a random variable on any sample space Ω which is the product of two sets $\Omega_1 \times \Omega_2$.

Joint random variables do induce probability distributions on Ω_1 and on Ω_2. If $E \subseteq \Omega_1$, define $P(E)$ to be the probability in Ω of the set $E \times \Omega_2$. That defines $P : \Omega_1 \to \mathbf{R}$ which satisfies the axioms for a probability distributions. Similarly, you can define $P : \Omega_2 \to \mathbf{R}$ by declaring for $F \subseteq \Omega_2$ that $P(F) = P(\Omega_1 \times F)$. If it happens that $P(E \times F) = P(E)P(F)$ for all $E \subseteq \Omega_1$ and $F \subseteq \Omega_2$, then the distribution on $\Omega = \Omega_1 \times \Omega_2$ is the product of the distributions on Ω_1 and Ω_2. But that doesn't always happen, and that's what we're interested in.

A discrete example. Deal a standard deck of 52 cards to four players so that each player gets 13 cards at random. Let X be the number of spades that the first player gets and Y be the number of spades that the second player gets. Let's compute the probability mass function $f(x, y) = P(X{=}x \text{ and } Y{=}y)$, that probability that the first player gets x spades and the second player gets y spades.

There are $\binom{52}{13}\binom{39}{13}$ hands that can be dealt to these two players. There are $\binom{13}{x}\binom{39}{13-x}$ hands for the first player with exactly x spades, and with the remaining deck there are $\binom{13-x}{y}\binom{26+x}{13-y}$ hands for the second player with exactly y spades. Thus, there are

$$\binom{13}{x}\binom{39}{13-x}\binom{13-x}{y}\binom{26+x}{13-y}$$

ways of dealing hands to those two players with x and y spades. Thus,

$$f(x, y) = \frac{\binom{13}{x}\binom{39}{13-x}\binom{13-x}{y}\binom{26+x}{13-y}}{\binom{52}{13}\binom{39}{13}}.$$

You can write that expression in terms of trinomial coefficients, or in terms of factorials, but we'll leave it as it is. If you were a professional Bridge player, you might want to see a table of the values of $f(x, y)$.

It's intuitive that X and Y are not independent joint random variables. The more spades the first player has, the fewer the second will probably have. In order to show they're not independent, it's enough to find particular values of x

and y so that $f(x,y) = P(X{=}x \text{ and } Y{=}y)$ does not equal the product of $f_X(x) = P(X{=}x)$ and $f_Y(y) = P(Y{=}y)$. For example, we know they can't both get 7 spades since there are only 13 spades in all, so $f(7,7) = 0$. But since the first player can get 7 spades, $f_X(7) > 0$, also the second player can, so $f_Y(y) > 0$. Therefore $f(7,7) \neq f_Y(7)f_Y(7)$. We've proven that X and Y are not independent.

A continuous example. Let's choose a point from inside a triangle Δ uniformly at random. Let's take Δ to be half the unit square, namely the half with vertices at $(0,0)$, $(1,0)$, and $(1,1)$. Then

$$\Delta = \{(x,y) \,|\, 0 \leq y \leq x \leq 1\}.$$

The area of this triangle is $\frac{1}{2}$, so we can find the probability of any event E, a subset of Δ, simply by doubling its area. The joint probability density function is constantly $\frac{1}{2}$ inside Δ and 0 outside.

$$f(x,y) = \begin{cases} 2 & \text{if } 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

A continuous joint random variable (X,Y) is determined by its *cumulative distribution* function

$$F(x,y) = P(X \leq x \text{ and } Y \leq y).$$

We'll figure out it's value for this example.

First, if either x or y is negative, then the event $(X \leq x \text{ and } Y \leq y)$ completely misses the triangle, so it's actually empty, so its probability is 0.

Second, if x is between 0 and 1 and $x \leq y$, then the event $(X \leq x \text{ and } Y \leq y)$ is a triangle of width and height x, so its area is $\frac{1}{2}x^2$. Doubling the area gives a probability of x^2.

Third, if x is between 0 and 1 and $x > y$, then the point (x,y) lies inside the triangle, and the event $(X \leq x \text{ and } Y \leq y)$ is a trapezoid. Its bottom length is x, top length is $x-y$, and height is y, so its area is $xy - y^2/2$. Therefore, the probability of this event is $2xy - y^2$.

There remain two more cases to be considered, but their analyses are omitted.

