14.30 Introduction to Statistical Methods in Economics

Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 7

Konrad Menzel

February 26, 2009

1 Joint Distributions of 2 Random Variables X, Y (ctd.)

1.1 Continuous Random Variables

If X and Y are continuous random variables defined over the same sample space S. The joint p.d.f. of (X, Y), $f_{XY}(x, y)$ is a function such that for any subset A of the (x, y) plane,

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

As in the single-variable case, this density must satisfy

$$f_{XY}(x, y) \geq 0 \quad \text{for each } (x, y) \in \mathbb{R}^2$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)dxdy = 1$$

Note that

- any single point has probability zero
- any one-dimensional curve on the plane has probability zero

Example 1 *A UFO appears at a random location over Wyoming, which - ignoring the curvature of the Earth - can be described quite accurately as a rectangle of 276 times 375 miles. The position of the UFO is uniformly distributed over the entire state, and can be expressed as a random longitude X (ranging from -111 to -104 degrees) and latitude Y (with values between 41 and 45 degrees).*
This means that the joint density of the coordinates is given by

$$f_{XY}(x, y) = \begin{cases} \frac{1}{28} & \text{if } -111 \leq x \leq -104 \text{ and } 41 \leq y \leq 45 \\ 0 & \text{otherwise} \end{cases}$$

If the UFO can be seen from a distance of up to 40 miles, what is the probability that it can be seen from Casper, WY (which is roughly in the middle of the state)?
Let's look at the problem graphically: This suggests that the set of locations for which the UFO can be seen from Casper can be described as a circle with a 40-mile radius around Casper. Also, for the uniform density, the probability of the UFO showing up in a region A (i.e. the integral of a constant density over

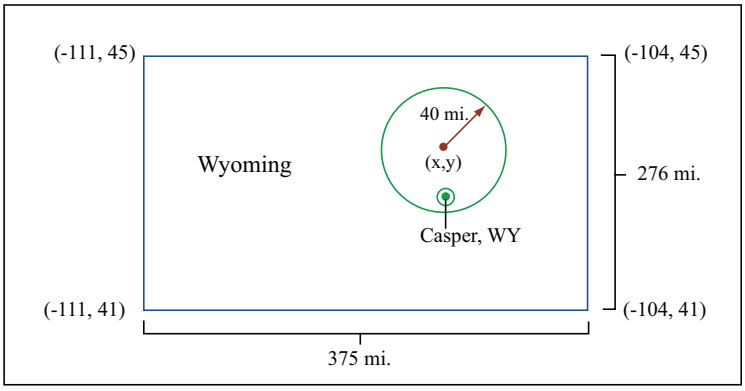

Image by MIT OpenCourseWare.

Figure 1: The UFO at (x, y) can be seen from Casper, WY

A) of the state is proportional to the area of A. Therefore, we don't have to do any integration, but finding the probability reduces to a purely geometric exercise. We can calculate the probability as

$$P(\text{"less than 40 miles from Casper"}) = \frac{\text{Area("less than 40 miles from Casper")}}{\text{Area("All of Wyoming")}} = \frac{40^2\pi}{375 \cdot 276} \approx 4.9\%$$

You should notice that for the uniform distribution, there is often no need to perform complicated integration, but you may be able to treat everything as a purely geometric problem.

Unlike in the last example, typically, there's no way around integrating the density function in order to obtain probabilities, since any nonconstant density re-weights different regions in terms of probability mass. We'll do this in the clean, systematic fashion in the following example:

Example 2 *Suppose you have 2 spark plugs in your lawn mower, and let X be the life of spark plug 1, and Y the life of spark plug 2. Suppose we can describe the distribution by*

$$f_{XY}(x, y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Figure 2 on page 2 shows what the joint density looks like.

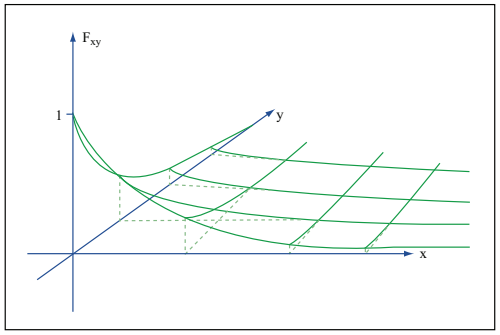

Image by MIT OpenCourseWare.

