1 Distribution Theory

Last Updated: January 28, 2009

1.1 Multivariate Distributions

Let C be a sample space and X_1 and X_2 be functions such that $X_1(c) = x_1$ and $X_2(c) = x_2$. We say that the pair (X_1, X_2) is a **random vector** where the space of (X_1, X_2) is given by $D = \{(x_1, x_2) : x_1 = X_1(c), x_1 = X_1(c), c \in C\}$.

Example 1 *A coin is tossed 3 times. Define X_1 to be the number of heads on the first two tosses and X_2 to be the number of heads on all three tosses. Then the sample space will be given by*

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

and the random vector will induce the sample space

$$D = \{(0,0), (0,1), (1,1), (1,2), (2,2), (2,3)\}.$$

All of the intuition behind random variables will carry over to random vectors. For example, the CDF of (X_1, X_2) is given by

$$F_{X_1X_2}(x_1, x_2) = P\left(\{X_1 \le x_1\} \cap \{X_2 \le x_2\}\right).$$

If (X_1, X_2) is absolutely continuous then

$$F_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1X_2}(t_1, t_2) dt_1 dt_2$$

where $f_{X_1X_2}(x_1, x_2)$ is the PDF of the random vector. In addition, provided that the PDF is continuous at (x_1, x_2), we will have that

$$\frac{\partial^2 F_{X_1X_2}(x_1, x_2)}{\partial x_1 x_2} = f_{X_1X_2}(x_1, x_2).$$

In addition, we will also have that $\int\int_D f_{X_1X_2}(x_1, x_2) dx_1 dx_2 = 1$ where D is the support of the random vector and that $f_{X_1X_2}(x_1, x_2) \geq 0$.

So, how do we go from the joint distribution of (X_1, X_2) to either the distribution of X_1 or X_2? To do this, first note that

$$\begin{aligned} \{X_1 &\leq x_1\} = \{X_1 \leq x_1\} \cap \{-\infty < X_2 < \infty\} \\ &= \{X_1 \leq x_1, -\infty < X_2 < \infty\}. \end{aligned}$$

Consequently, we will have that

$$F_{X_1}(x_1) = P(X_1 \leq x_1, -\infty < X_2 < \infty) = \lim_{x_2 \to \infty} F_{X_1X_2}(x_1, x_2).$$

Note that the last equality follows because the probability of a limit is the limit of a probability.

Next, note that for a discrete random vector, we will have that

$$F_{X_1}(x_1) = \sum_{w_1 \leq x_1} \sum_{-\infty < x_2 < \infty} p_{X_1X_2}(w_1, x_2).$$

We call $p_{X_1}(w_1) \equiv \sum_{-\infty < x_2 < \infty} p_{X_1X_2}(w_1, x_2)$ the **marginal distribution** of X_1 evaluated at w_1.

Example 2 *Consider the following joint distribution for a discrete random vector*

$X_2 \backslash X_1$	1	2	3	$p(X_2)$
1	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{2}$
2	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{2}$
$p(X_1)$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{2}$	

Note that to obtain the marginal distribution of X_1, we sum vertically over the support of X_2 and to obtain the marginal distribution of X_2, we sum horizontally over the support of X_1.

The procedure to obtain marginal distributions is analogous for continuous random variables:

$$\begin{aligned} F_{X_1}(x_1) &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1X_2}(w_1, x_2) dw_1 dx_2 \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1X_2}(w_1, x_2) dx_2 dw_1. \end{aligned}$$

The marginal distribution for X_1 is given by $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1X_2}(w_1, x_2) dx_2$. As a general rule, we always integrate out the "one that we do not want" when calculating marginal distributions. So, if we have a random vector (X, Y) and if we want the marginal distribution of Y then we will integrate out over the support of X.

Example 3 *Consider*

$$f(x_1, x_2) = \begin{cases} 8x_1x_2 \text{ for } 0 < x_1 < x_2 < 1 \\ 0 \text{ otherwise} \end{cases}$$

Then, we will have that

$$\begin{aligned} E[X_1X_2^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2^2 8x_1x_2 dx_1dx_2 = 8\int_0^1\int_0^{x_2} x_1^2x_2^3 dx_1dx_2 \\ &= \frac{8}{3}\int_0^1 x_2^6 dx_2 = \frac{8}{3} * \frac{1}{7} = \frac{8}{21}. \end{aligned}$$

We can also define the expectation of a random vector $X = (X_1, X_2)$ as $E[X] = (E[X_1], E[X_2])$.

1.2 Transformations

Now let's think about how we would calculate the distribution of the transformation of a random vector. To do this, we start with an example.

