Economics 520, Fall 2006

Lecture Note 7: Joint Distributions, Conditional Distributions, and Independence of Random Variables (CB 4.1-4.2, 4.6)

Definition 1 *A N–dimensional random vector is a function from the sample space to R^N.*

Thus a random variable can be thought of as a collection of random variables. We sometimes just use "random variable" loosely to denote any random vector (not just a scalar).

Definition 2 *Let (X, Y) be a discrete bivariate random vector. Then the function $f_{X,Y}(x, y)$ from R^2 to R, defined by*

$$f_{X,Y}(x, y) = Pr\left(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}\right),$$

is the joint probability mass function of (X, Y).

Example: Toss a coin three times. The sample space is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

An example of a bivariate random variable is (X, Y), where $X(\omega)$ is the number of heads in the first two tosses, and $Y(\omega)$ is the number of heads in the last two tosses. Assuming the coin is a fair one, and the probability of heads is $1/2$ at every toss, with different tosses independent events, the joint PMF is

$$f_{X,Y}(x, y) = \begin{cases} 1/8 & \text{if } (x, y) = (0, 0) & (\omega \in \{TTT\}), \\ 1/8 & \text{if } (x, y) = (0, 1) & (\omega \in \{TTH\}), \\ 1/8 & \text{if } (x, y) = (1, 0) & (\omega \in \{HTT\}), \\ 2/8 & \text{if } (x, y) = (1, 1) & (\omega \in \{HTH, THT\}), \\ 1/8 & \text{if } (x, y) = (1, 2) & (\omega \in \{THH\}), \\ 1/8 & \text{if } (x, y) = (2, 1) & (\omega \in \{HHT\}), \\ 1/8 & \text{if } (x, y) = (2, 2) & (\omega \in \{HHH\}), \\ 0 & \text{otherwise.} \end{cases}$$

Although we have defined two random variables on Ω, we can define probability mass functions for each random variable in the usual way. Ignoring the definition of $Y(\omega)$, the PMF for x is

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0 & (\omega \in \{TTT, TTH\}), \\ 2/4 & \text{if } x = 1 & (\omega \in \{HTH, HTT, THH, THT\}), \\ 1/4 & \text{if } x = 2 & (\omega \in \{HHT, HHH\}), \\ 0 & \text{elsewhere.} \end{cases}$$

Similarly we can calculate the PMF for Y. When working with joint random variables we typically refer to these single–variable probability mass functions as marginal probability mass functions. In general they can be calculated from the joint PMF as

$$f_X(x) = Pr\left(\{\omega \in \Omega : X(\omega) = x\}\right)$$

$$= Pr\left(\{\omega \in \Omega : X(\omega) = x, -\infty < Y(\omega) < \infty\}\right) = \sum_y f_{X,Y}(x, y),$$

and similarly

$$f_Y(y) = \sum_x f_{X,Y}(x, y).$$

Given the joint PMF we can calculate probabilities involving both random variables. For example,

$$Pr(X \leq 1, Y \geq 1) = Pr\left(\{\omega \in \Omega : X(\omega) \leq 1, Y(\omega) \geq 1\}\right) = \sum_{x \leq 1, y \geq 1} f_{X,Y}(x, y)$$

$$= f(0, 1) + f(1, 1) + f(1, 2) = 1/2.$$

□

Alternatively we can have joint continuous random variables:

Definition 3 *Let (X, Y) be a continuous bivariate random vector. Then the function $f_{X,Y}(x, y)$ from R^2 to R is the joint probability density function of (X, Y) if it satisfies*

$$\int_A f_{X,Y}(x, y)dxdy = Pr(\omega \in \Omega | (X(\omega), Y(\omega)) \in A),$$

for all sets A.

Example: Pick a point randomly in the triangle $(0, 0), (0, 1), (1, 0)$. Let X be the x–coordinate and y be the y–coordinate. A reasonable choice for the joint PDF appears to be

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

for $\{(x, y) : 0 < x < 1, 0 < y < 1, x + y < 1\}$ and zero otherwise (that is, constant on the domain). What is the appropriate value for c? We know that the probability of the entire sample space is equal to one, so:

$$1 = Pr(\omega \in \Omega) = Pr((X, Y) \in \{(x, y) : 0 < x < 1, 0 < y < 1, x + y < 1\})$$

$$= \int_{\{(x,y)|0<x<1,0<y<1,x+y<1\}} f_{X,Y}(x, y)dxdy = \int_0^1 \int_0^{1-y} cdxdy$$

$$= \int_0^1 c \cdot (1 - y)dy = c/2.$$

Hence $c = 2$. (Note that we can write such integrals in different ways:

$$\int_{\{(x,y)|0<x<1,0<y<1,x+y<1\}} g(x, y)dxdy = \int_0^1 \int_0^{1-y} g(x, y)dxdy = \int_0^1 \int_0^{1-x} g(x, y)dydx,$$

depending on the order of integration, without changing the result.) The marginal PDF's are in this case defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy = \int_0^{1-x} 2dy = 2 - 2x,$$

for $0 < x < 1$ and zero otherwise. Similarly

$$f_Y(y) = 2 - 2y, \qquad 0 < y < 1, \text{ and zero otherwise.}$$

□.

