Econometrics: Lecture 3

Hang Miao

Rutgers University

January 21, 2021

Overview

Joint Probability Distribution

Motivation

In economics, we are often interested in two (or more) random variables at the same time. For example, wage and the education level, crime rate and the number of police, the level of air pollution and rate of respiratory illness in cities, or the number of Facebook friends and the age of Facebook members. Joint distributions allow us to reason about the relationship between multiple events.

Ways to Define a Joint Probability Distribution

The joint probability distribution can be expressed either in terms of a **joint cumulative distribution function (joint CDF)** or in terms of a **joint probability density function (joint PDF)** (in the case of continuous variables) or **joint probability mass function (joint PMF)** (in the case of discrete variables).

Joint Cumulative Distribution Function

Definition (Bivariate Joint Cumulative Distribution Function)

For a pair of random variables X, Y, the joint cumulative distribution function (joint CDF) F_{XY} is given by

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

Definition (Multivariate Joint Cumulative Distribution Function)

For N random variables $X_1, \ldots, X_N$, the joint cumulative distribution function (joint CDF) $F_{X_1,\ldots,X_N}$ is given by

$$F_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = \Pr(X_1 \leq x_1, \ldots, X_N \leq x_n)$$

Discrete: Bivariate Joint Probability Mass Function

Definition (Probability Mass Function (PMF))

If X and Y are discrete random variables, then the joint probability distribution of could be defined by assign the probability to each pair of values (x, y) within the support of $X$and$Y$. The function assigning the value is called joint probability mass function (joint PMF):

$$p_{X,Y}(x, y) = \Pr(X = x \text{ and } Y = y)$$

where

$$\text{where} \sum_i \sum_j \Pr(X = x_i \text{ and } Y = y_j) = 1$$

Definition (Cumulative Distribution Function (CDF))

$$F_{X,Y}(x, y) = \Pr(X \leq x \text{ and } Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} p_{X,Y}(s, t)$$

Discrete: Multivariate Joint Probability Mass Function

Definition (Probability Mass Function (PMF))

The generalization of the preceding two-variable case is the joint probability distribution of n discrete random variables $X_1, X_2, \ldots, X_n$ which is:

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \Pr(X_1 = x_1 \text{ and } \ldots \text{ and } X_n = x_n)$$

$$\text{where} \sum_i \sum_j \cdots \sum_k \Pr(X_1 = x_{1i}, X_2 = x_{2j}, \ldots, X_n = x_{nk}) = 1.$$

Definition (Cumulative Distribution Function (CDF))

$$\begin{aligned} F_{X_1,\ldots,X_N}(x_1, \ldots, x_N) &= \Pr(X_1 \leq x_1, \ldots, X_N \leq x_n) \\ &= \sum_{s \leq x_1} \sum_{t \leq x_2} \cdots \sum_{u \leq x_n} p_{X_1,\ldots,X_n}(s, t, \ldots, u)) \end{aligned}$$

Continuous: Bivariate Joint Probability Density Function

Probability Distribution Can not defined by PMF

Same as the univariate case, if both random variable X, Y takes a continuum of outcome, the probability for any pair of (x, y) is 0. Thus we can not define the joint probability distribution function by assigning the probability to each pair of (x, y) as what we did in discrete case.

Definition (Probability Density Function (PDF))

While if we have the joint cumulative distribution function for the joint probability distribution, the joint probability density function $f_{X,Y}(x, y)$ for two continuous random variables could be defined as the derivative of the joint cumulative distribution function:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

$$\text{where} \int_x \int_y f_{X,Y}(x,y)\, dy\, dx = 1$$

Continuous: Multivariate Joint Probability Density Function

Definition (Probability Density Function (PDF))

The generalization of the preceding two-variable case is the joint probability distribution of n discrete random variables $X_1, X_2, \ldots, X_n$ which is:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}$$

$$\text{where} \int_{x_1} \cdots \int_{x_n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_1 \ldots\ dx_n = 1$$

Definition (Cumulative Distribution Function (CDF))

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \Pr(X_1 \leq x_1,\ldots,X_n \leq x_n) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(s,\ldots,t)\, dx_s \ldots\ dx_t \end{aligned}$$

Marginal Distribution

Derivatives of Joint Probability Distribution

When the joint probability distribution is defined, for each random variable of this joint distribution, there is another probability distribution which could be naturally derived from the original joint distribution and is called **marginal distribution**.

Definition (Marginal Distribution: Discrete Bivariate)

The marginal distribution giving the probabilities for any one of the variables with no reference to any specific ranges of values for the other variables. For discrete bivariate random variable X, Y, the marginal probability mass function is defined as follow:

$$\Pr(X = x) = \sum_y \Pr(X = x, Y = y) = \sum_y \Pr(X = x \mid Y = y) \Pr(Y = y)$$

$$\Pr(Y = y) = \sum_x \Pr(X = x, Y = y) = \sum_x \Pr(Y = y \mid X = x) \Pr(X = x)$$

Marginal Distribution

Derivatives of Joint Probability Distribution

When the joint probability distribution is defined, for each random variable of this joint distribution, there is another probability distribution which could be naturally derived from the original joint distribution and is called **marginal distribution**.

