University of Nottingham
UK | CHINA | MALAYSIA

Introductory Econometrics BUSI2053

Probability for Econometrics

Lecture Outline

- **Maths Primer**
- **Random Variables**
- **Probability Distributions**
- **Expected Value, Variance, Covariance and their rules**
- **Continuous RV and Normal Distribution**

- **Suggested Reading:**
  - Probability Primer: Hill, R.C., Griffiths W.E. and Lim, G.C. Principles of Econometrics, fourth edition, Wiley, 2012 (pp. 17-34)
  - Appendix A: Gujarati, D.N. and Porter D.C. Basic econometrics, 5th ed., McGraw-Hill, 2009 (pp. 801-836)
  - Review: Dougherty, Christopher. Introduction to econometrics. 4th ed. Oxford University Press, 2011 (pp.5-72)
  - Chapter 2 & Appendix 1.1: Westhoff, F. An introduction to econometrics: a self-contained approach, MIT Press, 2013

Maths Primer (Summation $\sum$)

- In Probability and Statistics, one of the math notations that will be used a great deal is the summation. The summation symbol, denoted sigma (Σ), was introduced by the Swiss polymath Leonhard Euler in the 18th century.

- The sum of the first n natural numbers can be written as: $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n$, where the numbers 1 and n are the **lower limit** and **upper limit** of summation.

- If we want to add up n observations of x together: $\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$

X = variable; x = values of the variable

- If x is a constant: $\sum_{i=1}^{n} 2 = 2 + 2 + 2 + \cdots + 2 = 2n$

- Multiply x by a constant a: $a\sum_{i=1}^{n} x_i = ax_1 + ax_2 + ax_3 + \cdots + ax_n$

- A place you will see this a lot is with the mean. The mean essentially sums up all of the values and multiplies it by $1/n$: $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1}{n} + \frac{x_2}{n} + \frac{x_3}{n} + \cdots + \frac{x_n}{n} = \bar{x}$

- E.g., the mean of the three values 1, 2, 3 : $\frac{1}{3} + \frac{2}{3} + \frac{3}{3} = \frac{6}{3} = 2$

Random Variables

- A **random variable** is a variable whose value is unknown until it is observed. i.e., not perfectly predictable (e.g., the value of bitcoin tomorrow, IE exam grade of a current Year 2 student)
- Each random variable has a set of possible values it can take, and each value has associated probability
  - A **discrete random variable** can take only a limited, or **countable**, number of values
  - An **indicator variable** taking the values one if yes, or zero if no to represent qualitative characteristics such as gender (male or female), or race (white or nonwhite)
  - A random variable that can have any value is treated as a **continuous random variable**
- A sequence of random numbers (x_1, x_2, x_3, …) must have 2 properties: mutually independent of each other and identically distributed (drawn from the same probability distribution), **iid**

slido

Please download and install the Slido app on all computers you use

Sallie is now 25 years old, single, outspoken, and very bright. When she was a student, she participated in anti-nuclear demonstrations. What do you think Sallie is doing after graduation? Choose the option with the highest probability.

ⓘ Start presenting to display the poll results on this slide.

Probability Distributions

- Probability: The Likelihood of a Particular Outcome of a Random Eexperiment (Process)
- Suppose we have data for attendance rates (×100%) and exam marks of a group of 10 students
- If we were to select one student from the table at random, that would constitute a random experiment

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

- $P(Y=65)=$
- $P(X=0.80)=$
- $P(Y \geq 75)=$
- $P(Y=65 \;\&\; X=0.80)=$
- $P(Y=65|X=0.80)=$
- $P(Y=65 \text{ or } X=0.80) =$

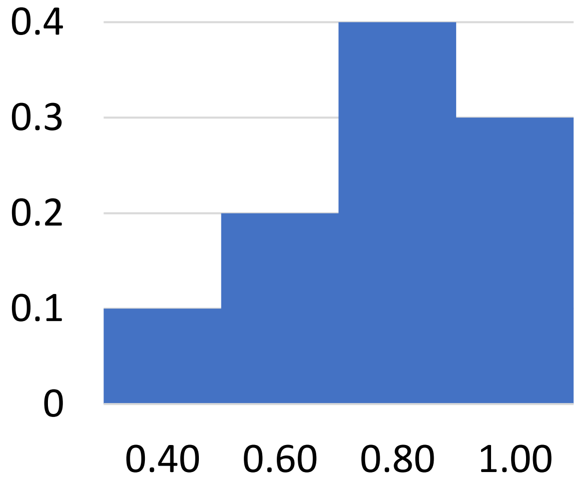
$=$

Probability Distributions

- The probabilities of possible outcomes can be summarised using a probability density function (*pdf*)
- The *pdf* for a discrete random variable indicates the probability of each possible value occurring
- The value of the probability density function $f(x)$ is the probability that the random variable X takes the value x, $f(x) = P(X = x)$ X = variable; x = values of the variable
- It must be true that $0 \leq f(x) \leq 1$
- $f(x_1) + f(x_2) + \ldots + f(x_n) = 1$

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

X	$f(x)$
0.40	
0.60	
0.80	
1.00	



Homework: Construct *pdf* of Y and scratch a density plot.