Expectations are calculated in pretty much the same way as before, but now involving double sums or integrals:

$$E[r(X,Y)] = \int_x \int_y r(x,y) f_{X,Y}(x,y) dy dx,$$

for continuous bivariate random variables, and

$$E[r(X,Y)] = \sum_x \sum_y r(x,y) f_{X,Y}(x,y),$$

for discrete bivariate random variables.

Example (ctd): Consider the expected value of XY in the previous example:

$$E(XY) = \int_0^1 \int_0^{1-x} 2xy dy dx = \int_0^1 xy^2 \Big|_0^{1-x} dx = \int_0^1 x(1-x)^2 dx$$

$$= \int_0^1 (x^3 - 2x^2 + x) dx = \frac{1}{4}x^4 - \frac{2}{3}x^3 + \frac{1}{2}x^2 \Big|_0^1 = \frac{1}{12}.$$

We can express the expected value of X in two ways:

$$E(X) = \int \int x f_{X,Y}(x,y) dy dx,$$

or

$$E(X) = \int x f_X(x) dx,$$

where

$$f_X(x) = \int f_{X,Y}(x,y) dy.$$

Either way we get

$$E(X) = \int_0^1 \int_0^{1-x} 2x dy dx = \int_0^1 2xy \Big|_0^{1-x} dx = \int_0^1 (2x - 2x^2) dx$$

$$= x^2 - \frac{2}{3}x^3 \Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3}.$$

□.

Recall how the conditional probability of events was defined:

$$Pr(E_1|E_2) = \frac{Pr(E_1 \cap E_2)}{Pr(E_2)},$$

provided $Pr(E_2) > 0$. We can do the same thing for bivariate discrete random variables:

Definition 4 *Given two discrete random variables X and Y with joint probability mass function $f_{X,Y}(x, y)$, the conditional probability mass function for X given $Y = y$ is*

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

provided $f_Y(y) > 0$.

Example (ctd): Consider the distribution of X given that $Y = 1$. In that case we are conditioning on $\omega \in \{HTH, HHT, THT, TTH\}$

$$f_{X|Y}(x|Y = 1) = \begin{cases} 1/4 & x = 0 & (\omega \in \{TTH\}), \\ 2/4 & x = 1 & (\omega \in \{HTH, THT\}), \\ 1/4 & x = 2 & (\omega \in \{HHT\}), \\ 0 & \text{elsewhere.} \end{cases}$$

Note that this is the same as the marginal PMF of X. This is not true if we condition on $Y = 2$. In that case we condition on $\omega \in \{HHH, THH\}$:

$$f_{X|Y}(x|Y = 2) = \begin{cases} 1/2 & x = 1 & (\omega \in \{THH\}), \\ 1/2 & x = 2 & (\omega \in \{HHH\}), \\ 0 & \text{elsewhere.} \end{cases}$$

□

For continuous bivariate random variables things are a little more complicated because the probability that they take on a particular value is <u>zero</u>, and conditional probabilities are not defined when the probability of the conditioning event is zero. Nevertheless, the definition of conditional probability density functions is very similar to that of conditional probability mass functions:

Definition 5 *Given two continuous random variables X and Y with joint probability density function $f_{X,Y}(x, y)$, the conditional probability density function for X given $Y = y$ is*

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

To see why this works we first look at the conditional cumulative distribution function where we condition on $Y \in (y, y + \Delta)$, which does have positive probability for nonzero Δ:

$$Pr(X \leq x)|Y \in (y, y + \Delta)) = \frac{Pr(X < x, y < Y < y + \Delta)}{Pr(y < Y < y + \Delta)}$$

$$\frac{\int_{-\infty}^{x} \int_{y}^{y+\Delta} f_{X,Y}(u, v) dv du}{\int_{-\infty}^{\infty} \int_{y}^{y+\Delta} f_{X,Y}(u, v) dv du}.$$

Now take the limit as Δ goes to zero. In that case both numerator and denominator go to zero, so to get the limit we have to use l'Hôpital's rule and differentiate both numerator and denominator with respect to Δ. In that case we get:

$$F_{X|Y}(x|y) = \frac{\int_{-\infty}^{x} f_{X,Y}(u, y) du}{\int_{-\infty}^{\infty} f_{X,Y}(u, y) du} = \frac{\int_{-\infty}^{x} f_{X,Y}(u, y) du}{f_Y(y)}.$$

Take the derivative with respect to x to get the conditional probability density function:

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

Definition 6 *Let (X,Y) be a bivariate random vector with PMF/PDF $f_{X,Y}(x,y)$. Then the random variables X and Y are independent if*

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

where $f_X(x)$ and $f_Y(y)$ are the marginal PDF/PMF's of X and Y.