Definition (Marginal Distribution: Discrete Bivariate)

The marginal distribution giving the probabilities for any one of the variables with no reference to any specific ranges of values for the other variables. For discrete bivariate random variable X, Y, the marginal probability mass function is defined as follow:

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_{X|Y}(x \mid y) p_Y(y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_x p_{Y|X}(x \mid y) p_X(x)$$

Marginal Distribution

Definition (Marginal Distribution: Continuous Bivariate)

For continuous random variables, the marginal probability density function for random variable X can be written as $f_X(x)$ and Y can be written as $f_Y(y)$. $f_X(x)$ and $f_Y(y)$ is defined as follow:

$$f_X(x) = \int_y f_{X,Y}(x, y)\,\mathrm{d}y = \int_y f_{X|Y}(x \mid y)\,f_Y(y)\,\mathrm{d}y,$$

$$f_Y(y) = \int_x f_{X,Y}(x, y)\,\mathrm{d}x = \int_x f_{Y|X}(y \mid x)\,f_X(x)\,\mathrm{d}x,$$

Definition (Marginal Distribution: Multivariate)

For multivariate distributions, formulae similar to those above apply with the symbols X and/or Y being interpreted as vectors. In particular, each summation or integration would be over all variables except those contained in X.

Conditional Distribution

Derivatives of Joint Probability Distribution

After we define the joint probability distribution and the derived marginal probability distribution, the **conditional probability distribution function** could be defined based on the previous two distribution.

Difference between Marginal and Conditional Distribution

Contrasted with marginal distribution, conditional distribution gives the probabilities of various values of the variables in the subset contingent upon the values of the other variables while marginal distribution gives the probabilities of various values of the variables in the subset without reference to the values of the other variables.

Definition (Conditional Probability)

For any two events A and B, the conditional probability of A given B is defined as follow:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional Distribution: Discrete

Definition (Conditional Distribution: Discrete Bivariate)

If X and Y are discrete random variables, the conditional probability mass function of X given $Y = y$ is defined as follow:

$$
\begin{aligned}
p_{X|Y}(x \mid y) &= \Pr(X = x \mid Y = y) \\
&= \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(Y = y)} \\
&= \frac{p_{X,Y}(x, y)}{p_Y(y)}
\end{aligned}
$$

where $p_{X,Y}(x, y)$ is the joint probability mass function for random X and Y while $p_Y(y)$ is the marginal distribution function for random variable Y.

Joint Distribution Representation

$$p_{X,Y}(x, y) = p_{X|Y}(x \mid y)\ p_Y(y) = p_{Y|X}(y \mid x)\ p_X(x)$$

Conditional Distribution: Continuous

Definition (Conditional Distribution: Continuous Bivariate)

For continuous random variables X, Y, the conditional probability density function of X given $Y = y$ is defined as follow::

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

where $f_{X,Y}(x, y)$ is the joint probability density function for random X and Y while $f_Y(y)$ is the marginal distribution function for random variable Y.

Joint Distribution Representation

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

Conditional Cumulative Distribution

Definition (Conditional Cumulative Distribution Function)

The conditional PMF could in turn be used to defined conditional CDF if X, Y are discrete random variable:

$$F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y = y) = \sum_{s \leq x} p_{X|Y}(s \mid y)$$

Definition (Conditional Cumulative Distribution Function)

The conditional PDF could in turn be used to defined conditional CDF if X, Y are continuous random variable:

$$F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y = y) = \int_{-\infty}^{x} f_{X|Y}(s \mid y) \, \mathrm{d}s$$

Conditional Expectation

Definition (Conditional Distribution: Discrete Bivariate)

Suppose X and Y are discrete random variables, the conditional probability mass function of X given $Y = y$ is defined as $p_{X|Y}(x \mid y)$. It is therefore natural to define the conditional expectation of X given $Y = y$ as follow:

$$\mathrm{E}[\, X \mid Y = y] = \sum_x x \Pr(X = x \mid Y = y) = \sum_x \; x \; p_{X|Y}(x \mid y)$$

Definition (Conditional Expectation: Continuous Bivariate)

Suppose X and Y are continuous random variables, the conditional probability density function of X given $Y = y$ is defined as $f_{X|Y}(x \mid y)$. It is therefore natural to define the conditional expectation of X given $Y = y$ as follow:

$$\mathrm{E}[\, X \mid Y = y] = \int_x x \; f_{X|Y}(x \mid y) \; \mathrm{d}x$$

Independence

Definition (Independence)

The random variable X and Y are said to be independent if **for any sets of real numbers** A **and** B:

$$\Pr(X \in A \text{ and } Y \in B) = \Pr(X \in A)\ \Pr(Y \in B)$$

Definition (Independence)

It can be shown that the above independence definition is equivalent to the following by using CDF:

$$F_{X,Y}(a, b) = F_X(a)\ F_Y(b) \quad \forall a, \text{b}$$

Independence

Discrete Bivariate Distribution

If X and Y are discrete random variable, the condition for independence is equivalent to

$$p_{X,Y}(a,b) = p_X(a)\ p_Y(b) \quad \forall a, \text{ b}$$

Continuous Bivariate Distribution

If X and Y are continuous random variable, the condition for independence is equivalent to

$$f_{X,Y}(x,y) = f_X(x)\ f_Y(y) \quad \forall x, \text{ y}$$

Covariance

Definition (Covariance)

For two random variable X and Y, the covariance give us the information about the relationship between these two variable. It is denoted as $\mathrm{Cov}(X, Y)$ defined as:

$$\mathrm{Cov}(X, Y) = \mathrm{E}\Big[\big(X - \mathrm{E}[X]\big)\big(Y - \mathrm{E}[Y]\big)\Big]$$

Another Expression for Covariance

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

Covariance and Correlation

Properties of Covariance

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$
- $\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)$

Definition (Correlation)

Correlation coefficient is a measure of the degree of linearity between X and Y. It is denoted by $\rho(X, Y)$ defined as:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

It can be showed that $-1 \leq \rho(X, Y) \leq 1$