14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 10

Konrad Menzel

March 12, 2009

1 Functions of 2 or more Random Variables

Let's recap what we have already learned about joint distributions of 2 or more random variables, say $X_1, X_2, \ldots, X_n$

- if $X_1, \ldots, X_n$ are discrete, their joint p.d.f. is given by

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

- if $X_1, \ldots, X_n$ are continuous, their joint p.d.f. is a positive function $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ such that

$$P\Big((X_1, \ldots, X_n) \in D\Big) = \int_D \cdots \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)dx_1 \ldots dx_n$$

  for any $D \subset \mathbb{R}^n$.

- $X_1, \ldots, X_n$ are independent if

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdot \ldots P(X_n \in A_n)$$

  recall that this is equivalent to

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots f_{X_n}(x_n)$$

We are now going to look at how we can generalize from the univariate case discussed above to 2 or more dimensions.

As for the single-dimensional case we'll again distinguish three cases:

1. underlying variables $X_1, \ldots, X_n$ discrete
2. underlying variable $X_1, \ldots, X_n$ continuous
3. X continuous *and* $u(X_1, \ldots, X_n)$ is an n-dimensional one-to-one function

1.1 Discrete Case

Suppose $X_1, \ldots, X_n$ are discrete with joint p.d.f. $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, and $Y_1, \ldots, Y_m$ are given by m functions

$$\begin{aligned} Y_1 &= u_1(X_1, \ldots, X_n) \\ &\vdots \\ Y_m &= u_m(X_1, \ldots, X_n) \end{aligned}$$

If we let

$$A_y := \{(x_1, \ldots, x_n) : r_1(x_1, \ldots, x_n) = y_1, \ldots, u_m(x_1, \ldots, x_n) = y_m\}$$

then the joint p.d.f. of $Y_1, \ldots, Y_m$ is given by

$$f_{Y_1,\ldots,Y_m}(y_1, \ldots, y_m) = \sum_{(x_1,\ldots,x_n)\in A_y} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Example 1 ***(Sum of Binomial Random Variables)*** *Suppose $X \sim B(m, p)$ and $Y \sim B(n, p)$ are independent binomial random variables with p.d.f.s*

$$\begin{aligned} f_X(k) &= \binom{m}{k} p^k(1-p)^{m-k} \\ f_Y(k) &= \binom{n}{k} p^k(1-p)^{n-k} \end{aligned}$$

If we define $Z = X + Y$, what is the p.d.f. $f_Z(z)$? Since X is the number of successes in a sequence of m independent trials, and Y the number of successes in n trials, both with the same success probability, a first guess would be that then Z should just be the number of successes in $m + n$ trials with success probability p, i.e. $Z \sim B(m + n, p)$. This turns out to be true, but we'll first have to check this formally:

$$\begin{aligned} P(Z = z) &= P\left(\{X = 0, Y = z\} \text{ or } \{X = 1, Y = z - 1\} \ldots \text{or } \{X = z, Y = 0\}\right) \\ &= \sum_{k=0}^{z} P(X = k, Y = z - k) = \sum_{k=0}^{z} P(X = k)P(Y = z - k) \\ &= \sum_{k=0}^{z} \binom{m}{k} p^k(1-p)^{m-k} \binom{n}{z-k} p^{z-k}(1-p)^{n-z+k} \\ &= \sum_{k=0}^{z} \binom{m}{k}\binom{n}{z-k} p^z(1-p)^{n-z} \end{aligned}$$

Now, the term $p^z(1-p)^{n-z}$ doesn't depend on k, so we can pull it out of the sum. On the other hand, I claim that

$$\sum_{k=0}^{z} \binom{m}{k}\binom{n}{z-k} = \binom{m+n}{z}$$

We can in fact show this using counting rules: by the multiplication rule and the formula for combinations, the term $\binom{m}{k}\binom{n}{z-k}$ corresponds to the number of different sets we can draw which contain k elements from a group with m members, and $z-k$ elements from another group with n members. Summing over all values of k, we get the total number of ways in which we can draw a set of z elements from both

sets combined (i.e. a set with $m+n$ members), which, according to the formula for combinations, equals $\binom{m+n}{z}$*, which is the right-hand side of the equality we wanted to prove.*
Putting bits and pieces together,

$$P(Z=z) = \binom{m+n}{z} p^z (1-p)^{n-z}$$

so that indeed, $Z \sim B(m+n, p)$.
As a cautious note, in general the sum Z of two independent random variables X and Y from the same family of distributions - in this case the binomial - will not belong to that same family. In that respect, the binomial distribution is a very special case, and there are only very few other commonly used distributions which have that property. E.g. if $X \sim B(m, p_X)$ and $Y \sim B(n, p_Y)$ with $p_X \neq p_Y$, the derivation above is not going to work anymore.

