2: Joint Distributions

Bertille Antoine
(adapted from notes by Brian Krauth and Simon Woodcock)

In econometrics we are almost always interested in the relationship between two or more random variables. For example, we might be interested in the relationship between interest rates and unemployment. Or we might want to characterize a firm's technology as the relationship between its employment, capital stock, and output. Examples abound.

The **joint density function** of two random variables X and Y is denoted $f_{XY}(x,y)$. In the discrete case, $f_{XY}(x,y) = \Pr(X = x, Y = y)$. More generally, $f_{XY}(x,y)$ is defined so that

$$\Pr(a \leq X \leq b, c \leq Y \leq d) = \begin{cases} \sum\limits_{a\leq x\leq b} \sum\limits_{c\leq y\leq d} f_{XY}(x,y) & \text{if } X,Y \text{ are discrete,} \\ \int_a^b \int_c^d f_{XY}(x,y)\,dydx & \text{if } X,Y \text{ are continuous.} \end{cases}$$

As in the univariate case, joint densities satisfy:

$$f_{XY}(x,y) \geq 0 \tag{1}$$

$$\begin{array}{cc} \sum_x \sum_y f_{XY}(x,y) = 1 & \text{if } X,Y \text{ are discrete,} \\ \int_X \int_Y f_{XY}(x,y)\,dydx = 1 & \text{if } X,Y \text{ are continuous.} \end{array} \tag{2}$$

We define the cumulative distribution analogously. It describes the probability of a joint event:

$$\begin{aligned} F_{XY}(x,y) &= \Pr(X \leq x, Y \leq y) \\ &= \begin{cases} \sum\limits_{X\leq x} \sum\limits_{Y\leq y} f_{XY}(x,y) & \text{if } X,Y \text{ are discrete,} \\ \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s,t)\,dtds & \text{if } X,Y \text{ are continuous.} \end{cases} \end{aligned}$$

Marginal Distributions

A **marginal probability density** describes the probability distribution of one random variable. We obtain the marginal density from the joint density by summing or integrating out the other variable(s):

$$f_X(x) = \begin{cases} \sum_y f_{XY}(x,y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} f_{XY}(x,t)\,dt & \text{if } Y \text{ is continuous} \end{cases}$$

and similarly for $f_Y(y)$.

Example 1 *Define a joint pdf by*

$$f_{XY}(x,y) = \begin{cases} 6xy^2 & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Verify for yourself that (1) and (2) are satisfied. Let's compute the marginal density of X. First note that for $x \geq 1$ or $x \leq 0$, $f_{XY}(x,y) = 0$ for all values of y. Thus,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = 0 \quad \text{for } x \geq 1 \text{ or } x \leq 0.$$

For $0 < x < 1$, $f_{XY}(x,y)$ is nonzero only if $0 < y < 1$. Thus,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^1 6xy^2 dy = 2xy^3\Big|_0^1 = 2x \quad \text{for } 0 < x < 1.$$

Definition 2 *Random variables X and Y are* ***independent*** *iff*

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

An equivalent definition involves the cdfs: X and Y are independent iff $F_{XY}(x,y) = F_X(x)\, F_Y(y)$. If two random variables are independent, knowing the value of one provides no information about the value of the other. We'll see another (equivalent) formal definition of independence in a minute.

Expectations, Covariance, and Correlation

The moments of variables in a joint distribution are defined with respect to the marginal distributions. For example,

$$\begin{aligned} E[X] &= \int_X x f_X(x)\,dx \\ &= \int_X \int_Y x f_{XY}(x,y)\,dydx \\ Var[X] &= \int_X (x - E[X])^2 f_X(x)\,dx \\ &= \int_X \int_Y (x - E[X])^2 f_{XY}(x,y)\,dydx. \end{aligned}$$

Using summation instead of integration yields analogous results for the discrete case.

For any function $g(X,Y)$ of the random variables X and Y, we compute the expected value as:

$$E[g(X,Y)] = \int_X \int_Y g(x,y)\, f_{XY}(x,y)\,dydx \tag{3}$$

and similarly for the discrete case. The covariance between X and Y, usually denoted σ_{XY}, is a special case of (3).