We can summarize the cumulative distribution function as

$$F(x,y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ x^2 & \text{if } 0 \leq x \leq 1 \text{ and } x \leq y \\ 2xy - y^2 & \text{if } 0 \leq x \leq 1 \text{ and } x > y \\ 2y - y^2 & \text{if } x > 1 \text{ and } 0 \leq y \leq 1 \\ 1 & \text{if } x > 1 \text{ and } y > 1 \end{cases}$$

Generally speaking, joint cumulative distribution functions aren't used as much as joint density functions. Typically, joint c.d.f.'s are much more complicated to describe, just as in this example.

Joint distributions and density functions. Density functions are the usual way to describe joint continuous real-valued random variables.

Let X and Y be two continuous real-valued random variables. Individually, they have their own cumulative distribution functions

$$F_X(x) = P(X \leq x) \qquad F_Y(y) = P(Y \leq y),$$

whose derivatives, as we know, are the probability density functions

$$f_X(x) = \frac{d}{dx}F_X(x) \qquad f_Y(y) = \frac{d}{dy}F_Y(y).$$

Furthermore, the cumulative distribution functions can be found by integrating the density functions

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt.$$

There is also a joint cumulative distribution function for (X,Y) defined by

$$F(x,y) = P(X \leq x \text{ and } Y \leq y).$$

The joint probability density function $f(x,y)$ is found by taking the derivative of F twice, once with respect to each variable, so that

$$f(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y).$$

(The notation ∂ is substituted for d to indicate that there are other variables in the expression that are held constant while the derivative is taken with respect to the given variable.) The joint cumulative distribution function can be recovered from the joint density function by integrating twice

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,dt\,ds.$$

(When more that one integration is specified, the inner integral, namely $\int_{-\infty}^{y} f(s,t)\,dt$, is written without parentheses around it, even though the parentheses would help clarify the expression.)

Furthermore, the individual cumulative distribution functions are determined by the joint distribution function.

$$\begin{aligned} F_X(x) &= P(X \le x \text{ and } Y \le \infty) \\ &= \lim_{y\to\infty} F(x,y) = F(x,\infty) \end{aligned}$$

$$\begin{aligned} F_Y(y) &= P(X \le \infty \text{ and } Y \le y) \\ &= \lim_{x\to\infty} F(x,y) = F(\infty,y) \end{aligned}$$

Likewise, the individual density functions can be found by integrating joint density function.

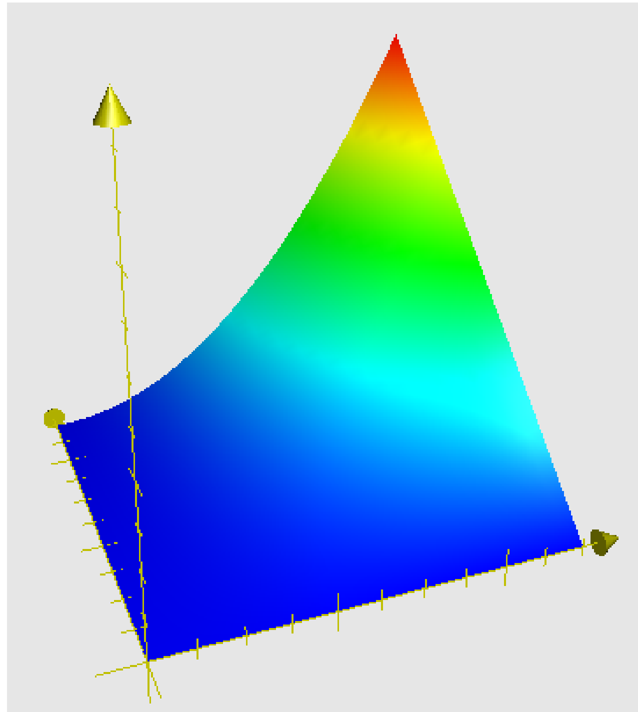
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad f_Y(x) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

These individual density functions f_X and f_y are often called *marginal density functions* to distinguish them from the joint density function $f_{(X,Y)}$. Likewise the corresponding individual cumulative distribution functions F_X and F_Y are called *marginal cumulative distribution functions* to distinguish them form the joint c.d.f $F_{(X,Y)}$.

Another continuous example. The last example was a uniform distribution on a triangle. For this example, we'll go back to the unit square, but make the distribution nonuniform. We'll describe the distribution via a joint density function

$$f(x,y) = 6x^2y$$

if (x,y) is in the unit square, that is, x and y are both between 0 and 1. Outside that square $f(x,y)$ is 0. The graph $z = f(x,y)$ of this function is a surface sitting above the unit square.