Expected Value of a Random variable

- Two key features of a probability distribution:
  - center (location) measured by the **mean**
  - width(dispersion) measured by *variance* (or **standard deviation**)
- The mean of a random variable is given by its **mathematical expectation** or **expected value**
  - If X is a discrete random variable,

$$E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_n P(X = x_n)$$

$$E(X) = \mu = \sum_{i=1}^{n} x_i \, f(x_i)$$

X	$f(x)$
0.40	0.1
0.60	0.2
0.80	0.4
1.00	0.3

Based on the probability density function we created: how can we predict the expected attendance rate of a randomly selected student?

Expected Value of a Random variable

- For the population in the table, the expected value of X is:

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

X	$P(X)=f(x_i)$	$xf(x_i)$
0.40	0.1	
0.60	0.2	
0.80	0.4	
1.00	0.3	
$\sum x_i f(x_i)$		

$E(X) = [0.40 \times P(X = 0.40)] + [0.60 \times P(X = 0.60)] + [0.80 \times P(X = 0.80)] + [1.00 \times P(X = 1.00)]$

$E(X) = (0.40 \times 0.1) + (0.60 \times 0.2) + (0.80 \times 0.4) + (1.00 \times 0.3)$

$= 0.04 + 0.12 + 0.32 + 0.30 = 0.78$

Expected Value of a Random variable (Homework)

- Find the expected value of *Y*.

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

Y	$P(Y)=f(y_i)$	$yf(y_i)$
35	**0.1**	**3.5**
45	**0.1**	**4.5**
55	**0.2**	**11**
65	**0.3**	**19.5**
75	**0.2**	**15**
85	**0.1**	**8.5**
$\sum y_i f(y_i)$		**62**

$$E(Y) = 62$$

Conditional Expectation

- For a discrete random variable, the calculation of conditional expected value uses the conditional probability density function $f(y|x)$:

$$E(Y|X=x) = \sum_{y} y f(y|x)$$

- For our data, the values of X~40% - 100%. To find expected exam mark of students with 80% attendance:

$$\mathrm{E}(Y|X=0.80) = \sum_{Y=55}^{75} y f(y|0.80)$$

$$= [55 \times f(55|0.80)] + [65 \times f(65|0.80)] + [75 \times f(75|0.80)]$$
$$= (55 \times 0.5) + (65 \times 0.25) + (75 \times 0.25) = 27.5 + 16.25 + 18.75$$
$$= 62.5$$

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

Conditional Expectation (Homework)

- **Calculate $E(Y|X{=}1.00)$**

$$\mathrm{E}(Y|X = 1.00) = \sum_{Y=65}^{85} yf(y|1.00)$$

$$= [65 \times f(65|1)] + [85 \times f(85|1)] + [75 \times f(75|1)]$$
$$= (65 \times 0.33) + (85 \times 0.33) + (75 \times 0.33) = 75$$

Attendance (X)	Exam mark (Y)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

Answer: $E(Y|X{=}1.00){=}75$

Rules for Expected Values

- If $g(X)$ is a function of the random variable X, then $g(X)$ is also random
  - If X is a discrete random variable, then the expected value of $g(X)$ is obtained using:

$$E[g(X)] = g(x_1)f(x_1) + \ldots + g(x_n)f(x_n) = \sum_{i=1}^{n} g(x_i)f(x_i)$$

  - For example, if $g(X) = X^2$:

$$E(X^2) = x_1^2 f(x_1) + \ldots + x_n^2 f(x_n) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

Quiz: Is $E(X^2) = [E(X)]^2$?

Rules for Expected Values

- Let X and Z be random variables:

$$E(X+Z) = E(X) + E(Z)$$

“the expected value of a sum is the sum of the expected values”

- If a is a constant: $E(a) = a$

$$E(aX) = aE(X)$$

- Then: $E(aX + bZ + c) = ?$

- If X and Z are independent**:** $E(XZ) = E(X)\,E(Z)$

Fixed & Random components of a random variable

Population mean of Y: $E(Y) = \mu_Y$

Hence y_i can be decomposed into fixed and random components:

$$y_i = \mu_Y + e_i$$

In observation i, the random component is given by

$$e_i = y_i - \mu_Y$$

Note that the expected value of e_i is zero:

$$E(e_i) = E(y_i - \mu_Y) = E(y_i) + E(-\mu_Y) = \mu_Y - \mu_Y = 0$$

Exam mark (Y)	***e***	
65	65 – 62	3
35	35 – 62	-27
85	85 – 62	23
65	65 – 62	3
65	65 – 62	3
75	75 – 62	13
75	75 – 62	13
55	55 – 62	-7
55	55 – 62	-7
45	45 – 62	-17
$E(e_i)$		0

$$E(Y) = \mu_Y = 62$$

Variance of a Random Variable

- The variance of a discrete or continuous random variable X :

$$E[(X-\mu)^2]$$

- Algebraically:

$$\begin{aligned}\mathrm{var}(X) = \sigma_X^2 &= E(X-\mu)^2 \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