Example 4 *Consider*

$$f(x, y) = \begin{cases} 1 \text{ for } 0 < x, y < 1 \\ 0 \text{ otherwise} \end{cases}$$

Define $z = x + y$. Then we will have that

$$F_z(z) = \begin{cases} 0 \text{ for } z \leq 0 \\ \int_0^z \int_0^{z-x} dydx = \frac{z^2}{2} \text{ for } 0 < z < 1 \\ 1 - \int_{z-1}^1 \int_{z-x}^1 dydx = 1 - \frac{(2-z)^2}{2} \text{ for } 1 \leq z < 2 \\ 1 \text{ for } z \geq 2 \end{cases}$$

and, so, we obtain that

$$f_z(z) = \begin{cases} z \text{ for } 0 < z < 1 \\ 2 - z \text{ for } 1 \leq z < 2 \end{cases}.$$

Now, let's generalize what we just did. Let (X_1, X_2) have a continuous PDF given by $f_{X_1X_2}(x_1, x_2)$ with support give by S. Define the one-one mapping

$$\begin{aligned} y_1 &= u_1(x_1, x_2) \\ y_2 &= u_2(x_1, x_2) \end{aligned}$$

where the mapping is from S to T. Because the mapping is one-to-one, we can write that

$$\begin{aligned} x_1 &= w_1(y_1, y_2) \\ x_2 &= w_2(y_1, y_2) \end{aligned}$$

and, so the Jacobian of this transformation is

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

If we let $A \subset S$ and B be the mapping of A, then we will have that

$$\begin{aligned} P((Y_1, Y_2) &\in B) = P((X_1, X_2) \in A) = \int\int_A f_{X_1X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int\int_B f_{X_1X_2}(w_1(y_1, y_2), w_2(y_1, y_2))|J| dy_1 dy_2. \end{aligned}$$

Thus, we will have that

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(w_1(y_1, y_2), w_2(y_1, y_2))|J| \text{ for } (y_1, y_2) \in T.$$

Example 5 *Now, let's apply this theorem to the previous example. Define*

$$\begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= X_1 - X_2 \end{aligned}$$

which implies that

$$\begin{aligned} X_1 &= \frac{1}{2}(Y_1 + Y_2) \\ X_2 &= \frac{1}{2}(Y_1 - Y_2). \end{aligned}$$

So, we will have that

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = \left| -\frac{1}{4} - \frac{1}{4} \right| = \frac{1}{2}.$$

Next, let's note the following set of inequalities

$$\begin{aligned} 0 &< x_1 < 1, 0 < x_2 < 1 \Leftrightarrow \\ 0 &< \frac{1}{2}(y_1 + y_2) < 1, 0 < \frac{1}{2}(y_1 - y_2) < 1 \Leftrightarrow \\ -y_1 &< y_2, y_2 < 2 - y_1, y_2 < y_1, y_2 > y_1 - 2. \end{aligned}$$

The last two sets of inequalities defines T or the support of (Y_1, Y_2). The distribution of the *transformation is then given by*

$$f_{Y_1Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2} \textit{ for } (y_1, y_2) \in T \\ 0 \textit{ otherwise} \end{cases}.$$

We can now calculate the marginal distribution of Y_1 as

$$f_{Y_1}(y_1) = \begin{cases} \int_{-y_1}^{y_1} \frac{1}{2} dy_2 = y_1 \textit{ for } 0 \leq y_1 < 1 \\ \int_{y_1-2}^{2-y_1} \frac{1}{2} dy_2 = 2 - y_1 \textit{ for } 1 \leq y_1 \leq 2 \end{cases}.$$

1.3 Conditional Distributions and Expectations

First, let's consider the discrete case. Let (X_1, X_2) be discrete random variables. Then we will have that

$$P(X_2 = x_2 | X_1 = x_1) = \frac{P(X = x_1, X = x_2)}{P(X = x_1)} = \frac{p_{X_1X_2}(x_1, x_2)}{p_{X_1}(x_1)}.$$

This is the definition of the **conditional distribution** of X_2 given X_1. Note conditional distributions have the same properties as univariate distributions e.g.

$$\sum_{x_2} p_{X_2|X_1}(x_2|x_1) = \sum_{x_2} \frac{p_{X_1X_2}(x_1, x_2)}{p_{X_1}(x_1)} = \frac{1}{p_{X_1}(x_1)} \sum_{x_2} p_{X_1X_2}(x_1, x_2) = 1.$$

A similar discussion holds for when (X_1, X_2) is continuous. In this scenario, we define the conditional distribution as

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} \text{ for } f_{X_1}(x_1) > 0.$$

It is also easy to show that the object integrates to unity. The main point to remember with conditional distributions is that they will behave like marginal distributions when calculating expectations and probabilities.