The volume under that surface is 1, as it has to be for f to be a probability density function. To find the probability that (X,Y) lies in an event E, a subset of the unit square, just find the volume of the solid above E and below the surface $z = f(x,y)$. That's a double integral

$$P((X,Y) \in E) = \iint_E f(x,y)\,dx\,dy.$$

For example, suppose we want to find the probability $P(0 \le X \le \frac{3}{4} \text{ and } \frac{1}{3} \le Y \le 1)$. The event E is $(0 \le X \le \frac{3}{4} \text{ and } \frac{1}{3} \le Y \le 1)$, and that describes a rectangle. The double integral giving the probability is

$$P(E) = \int_0^{3/4} \int_{1/3}^{1} 6x^2y\,dy\,dx.$$

Evaluate that integral from the inside out. The inner integral is

$$\begin{aligned} \int_{1/3}^{1} 6x^2y\,dy &= 6x^2 \int_{1/3}^{1} y\,dy \\ &= 6x^2(y^2/2)\Big|_{y=1/3}^{1} \\ &= 6x^2(1-\tfrac{1}{9})/2 = \tfrac{8}{3}x^2 \end{aligned}$$

Next, evaluate the outer integral.

$$\begin{aligned} P(E) &= \int_0^{3/4} \tfrac{8}{3}x^2\,dx \\ &= \tfrac{8}{9}x^3 \Big|_0^{3/4} = \tfrac{3}{8} \end{aligned}$$

Let's do more with this example. We're given the joint density function $f(x, y)$. Let's find the marginal density functions for X and Y.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\ &= \int_0^1 6x^2y\,dy \\ &= 6x^2y^2/2 \Big|_{y=0}^{1} \\ &= 3x^2 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= \int_0^1 6x^2y\,dx \\ &= 2x^3y \Big|_{x=0}^{1} \\ &= 2y \end{aligned}$$

In summary, the joint density is $f(xy) = 6x^2y$ over the unit square. The marginal density functions are $f_x(x) = 3x^2$ and $f_Y(y) = 2y$.

Note: it so happens in this example that the joint density function $f(x, y)$ is the product of the two marginal density functions $f_X(x,y)f_Y(x,y)$. That doesn't always happen. But it does happen when the random variables X and Y are independent, which is discussed next.

Independent continuous random variables. In the case of continuous real random variables, we can characterize independence in terms of density functions. Random variables X and Y will be independent when the events $X \leq x$ and $Y \leq y$ are independent for all values of x and y. That means

$$P(X \leq x \text{ and } Y \leq y) = P(X \leq x)\,P(Y \leq y),$$

from which it follows that the joint cumulative distribution function is the product of the marginal cumulative distribution functions

$$F(x,y) = F_X(x)\,F_Y(y).$$

Take the partial derivatives $\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y}$ of both sides of that equation to conclude

$$f(x,y) = f_X(x)\,f_Y(y),$$

the joint probability density function is the product of the marginal density functions

Another example. We've seen two examples so far. One was when the probability was uniform over a triangle, and the other had X and Y independent. This example is of a nonuniform probability where the variables aren't independent.

Let

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

You can show that the double integral

$$\int_0^1 \int_0^1 (x+y)\,dy\,dx$$

equals 1 as follows. First, evaluate the inner integral.

$$\int_0^1 (x+y)\,dy = xy + \tfrac{1}{2}y^2 \Big|_{y=0}^{1} = x + \tfrac{1}{2}$$

Then substitute that in the outer integral and evaluate it.

$$\int_0^1 (x+\tfrac{1}{2})\,dx = \tfrac{1}{2}x^2 + \tfrac{1}{2} \Big|_{x=0}^{1} = 1$$

Therefore f is a probability density function.

Next, let's compute the two marginal density functions f_X and f_Y, so X and Y. In fact, the first computation we performed, evaluating the inner integral, gave us the marginal density function for X. In general, the marginal density function for X can be found as the integral

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

which in this case was, for $x \in [0, 1]$,

$$f_X(x) = \int_0^1 (x + y)\, dy = x + \tfrac{1}{2}.$$

Also for this example, since $f(x, y)$ is symmetric in x and y, therefore the marginal density function for Y is $f_Y(y) = y + \frac{1}{2}$ for $y \in [0, 1]$.

For X and Y to be independent, the joint density function $f(x, y)$ would have to equal the product $f_X(x) f_Y(y)$ of the two marginal density functions, but

$$x + y \neq (x + \tfrac{1}{2})(y + \tfrac{1}{2}).$$

Therefore, X and Y are not independent.

Later, we'll develop the concept of correlation which will quantify how related X and Y are. Independent variables will have 0 correlation, but if a larger value of X indicates a larger value of Y, then they will have a positive correlation.

Math 217 Home Page at
`http://math.clarku.edu/~djoyce/ma217/`