Figure 2: Joint Density of Lives X and Y of Sparkplugs 1 and 2

In fact, this density can be derived from the assumption that the spark plugs fail independently of one another at a fixed rate λ, which doesn't change over their lifetime.
If the lawn mower is going to work as long as either spark plug works, what is the probability that the lawn mower fails within 1000 hours?

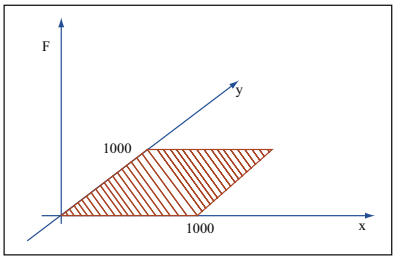

Image by MIT OpenCourseWare.

Figure 3: The Event "Lawn Mower Fails before 1000 hrs." in first situation

$$\begin{aligned} P(X \leq 1000, Y \leq 1000) &= \int_0^{1000} \int_0^{1000} \lambda^2 e^{-\lambda(x+y)} dy dx \\ &= \int_0^{1000} \int_0^{1000} \lambda^2 e^{-\lambda x} e^{-\lambda y} dy dx \\ &= \int_0^{1000} \lambda e^{-\lambda x} \left(\int_0^{1000} \lambda e^{-\lambda y} dy \right) dx \\ &= \int_0^{1000} \lambda e^{-\lambda x} \left(1 - e^{-1000\lambda}\right) dx \\ &= \left(1 - e^{-1000\lambda}\right)^2 \end{aligned}$$

What is that probability if the second spark plug is only used if the first one fails, i.e. how do we calculate $P(X + Y \leq 1000)$? Note that this only changes the "event" we care about, i.e. the region of $\mathbb{R}^2$ we

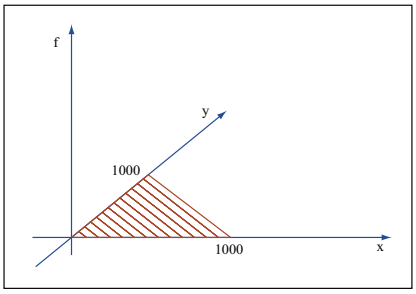

Image by MIT OpenCourseWare.

Figure 4: The Event "Lawn Mower Fails before 1000 hrs." in second situation

integrate over, but we still integrate the same density.

$$\begin{aligned} P(X + Y \leq 1000) &= \int_0^{1000} \left[\int_0^{1000-x} \lambda^2 e^{-\lambda x} e^{-\lambda y} dy \right] dx \\ &= \int_0^{1000} \lambda e^{-\lambda x} \left[\int_0^{1000-x} \lambda e^{-\lambda y} dy \right] dx \end{aligned}$$

$$
\begin{aligned}
&= \int_0^{1000} \lambda e^{-\lambda x}\left[1 - e^{-\lambda(1000-x)}\right] dx \\
&= \int_0^{1000} \lambda\left[e^{-\lambda x} - e^{-1000\lambda}\right] dx \\
&= 1 - e^{-1000\lambda} - 1000\lambda e^{-1000\lambda} = 1 - (1 + 1000\lambda)e^{-1000\lambda}
\end{aligned}
$$

Again, events over continuous bivariate random variables correspond to areas in the plane, and we find probabilities by integrating the density over those areas.

2 Joint c.d.f. of 2 Random Variables X, Y

I'll just give definitions. We are not going to use this a lot in this class, but you should have seen this.