1.2 Continuous Case

Suppose $X_1, \ldots, X_n$ are continuous with joint p.d.f. $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, and Y (let's stick to only one variable to keep notation simple) is given by a function

$$Y = u(X_1, \ldots, X_n)$$

If we let

$$B_y := \{(x_1, \ldots, x_n) : u(x_1, \ldots, x_n) \leq y\}$$

then the p.d.f. of Y is given by

$$F_Y(y) = \int_{(x_1,\ldots,x_n)\in B_y} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

2 Change of Variables Formula for One-to-One Transformations

This is again a special case, which works only for continuous variables: Let A be the support of $X_1, \ldots, X_n$, i.e.

$$P\Big((x_1, \ldots, x_n) \in A\Big) = 1$$

and B the *induced* support of $Y_1, \ldots, Y_n$, i.e.

$$(Y_1, \ldots, Y_n) \in B \Leftrightarrow (X_1, \ldots, X_n) \in A$$

Suppose $Y_1, \ldots, Y_n$ are generated from $X_1, \ldots, X_n$ from a differentiable *one-to-one* transformation

$$\begin{aligned} Y_1 &= u_1(X_1, \ldots, X_n) \\ &\vdots \\ Y_n &= u_n(X_1, \ldots, X_n) \end{aligned}$$

i.e. every value of $(x_1, \ldots, x_n) \in A$ is mapped to a *unique* element $(y_1, \ldots, y_n) \in B$. We can then define the inverse $[s_1(x_1, \ldots, x_n), \ldots, s_n(x_1, \ldots, x_n)]'$ such that

$$\begin{aligned} X_1 &= s_1(Y_1, \ldots, Y_n) \\ &\vdots \\ X_n &= s_n(Y_1, \ldots, Y_n) \end{aligned}$$

If $s_1(\cdot), \ldots, s_n(\cdot)$ are differentiable on B, we define the matrix

$$J = \begin{pmatrix} \frac{\partial}{\partial y_1} s_1 & \ldots & \frac{\partial}{\partial y_n} s_1 \\ \vdots & & \vdots \\ \frac{\partial}{\partial y_1} s_n & \ldots & \frac{\partial}{\partial y_n} s_n \end{pmatrix}$$

This matrix of partial derivatives is also called the *Jacobian* of the inverse transformation. For those of you who didn't take Linear Algebra, it's sufficient if you can work with the 2-by-2 case. You should also know that for a two-by-two matrix A, the determinant is given by

$$\det(A) = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Proposition 1 *If the mapping of $X_1, \ldots, X_n$ to $Y_1, \ldots, Y_n$ as outlined above is one-to-one and has a differentiable inverse $s_1(\cdot), \ldots, s_n(\cdot)$, then the joint p.d.f. of $Y_1, \ldots, Y_n$ is given by*

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \begin{cases} f_{X_1,\ldots,X_n}(s_1(y), \ldots, s_n(y)) \cdot |\det(J)| & \text{if } y \in B = \text{support}(Y_1, \ldots, Y_n) \\ 0 & \text{otherwise} \end{cases}$$

2.1 Linear Transformation

Let $\mathbf{X}$ be a vector of random variables, i.e. $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, and

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \mathbf{AX}$$

for an $n \times n$ matrix $\mathbf{A}$ with $\det(\mathbf{A}) \neq 0$. Then the linear mapping $\mathbf{Y} = \mathbf{AX}$ is one-to-one (since the matrix has an inverse), and we can find the joint distribution of Y using the change of variables formula

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)|\det(\mathbf{A}^{-1})| = \frac{1}{|\det(\mathbf{A})|} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