Example 6 *Let X_1 and X_2 have the PDF*

$$f_{12}(x_1, x_2) = \begin{array}{l} 2 \text{ for } 0 \leq x_1 \leq x_2 \leq 1 \\ 0 \text{ otherwise} \end{array} .$$

Then we will have that

$$\begin{aligned} f_1(x_1) &= \int_{x_1}^{1} 2dx_2 = 2(1 - x_1) \\ f_2(x_2) &= \int_{0}^{x_2} 2dx_1 = 2x_2. \end{aligned}$$

This then gives us that

$$f_{2|1}(x_2|x_1) = \frac{f_{12}(x_1, x_2)}{f_1(x_1)} = \frac{1}{(1 - x_1)} \text{ for } x_1 \leq x_2 \leq 1$$

which is a uniform distribution on $[x_1, 1]$. It is trivial to verify that this distribution integrates to unity. We conclude this example with the following calculation

$$E\left[X_2|X_1 = \frac{1}{2}\right] = \int_{\frac{1}{2}}^{1} \frac{x_2}{1 - \frac{1}{2}} dx_2 = 2\int_{\frac{1}{2}}^{1} x_2 dx_2 = x_2^2|_{\frac{1}{2}}^{1} = 1 - \frac{1}{4} = \frac{3}{4}.$$

Example 7 *Let X be uniformly distributed on $[0, 1]$. First, draw $X = x$. Next, draw Y from $[x, 1]$. Let's calculate the marginal distribution of Y. We start by noting that the conditional* *distribibition of* Y *given* $X = x$ *is given by*

$$f_{Y|X}(y|x) = \frac{1}{1-x} \text{ for } x \le y \le 1.$$

We can recover the joint distribution of (X, Y) *by calculating*

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x) = \frac{1}{1-x} \text{ for } 0 \le x \le y \le 1.$$

Let's verify that this is, indeed, a PDF:

$$\int_0^1 \int_x^1 f_{X,Y}(x, y) dy dx = \int_0^1 \int_x^1 \frac{1}{1-x} dy dx = \int_0^1 dx = 1.$$

The marginal distribution of Y *can be recovered now via*

$$f_Y(y) = \int_0^y \frac{1}{1-x} dx = -\ln(1-x)|_0^y = -\ln(1-y) \text{ for } 0 \le y \le 1.$$

An important result is the Law of Iterated Expectations.

Theorem 8 *(Law of Iterated Expectations)* $E\left[E\left[X_2|X_1\right]\right] = E[X_2]$

Proof.

$$\begin{aligned} E\left[X_2\right] &= \int\int x_2 f(x_1, x_2) dx_1 dx_2 = \int \underbrace{\int x_2 f_{2|1}(x_2|x_1) dx_2}\, f_1(x_1) dx_1 \\ &= \int E\left[X_2|X_1 = x_1\right] f_1(x_1) dx_1 = E\left[E\left[X_2|X_1\right]\right] \end{aligned}$$

■

The Law of Iterated Expectations is not just important for econometrics and probability theory, it is also important for economic theory. To see this, let's work through an example based on Deaton (1992). Our starting point is the Permanent Income Hypothesis which states that agents will tend to save out of transitory income gains, but consume out of permanent gains. One way of understanding this is through the equation

$$s_t = -\sum_{k=1}^{\infty} (1+r)^{-k} E(\Delta y_{t+k} | I_t)$$

where s_t is savings at time t, r is the interest rate, y_t is income at time t and I_t is the agent's information set at time t. Essentally, what this equation states is that agents will tend to dissave when they expect their incomes to rise in the future. Of course, a crucial issue is to understand how both agents and the econometrician calculate expectations of future income. Typically, it is assumed the two have the same information set. It is then assumed that income follows a univariate stochastic process and then expectations are calculated accordingly.

Now, suppose that the econometrician does not know what the agent does. In particular, the econometrician's informations set is given by $H_t \subseteq I_t$. If we assume that savings is observed by the econometrician, we can take expectations of the equation above conditional on H_t and apply the Law of Iterated Expectations to obtain

$$s_t = -\sum_{k=1}^{\infty} (1+r)^{-k} E(\Delta y_{t+k} | H_t)$$

A problem now arises in that we are (possibly) using savings to forcast future income, whereas before we were assuming that income followed a univariate stochastic process which was both known to the agent and the econometrician. However, if income follows a univariate process that is unkown to the econometrician, then savings may be informative of future income growth. As pointed out by Deaton (1992), the realization that the econometrician knows less than the agent raises the possibility that the econometrics may need to be done differently.