Definition 1 *The joint c.d.f. for random variables (X, Y) is defined as the function $F_{XY}(x, y)$ for $(x, y) \in \mathbb{R}^2$*

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

We compute probabilities from joint c.d.f.s as follows

$$P(a \leq X \leq b, c \leq Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

We have to add in the last term because in a sense, it got subtracted off twice before.
Joint c.d.f.s are related to p.d.f.s in the following way: for continuous random variables,

$$
\begin{aligned}
F_{XY}(x, y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u, v)dudv \\
f_{XY}(x, y) &= \frac{\partial^2}{\partial y \partial x} F_{XY}(x, y)
\end{aligned}
$$

In the discrete case,

$$F_{XY}(x, y) = \sum_{u \leq x}\sum_{v \leq y} f_{XY}(u, v)$$

3 Marginal p.d.f.s

If we have a *joint* distribution, we may want to recover distribution of one variable X.
If X and Y are *discrete* random variables with joint p.d.f. F_{XY}, then

$$
\begin{aligned}
f_X(x) &= \sum_{\text{all } y} f_{XY}(x, y) \\
f_Y(y) &= \sum_{\text{all } x} f_{XY}(x, y)
\end{aligned}
$$

If X and Y are continuous, we'll essentially replace summation by integration, so that

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y)dy \\
f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y)dx
\end{aligned}
\tag{1}
$$

Example 3 *This example is based on real-world data extra-marital affairs collected by the* Redbook *magazine in 1977.*[1] *In the survey, individuals were asked to rate their marriage on a scale from 1 (unhappy) to 3 (happy), and to report the number of extra-marital affairs, divided by the number of years married. For now let's look at the joint distribution of "marriage quality", X, with duration of marriage in years, Y. We can start from the "cell" probabilities given by the joint p.d.f., and then fill in the marginal p.d.f.s on the left and at the bottom of the table:*

It is interesting to note that, even though the marginal distributions are relatively even, the joint distribution seems to be concentrated along the bottom left/top right diagonal, with the joint p.d.f. taking much lower values in the top-left and bottom-right corners of the table.

		Y			
	f_{XY}	1	8	12	f_X
	1	4.66%	11.48%	12.98%	29.12%
X	2	5.16%	14.81%	12.31%	32.28%
	3	13.48%	16.47%	8.65%	38.60%
	f_Y	23.30%	42.76%	33.94%	100.00%

Example 4 *Recall the example with the two spark plugs from last time. The joint p.d.f. was*

$$f_{XY}(x,y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_0^{\infty} \lambda^2 e^{-\lambda(x+y)}dy \\ &= \lambda e^{-\lambda x}\int_0^{\infty} \lambda e^{-\lambda y}dy = \lambda e^{-\lambda x}[1-0] = \lambda e^{-\lambda x} \end{aligned}$$

Similarly,

$$f_Y(y) = \lambda e^{-\lambda y}$$

4 Independence

Recall that we said that two *events* A and B were independent if $P(AB) = P(A)P(B)$. Now we'll define a similar notion for random variables.

Definition 2 *We say that the random variables X and Y are* independent *if for* any *regions $A, B \subset \mathbb{R}$,*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Note that this requirement is very strict: we are looking at events of the type $X \in A$ and $Y \in B$ and require that *all* pairs of them are mutually independent.
This definition is not very practical per se, because it may be difficult to check, however if X and Y are independent, it follows from the definition that in particular

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$$

From this, it is possible to derive the following condition which is usually much easier to verify

[1] Data available at http://pages.stern.nyu.edu/ wgreene/Text/Edition6/tablelist6.htm

Proposition 1 *X and Y are independent if and only if their joint and marginal p.d.f.s satisfy*

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

PROOF: For discrete random variables, this follows directly from applying the definition to $A = \{x\}$ and $B = \{y\}$. For continuous random variables, we can show that if X and Y are independent, we can differentiate the equation

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

on both sides in order to obtain

$$f_{XY}(x,y) = \frac{\partial^2}{\partial y \partial x} F_{XY}(x,y) = \frac{\partial^2}{\partial y \partial x} \left[F_X(x)F_Y(y)\right] = \frac{\partial}{\partial y} f_X(x)F_Y(y) = f_X(x)f_Y(y)$$

Conversely, if the product of the marginal p.d.f.s equals the joint p.d.f., we can integrate

$$\begin{aligned} P(X \in A, Y \in B) &= \int_A \int_B f_{XY}(x,y)dydx = \int_A \int_B f_X(x)f_Y(y)dydx \\ &= \left(\int_A f_X(x)dx\right)\left(\int_B f_Y(y)dy\right) \end{aligned}$$

so that the condition on the marginals implies independence, and we've proven both directions of the equivalence □