Definition 3 *The* ***covariance*** *between two random variables X and Y is*

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]\,E[Y]. \end{aligned}$$

Definition 4 *The **correlation** between two random variables X and Y is*

$$\rho_{XY} = \frac{Cov\left[X,Y\right]}{\sigma_X \sigma_Y}$$

where σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

The covariance and correlation are almost equivalent measures of the association between two variables. Positive values imply co-movement of the variables (positive association) while negative values imply the reverse. However, the scale of the covariance is the product of scales of the two variables. This sometimes makes it difficult to determine the magnitude of association between two variables. In contrast, $-1 \leq \rho_{XY} \leq 1$ is scale-free, which frequently makes it a preferred measure of association.

Theorem 5 *If X and Y are independent, then $\sigma_{XY} = \rho_{XY} = 0$.*

Proof. When X and Y are independent, $f_{XY}(x,y) = f_X(x) f_Y(y)$. Thus if $E[X] = \mu_X$ and $E[Y] = \mu_Y$

$$\begin{aligned}
\sigma_{XY} &= \int_X \int_Y (x - \mu_X)(y - \mu_Y) f_X(x) f_Y(y)\, dy dx \\
&= \int_X (x - \mu_X) f_X(x)\, dx \int_Y (y - \mu_Y) f_Y(y)\, dy \\
&= E[X - \mu_X] E[Y - \mu_Y] \\
&= 0.
\end{aligned}$$

■

Note that the converse of Theorem 5 is not true: in general $\sigma_{XY} = 0$ does **not** imply that X and Y are independent. An important exception is when X and Y have a bivariate normal distribution (below).

Proposition 6 *Some useful results on expectations in joint distributions:*

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

$$Var[aX + bY + c] = a^2 Var[X] + b^2 Var[Y] + 2ab Cov[X,Y]$$

$$Cov[aX + bY, cX + dY] = ac Var[X] + bd Var[Y] + (ad + bc) Cov[X,Y]$$

and if X and Y are independent, we have

$$E[g_1(X) g_2(Y)] = E[g_1(X)] E[g_2(Y)].$$

Bivariate Transformations

Last day we discussed transformations of random variables in the univariate case. Things are a little more complicated in the bivariate case, but not much.

Suppose that X_1 and X_2 have joint distribution $f_X(x_1, x_2)$, and that $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ are functions of X_1 and X_2. Let $\mathcal{A} = \{(x_1, x_2) : f_X(x_1, x_2) > 0\}$ and $\mathcal{B} = \{(y_1, y_2) : y_1 = g_1(x_1, x_2), y_2 = g_2(x_1, x_2) \text{ for some } (x_1, x_2) \in \mathcal{A}\}$. Assume that g_1 and g_2 are continuous, differentiable, and define a one-to-one transformation of $\mathcal{A}$ onto $\mathcal{B}$. The latter assumption implies the inverses of g_1 and g_2 exist. (Note: we're assuming that g_1 and g_2 are "one-to-one and onto" functions; they are onto by the definition of $\mathcal{B}$. A sufficient, but stronger, condition for the inverses to exist is that g_1 and g_2 are monotone). Denote the inverse transformations by $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$. We define the **Jacobian Matrix** of the transformations as the matrix of partial derivatives:

$$\mathbf{J} = \left[\begin{array}{cc} \frac{\partial h_1(y_1,y_2)}{\partial y_1} & \frac{\partial h_1(y_1,y_2)}{\partial y_2} \\ \frac{\partial h_2(y_1,y_2)}{\partial y_1} & \frac{\partial h_2(y_1,y_2)}{\partial y_2} \end{array} \right].$$

Then the joint distribution of Y_1 and Y_2 is

$$f_Y(y_1, y_2) = f_X(h_1(y_1, y_2), h_2(y_1, y_2)) \operatorname{abs}(|\mathbf{J}|).$$

The distribution only exists if the Jacobian has a nonzero determinant, i.e., if the transformations h_1 and h_2 (and hence g_1 and g_2) are functionally independent.