Result 1 (FACTORIZATION THEOREM) *The two random variables X and Y are independent if and only if the joint PDF/PMF can be written as*

$$f_{X,Y}(x,y) = g(x) \cdot h(y),$$

for all $-\infty < x < \infty$, and $-\infty < y < \infty$.

Proof:
If X and Y are independent, then we can choose $g(x) = f_X(x)$, and $h(y) = f_Y(y)$ and the result is trivial. Now suppose we can factor the joint density of X and Y as $f_{X,Y}(x,y) = g(x) \cdot h(y)$. Then the marginal density of X is

$$f_X(x) = \int_y f_{X,Y}(x,y)dy = g(x) \int_y h(y)dy.$$

Similarly

$$f_Y(y) = \int_x f_{X,Y}(x,y)dx = h(y) \int_x g(x)dx.$$

The product is equal to

$$f_X(x) \cdot f_Y(y) = g(x)h(y) \int_x g(x)dx \int_y h(y)dy$$

$$= f_{X,Y}(x,y) \cdot \int_x \int_y f_{X,Y}(x,y)dxdy = f_{X,Y}(x,y),$$

which completes the proof. $\square$

Result 2 *Let X and Y be two independent random variables. Then*
(i) $f_{X|Y}(x|y) = f_X(x)$, and
(ii) $f_{Y|X}(y|x) = f_Y(y)$.

Example (ctd): In the previous example we had

$$f_{X,Y}(x,y) = 2, \qquad 0 < x < 1,\ 0 < y < 1,\ x + y < 1, \qquad \text{and zero elsewhere.}$$

It may appear that X and Y are independent by the previous results: choose $g(x) = 2$ and $h(y) = 1$. The reason this does not work is because of the restrictions on the values of X and Y. To see this write

$$f_{X,Y}(x,y) = 2 \cdot 1\{0 < x < 1\} \cdot 1\{0 < y < 1\} \cdot 1\{x + y < 1\},$$

for all $-\infty < x < \infty$ and $-\infty < y < \infty$. This indicator function $1\{x + y < 1\}$ cannot be seperated into a function of x and a function of y and therefore X and Y are not independent. $\square$

Consider two random variables X and Y with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively. The covariance of X and Y is defined as

$$C(X,Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)],$$

and the correlation coefficient as

$$\rho_{X,Y} = C(X,Y)/(\sigma_X \sigma_Y).$$

The mean and variance of the sum $X + Y$ are

$$E[X + Y] = \mu_X + \mu_Y,$$

and

$$V(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2 \cdot C(X,Y).$$

If the two random variables are independent, the covariance is zero (note however, that zero covariance does **NOT** necessarily imply independence), and the variance simplifies to the sum of the variances $\sigma_X^2 + \sigma_Y^2$.

An important result on expectations of pairs of random variables is the following:

Result 3 (LAW OF ITERATED EXPECTATIONS) *The expected value of X is*

$$E(r(X,Y)) = E(E(r(X,Y)|X)).$$

Proof: Consider the continuous case. By definition,

$$E(r(X,Y)) = \int_x \int_y r(x,y) f_{YX}(x,y) dy dx = \int_x \int_y r(x,y) f_{Y|X}(y|x) f_X(x) dy dx$$

$$= \int_x f_X(x) \left(\int_y r(x,y) f_{Y|X}(y|x) dy \right) dx = E(E(r(X,Y)|X)).$$

$\square$.

We use that to prove the following result.

Result 4 (DECOMPOSITION OF VARIANCE)

$$V(Y) = V(E(Y|X)) + E(V(Y|X)).$$

Proof: First write

$$Y - E(Y) = Y - E(Y|X) + E(Y|X) - E(Y) = h(Y,X) + g(X),$$

where $h(Y,X) = Y - E(Y|X)$ and $g(X) = E(Y|X) - E(Y)$. Now the covariance between $g(X)$ and $h(Y,X)$ is zero, which is shown by using the law of iterated expectations and the fact that $E(h(Y,X)|X) = 0$:

$$E(h(Y,X) \cdot g(X)) = E(E(h(Y,X) \cdot g(X)|X)) = E(g(X) \cdot E(h(Y,X)|X))$$

$$= E(g(X) \cdot 0) = 0.$$

Hence

$$V(Y) = V(h(Y,X)) + v(g(X)).$$

Also,

$$V(h(Y,X) = E(h(Y,X)^2) = E((Y - E(Y|X))^2) =$$
$$E(E((Y - E(Y|X))^2)|X) = E(V(Y|X)).$$

Finally,

$$V(g(X)) = V(E(Y(X)),$$

so that

$$V(Y) = V(h(Y,X)) + V(g(X)) = E(V(Y|X)) + V(E(Y|X)).$$

□.