Example 5 *Going back to the data on extra-marital affairs, remember that we calculated the marginal p.d.f.s of reported "marriage quality", X, and years married, Y as*

$$f_X(1) = 29.12\%, \quad f_X(2) = 32.28\%, \quad f_X(3) = 38.60\%$$

and

$$f_Y(1) = 23.30\%, \quad f_Y(8) = 42.76\%, \quad f_Y(12) = 33.94\%$$

What should the joint distribution look like if the two random variables were in fact independent? E.g.

$$\tilde{f}_{XY}(3,1) = f_X(3)f_Y(1) = 38.60\% \cdot 23.30\% = 8.99\%$$

The actual value of the joint p.d.f. at that point was 13.48*, so that apparently, the two variables are not independent. We can now fill in the remainder of the table under the assumption of independence: Comparing this to our last table we see some systematic discrepancies - in particular, the constructed*

		Y			
	$\tilde{f}_{XY}$	1	8	12	f_X
	1	6.78%	12.45%	9.88%	29.12%
X	2	7.52%	14.81%	10.96%	32.28%
	3	8.99%	16.50%	13.10%	38.60%
	f_Y	23.30%	42.76%	33.94%	100.00%

joint p.d.f. $\tilde{f}_{XY}$ is not as strongly concentrated on the diagonal, which seemed to be a noteworthy feature of the actual joint p.d.f..
But does this really mean that X and Y are not independent? One caveat is that we calculated the probabilities in the joint p.d.f. from a sample *of "draws" from the underlying distribution, so there is some uncertainty over how accurately we could measure the true cell probabilities. In the last part of the class, we will see a method of testing formally whether the differences between the "constructed" and the actual p.d.f. are large enough to suggest that the random variables X and Y are in fact* not *independent.*

Example 6 *Recall the example with the two spark plugs from last time. The joint p.d.f. was*

$$f_{XY}(x,y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and in the last section we derived the marginal p.d.f.s

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} \\ f_Y(y) &= \lambda e^{-\lambda y} \end{aligned}$$

Therefore, the product is

$$f_X(x) f_Y(y) = \lambda^2 e^{-\lambda x} e^{-\lambda y} = f_{XY}(x,y)$$

so that the lives of spark plug 1 and 2 are independent.

Remark 1 *The condition on the joint and marginal densities for independence can be restated as follows for continuous random variables: Whenever we can factor the joint p.d.f. into*

$$f_{XY}(x,y) = g(x)h(y)$$

where $g(\cdot)$ depends only *on x and $h(\cdot)$ depends only on y, then X and Y are independent. In particular, we don't have to calculate the marginal densities explicitly.*

Example 7 *Say, we have a joint p.d.f.*

$$f_{XY}(x,y) = \begin{cases} ce^{-(x+2y)} & \text{if } x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then we can choose e.g. $g(x) = ce^{-x}$ and $h(y) = e^{-2y}$, and even though these aren't proper densities, this is enough to show that X and Y are independent.

Example 8 *Suppose we have the joint p.d.f.*

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

Can X and Y be independent?
Even though in either case (i.e. whether $x^2 \leq y \leq 1$ holds or whether it doesn't) the p.d.f. factors into functions of x and y (for the zero part, that's trivially true), we can also see that the support of X depends on the value of Y, and therefore, X and Y can't be independent - e.g. if $X \geq \frac{1}{2}$, we must have $Y \geq \frac{1}{4}$, so that

$$P\left(X \geq \frac{1}{2}, Y \leq \frac{1}{4}\right) = 0 < P\left(X \geq \frac{1}{2}\right) P\left(Y \leq \frac{1}{4}\right)$$

Note that the joint support of two random variables has to be rectangular (possibly all of $\mathbb{R}^2$) in order for X and Y to be independent: if it's not, for some realizations of X, certain values of Y would be ruled out which could occur otherwise. But if that were true, knowing X does give us information about Y, so they can't be independent. However, this condition on the support alone does *not* imply independence.