Example 2 *To see how this matters in economics, suppose we have a simple (partial equilibrium) model for the market for orange juice in Boston. Firms are willing to supply quantity q_s as a linear function of price p with coefficients α_s and β_s*

$$q_s = \alpha_s + \beta_s p + u_s$$

where u_s is a random variable (say, sunshine hours in Florida). Consumers demand quantity q_d given another random shock u_d (say income):

$$q_d = \alpha_d - \beta_d p + u_d$$

In equilibrium, supply equals demand, i.e. prices are such that $q_s = q_d = q$, and price and quantity are jointly determined by the following relationship

$$\begin{bmatrix} 1 & -\beta_s \\ 1 & \beta_d \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} \alpha_s \\ \alpha_d \end{bmatrix} + \begin{bmatrix} u_s \\ u_d \end{bmatrix}$$

We may know or postulate the joint p.d.f. $f_U(u_s, u_d)$ of the shocks (u_d, u_s), from which we can derive the joint distribution of price and quantity. This joint p.d.f. is going to depend crucially on the Jacobian (which is the matrix on the left-hand side). In this case, $\det(J) = \beta_d + \beta_s$*, so that if supply and/or demand have a nontrivial slope, the transformation from the shocks to price and quantity is one-to-one, and the resulting joint p.d.f. is*

$$f_{PQ}(p, q) = f_U(u_1(p, q), u_2(p, q))|\beta_s + \beta_d|$$

This is a little too far beyond what we are going to cover in this class, but the Jacobian term $|\beta_s + \beta_d|$ captures the interdependence of price and quantity through the market equilibrium. It turns out to be the source of what in 14.32 will be called the "simultaneity problem" which makes it very difficult to estimate supply and demand separately from market outcomes. This is one of the fundamental problems in econometrics.

2.2 Distribution of $X + Y$ (Convolution)

Suppose X and Y are independent continuous random variables with p.d.f.s $f_X(x)$ and $f_Y(y)$, respectively, so that the joint p.d.f. of the random variables is $f_{XY}(x, y) = f_X(x)f_Y(y)$. What is the p.d.f. of $Z = X + Y$?

Example 3 *Remember that we already did an example like this in class: we were looking at the sum of the lives of two spark plugs in a lawnmower, and it turned out that the probability $P(X + Y \leq z)$ was the integral of the joint density $f_{XY}(x, y)$ over the triangular area defined by $\{(x, y) : y \leq z - x\}$. So the c.d.f. of Z is given by*

$$F_Z(z) = P(X+Y \leq z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{XY}(x, y)dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_X(x)f_Y(y)dydx = \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)dx$$

From this, we can derive the density of Z,

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

The random variable $Z = X + Y$ is also called the convolution *of X and Y. Note that the last formula is valid only for the sum of* independent *random variables.*

Example 4 *The discussion in the previous example was again along the lines of the "2-step" method, and one might wonder whether it would also be possible to use the shortcut through the formula for transformations of variables.*
The mapping from (X, Y) to Z is clearly not one-to-one, so we can't use the formula for transformations of variables right away. However, we can do the following "trick": define

$$\begin{aligned} Z = u_1(X, Y) &= X + Y \\ W = u_2(X, Y) &= Y \end{aligned}$$

Then, the inverse transformation is defined as

$$\begin{aligned} X = s_1(Z, W) &= Z - W \\ Y = s_2(Z, W) &= W \end{aligned}$$

Then,

$$\det \mathbf{J} = \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$$

Therefore

$$f_{ZW}(z,w) = f_{XY}(s_1(z,w), s_2(z,w)) = f_X(s_1(z,w))f_Y(s_2(z,w)) = f_X(z-w)f_Y(w)$$

We can now obtain the marginal p.d.f. of Z by integrating the joint p.d.f. over w

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-w)f_Y(w)dw$$

which is the same formula as that obtained from the previous derivation.

Example 5 *We already saw several instances of the exponential distribution (e.g. in the lawnmower example). Let X and Y be independent exponential random variables with marginal p.d.f.s*

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

By the last formula, the p.d.f. of $Z = X + Y$ is given by

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-w)f_Y(w)dw \\ &= \int_0^z e^{-(z-w)}e^{-w}dw \\ &= \int_0^z e^{-z}dw = ze^{-z} \end{aligned}$$

where integration limits in the second step come from the fact that the support of X and Y is restricted to the positive real numbers, i.e. for $z < 0$, $f_X(z)$ is zero, whereas for $z > w$, $f_Y(z-w)$ becomes zero.