1.4 Correlation

Suppose that we want to summarize the relationship between X and Y. First, we consider the following object:

$$E\left[\left(X-\mu_1\right)\left(Y-\mu_2\right)\right]=E\left[XY-X\mu_2-Y\mu_1+\mu_1\mu_2\right]=E\left[XY\right]-\mu_1\mu_2\equiv\sigma_{XY}$$

where $\mu_1 \equiv E[X]$ and $\mu_2 = E[Y]$. We call σ_{XY} the **covariance** between X and Y. Next, we define

$$\rho\equiv\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

to be the correlation between X and Y. Note the denomintor of the above object contains the standard deviations of X and Y, not their variances. One of the properties of the correlation is that its abolsute value is always less than or equal to unity. If $\rho = 1$ ($\rho = -1$), then we say that the random variables have a perfect positive (negative) correlation. If $\rho = 0$, then the random variables are uncorrelated. To see why we must have that $|\rho| \leq 1$, consider

$$\begin{aligned} h(\upsilon) &= E\left[\left((X-\mu_1)+\upsilon\,(Y-\mu_2)\right)^2\right] \\ &= E\left[(X-\mu_1)^2+2\upsilon\,(X-\mu_1)\,(Y-\mu_2)+\upsilon^2\,(Y-\mu_2)^2\right] \\ &= \sigma_X^2+2\upsilon\sigma_{XY}+\upsilon^2\sigma_Y^2. \end{aligned}$$

Note that for all υ, we must have that $h(\upsilon) \geq 0$. Accordingly, the above quadratic expression can have at most one real root which means that its discriminant cannot be postive and, thus, we will have that

$$(2\sigma_{XY})^2 - 4\sigma_X^2\sigma_Y^2 \leq 0 \Leftrightarrow \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1 \Leftrightarrow |\rho| \leq 1.$$

Next, suppose that the expectation of Y given $X = x$ is a linear function of X so that

$$E\left[Y|X=x\right] = a + bx.$$

Then it can easily be shown that

$$a = \mu_Y - b\mu_X \text{ and } b = \rho\frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}.$$

Students may recognize that the expression for b looks a lot like the formula for the OLS estimate of the slope parameter in a linear regression. In fact, b is the population analogue of the OLS estimator.

Example 9 *Let r_S denote the returns of a stock and let r_M denote the return on a broad market* *index. If we write*

$$E\left[r_S | r_M\right] = \alpha + \beta r_M$$

then we will have that

$$\beta = \rho \frac{\sigma_S}{\sigma_M}$$

where ρ is the correlation between the two returns and σ_S and σ_M are the variances of the returns of the stock and the market index. Because $|\rho| \leq 1$, we will have that

$$\sigma_S \geq |\beta| \, \sigma_M.$$

1.5 Independent Random Variables

We say that X_1 and X_2 are **independent random variables** if and only if $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ where $f_1(x_1)$ and $f_2(x_2)$ are the marginal distributions of X_1 and X_2. One of the implications of this definition is that the conditional distribution of a random variable will be equal to its marginal distribution which is completely analogous to what we saw with probabilities.

Example 10 *Consider the joint distribution:*

$$f(x_1, x_2) = \begin{array}{l} x_1 + x_2 \text{ for } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 \text{ otherwise} \end{array}.$$

The marginal distribution of X_1 is given by

$$f_1(x_1) = \int_0^1 (x_1 + x_2)dx_2 = x_1 + \frac{1}{2} \text{ for } 0 < x_1 < 1.$$

The marginal distribution of X_2 is the same. Since

$$f(x_1, x_2) \neq f_1(x_1)f_2(x_2)$$

we can conclude that these random variables are not independent.

As it turns out, if we can write $f(x_1, x_2) = g(x_1)h(x_2)$ for two functions $g(.)$ and $h(.)$ then X_1 and X_2 are independent. Clearly, the converse of this statement is also true. In addition, we will also have that two random variables are independent if and only if $F(x_1, x_2) = F_1(x_1)F_2(x_2)$. The proof is straightforward, so we omit it. Finally, another useful result which is also trivial to prove is that, for two independent random variables X_1 and X_2, we will have that

$$E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)].$$

Note that independent random variables always have zero correlation, but the reverse need not be true which we illustrate in the following example.

Example 11 *Let (X, Y) take on values in $\{(0,0), (1,1), (1,0), (1,-1)\}$ with equal probability. Clearly, we will have that*

$$\sigma_{XY} = E[XY] - E[X]E[Y] = 0 - \frac{3}{4} * 0 = 0$$

and, thus, these random variables are uncorrelated. However, they are not independent as

$$P(X = 0) = \frac{1}{4} \neq P(X = 0|Y = 1) = 0.$$