Example 7 *Let X_1 and X_2 be independent standard normal random variables. Consider the transformation $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ (that is, $g_1(x_1, x_2) = x_1 + x_2$ and $g_2(x_1, x_2) = x_1 - x_2$). Because X_1 and X_2 are independent, their joint pdf is the product of two standard normal densities:*

$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{2\pi} e^{\left(-x_1^2/2\right)} e^{\left(-x_2^2/2\right)}, \quad -\infty < x_1, x_2 < \infty.$$

So the set $\mathcal{A} = R^2$. The first step is to find the set $\mathcal{B}$ where $f_Y(y_1, y_2) > 0$. That is, the set of values taken by $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ as x_1 and x_2 range over the set $\mathcal{A} = R^2$. Notice we can solve the system of equations $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ for x_1 and x_2:

$$\begin{aligned} x_1 &= h_1(y_1, y_2) = \frac{y_1 + y_2}{2} \qquad (4) \\ x_2 &= h_2(y_1, y_2) = \frac{y_1 - y_2}{2}. \end{aligned}$$

This shows two things. First, for any $(y_1, y_2) \in R^2$ there is an $(x_1, x_2) \in \mathcal{A}$ (defined by h_1 and h_2) such that $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$. Thus $\mathcal{B} = R^2$. Second, since the solution (4) is unique, the transformation g_1 and g_2 is one-to-one and onto, and we can proceed. It is easy to verify that

$$\mathbf{J} = \left[\begin{array}{cc} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{array} \right]$$

and thus $|\mathbf{J}| = -\frac{1}{2}$. *Therefore*

$$\begin{aligned} f_Y(y_1, y_2) &= f_X(h_1(y_1, y_2), h_2(y_1, y_2))\, abs(|\mathbf{J}|) \\ &= \frac{1}{2\pi} \exp\left(-\frac{((y_1 + y_2)/2)^2}{2}\right) \exp\left(-\frac{((y_1 - y_2)/2)^2}{2}\right) \frac{1}{2} \\ &= \left(\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-y_1^2/4}\right) \left(\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-y_2^2/4}\right) \qquad -\infty < y_1, y_2 < \infty \end{aligned}$$

after some rearranging. Notice that the joint pdf of Y_1 *and* Y_2 *factors into a function of* Y_1 *and a function of* Y_2*. Thus they are independent. In fact, we note that the two functions are pdfs of* $N(0, 2)$ *random variables – that is,* Y_1, Y_2 *iid* $N(0, 2)$.

Conditional Distributions

When random variables are jointly distributed, we are frequently interested in representing the probability distribution of one variable (or some of them) as a function of others. This is called a **conditional distribution**. For example if X and Y are jointly distributed, the conditional distribution of Y given X describes the probability distribution of Y as a function of X. This conditional density is

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

Of course we can define the conditional distribution of X given Y, $f_{X|Y}(x|y)$ analogously.

Theorem 8 *If* X *and* Y *are independent,* $f_{Y|X}(y|x) = f_Y(y)$ *and* $f_{X|Y}(x|y) = f_X(x)$.

Proof. This follows immediately from the definition of independence. ■

The definition of a conditional density implies the important result

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

from which we can derive **Bayes' Rule:**

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

A **conditional expectation** or **conditional mean** is just the mean of the conditional distribution:

$$E[Y|X] = \int_Y y f_{Y|X}(y|x)\, dy$$

in the continuous case. Sometimes the conditional mean function $E[Y|X]$ is called the **regression** of y on x.

The **conditional variance** is just the variance of the conditional distribution:

$$\begin{aligned} Var[Y|X] &= E\left[(Y - E[Y|X])^2 |X\right] \\ &= \int_Y (y - E[Y|X])^2 f_{Y|X}(y|x)\, dy \end{aligned}$$

if Y is continuous. A convenient simplification is $Var[Y|X] = E[Y^2|X] - E[Y|X]^2$. Sometimes the conditional variance is called the **scedastic function**, and when it doesn't depend on X we say that Y is **homoscedastic.**

Some Useful Results

Theorem 9 (Law of Iterated Expectations) *For any two random variables X and Y, $E[X] = E[E[X|Y]]$.*

Proof. By definition,

$$\begin{aligned} E[X] &= \int_Y \int_X x f_{XY}(x,y)\,dxdy \\ &= \int_Y \left[\int_X x f_{X|Y}(x|y)\,dx \right] f_Y(y)\,dy \\ &= \int_Y E[X|y]\, f_Y(y)\,dy \\ &= E[E[X|Y]]. \end{aligned}$$

■

Proposition 10 (Decomposition of Variance) *For any two random variables X and Y, $Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$.*

The Bivariate Normal Distribution

The bivariate normal is a joint distribution that appears over and over in econometrics. The density is

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \left[\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{1-\rho^2} \right] \right)$$

$$\text{where } z_x = \frac{x-\mu_x}{\sigma_x}, \quad z_y = \frac{y-\mu_y}{\sigma_y}.$$

Here, μ_x, σ_x, μ_y, and σ_y are the means and standard deviations of the marginal distributions of X and Y, respectively, and ρ is the correlation between X and Y. The covariance is $\sigma_{xy} = \rho\sigma_x\sigma_y$. The bivariate normal has lots of nice properties. In particular, if $(X,Y) \sim N_2\left[\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho\right]$ then:

1. The marginal distributions are also normal:

$$\begin{aligned} f_X(x) &= N\left[\mu_x, \sigma_x^2\right], \\ f_Y(y) &= N\left[\mu_y, \sigma_y^2\right]. \end{aligned}$$

2. The conditional distributions are normal:

$$f_{Y|X}(y|x) = N\left[\alpha + \beta x, \sigma_y^2\left(1 - \rho^2\right)\right]$$

$$\text{where } \alpha = \mu_y - \beta\mu_x \text{ and } \beta = \frac{\sigma_{xy}}{\sigma_y^2}$$

and similarly for X. Note that the conditional expectation (i.e., the regression function) is linear, and the conditional variance doesn't depend on X (it's homoscedastic). These results motivate the **classical linear regression model (CLRM)**.

3. X and Y are independent if and only if $\rho = 0$. Note that the "only if" part implies that if $\rho = 0$, then X and Y are independent. This is the important exception where the converse to Theorem 5 is true. Furthermore, we know that when X and Y are independent the joint density factors into the product of marginals. Because the marginals are normal, this means the joint density factors into the product of two univariate normals. We saw an instance of this in Example 7.

The Multivariate Normal Distribution

We can generalize the bivariate normal distribution to the case of more than two random variables. Let $(x_1, x_2, ..., x_n)' = \mathbf{x}$ be a vector of n random variables, μ their mean vector, and $\mathbf{\Sigma} = E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'\right]$ their covariance matrix. The multivariate normal density is

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{\Sigma}|^{-1/2} \exp\left\{(-1/2)(\mathbf{x} - \mu)' \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu)\right\}.$$

A special case is the multivariate standard normal, where $\mu = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I}_n$ (here $\mathbf{I}_n$ is the identity matrix of order n) and

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} \exp\left\{-\mathbf{x}'\mathbf{x}/2\right\}.$$

Those properties of the bivariate normal that we outlined above have counterparts in the multivariate case. Partition $\mathbf{x}, \mu$, and $\mathbf{\Sigma}$ conformably as follows:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}.$$

Then if $[\mathbf{x}_1, \mathbf{x}_2]$ have a joint multivariate normal distribution:

1. The marginal distributions are also normal:

$$\begin{aligned} \mathbf{x}_1 &\sim N\left[\mu_1, \mathbf{\Sigma}_{11}\right], \\ \mathbf{x}_2 &\sim N\left[\mu_2, \mathbf{\Sigma}_{22}\right]. \end{aligned}$$

2. The conditional distributions are normal:

$$\mathbf{x}_1|\mathbf{x}_2 \sim N\left[\mu_{1.2}, \mathbf{\Sigma}_{11.2}\right]$$

$$
\begin{aligned}
\text{where } \mu_{1.2} &= \mu_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \mu_2) \\
\text{and } \mathbf{\Sigma}_{11.2} &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}
\end{aligned}
$$

by the partitioned inverse and partitioned determinant formulae (see Greene, equations (A-72) and (A-74)). Again, the conditional expectation is linear and the conditional variance is constant.

3. $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent if and only if $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21} = \mathbf{0}$.

Some more useful results:

Proposition 11 *If* $\mathbf{x} \sim \mathbf{N}[\mu, \mathbf{\Sigma}]$, *then* $\mathbf{Ax} + \mathbf{b} \sim N[\mathbf{A}\mu + \mathbf{b}, \mathbf{A\Sigma A}']$.

Proposition 12 *If* $\mathbf{x}$ *has a multivariate standard normal distribution, then*

$$
\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2 \sim \chi_n^2.
$$

Furthermore, if $\mathbf{A}$ *is any idempotent matrix, then* $\mathbf{x}'\mathbf{Ax} \sim \chi^2_{rank(\mathbf{A})}$.

Proposition 13 *If* $\mathbf{x} \sim N[\mu, \mathbf{\Sigma}]$ *then* $\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \mu) \sim N[\mathbf{0}, \mathbf{I}_n]$. *Furthermore,*

$$
(\mathbf{x} - \mu)' \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \sim \chi_n^2.
$$