- Note:

$$E(X) = \mu, \therefore 2\mu E(X) = 2\mu^2$$

Variance of a Random Variable

- For our example, $E(X) = \mu_x = 78\% = 0.78$

$$E(X^2) = x_1^2 f(x_1) + \ldots + x_4^2 f(x_n)$$
$$= \sum_{i=1}^{4} x_i^2 f(x_i) = ?$$

X	$P(X)=f(x_i)$	X^2	$x^2 f(x_i)$
0.40	0.1	0.16	
0.60	0.2	0.36	
0.80	0.4	0.64	
1.00	0.3	1	
$E(X^2)$			0.644

- Then, $\mathrm{var}(X) = \sigma_X^2 = E(X^2) - \mu^2 = 0.644 - (0.78)^2$
$$= 0.644 - 0.6084 = 0.0356$$

- The standard deviation: ?

Variance of a Random Variable (Homework)

- Using $E(Y)$=62 (from exercise in slide 9)

- Find $\mathrm{var}(Y) = \sigma_Y^2 = E(Y^2) - {\mu_Y}^2$ and standard deviation of Y.

$$\mathrm{var}(Y) = \sigma_Y^2 = E(Y^2) - {\mu_Y}^2 = 4045 - (62)^2$$
$$= 4045 - 3844 = 201$$

Y	$P(Y)=f(y_i)$	Y^2	$y^2 f(y_i)$
35	**0.1**	**1225**	**122.5**
45	**0.1**	**2025**	**202.5**
55	**0.2**	**3025**	**605,0**
65	**0.3**	**4225**	**1267.5**
75	**0.2**	**5625**	**1125**
85	**0.1**	**7225**	**722.5**
			4045

Answers: var(Y)=201, stdv(Y)= 14.18

Rules for Variance

- Let a be a constant, then:

$$\text{var}(aX) = a^2 \text{var}(X)$$

- To prove this;

$$\begin{aligned} \text{var}(aX) &= E[(aX - a\mu_X)]^2 \\ &= E[a^2(X - \mu_X)^2] \\ &= a^2E[(X - \mu_X)]^2 = a^2 \text{var}(X) \end{aligned}$$

- But $\text{var}(a) = 0$; why?

Rules for Variance

$$\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)$$

$$\operatorname{var}(X-Y) = \operatorname{var}(X) + \operatorname{var}(Y) - 2\operatorname{cov}(X,Y)$$

- If X and Y are independent, cov(X, Y) = 0, then:

$$\operatorname{var}(X \pm Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$

- But; $\operatorname{var}(a+X) = \operatorname{var}(a) + \operatorname{var}(X) + 2\operatorname{cov}(a,X) = \operatorname{var}(X)$

- If a and b are constants, then:

$$\begin{aligned}&\operatorname{var}(2aX+bY)\\ &= (2a)^2\operatorname{var}(X) + b^2\operatorname{var}(Y) + 2\times 2a\times b\operatorname{cov}(X,Y)\\ &= 4a^2\operatorname{var}(X) + b^2\operatorname{var}(Y) + 4ab\operatorname{cov}(X,Y)\end{aligned}$$

Covariance Between Two Random Variables

- The covariance between X and Y is a measure of linear association between them (See POE Appendix B1.5)

$$\text{cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$$

Note:

$$\text{cov}(X, X) = E[(X - \mu_X)(X - \mu_X)] = E(X^2) - \mu_X^2$$

$$\text{cov}(X, X) = \text{var}(X)$$

$$\begin{aligned} & E(X^2 - 2\mu_X X + \mu_X^2) \\ &= E(X^2) - 2\mu_X E(X) + \mu_X^2 \\ &= E(X^2) - 2\mu_X\mu_X + \mu_X^2 \\ &= E(X^2) - 2\mu_X^2 + \mu_X^2 \\ &= E(X^2) - \mu_X^2 \end{aligned}$$

Rules for Covariance

- If $Z = V + W$,

  $\text{Cov}(X, Z) = = \text{Cov}(X, [V+W]) = \text{Cov}(X, V) + \text{Cov}(X, W)$

- If $Z = bW$, where b is a constant,

  $\text{Cov}(X, Z) = \text{Cov}(X, bW) = b\text{Cov}(X, W)$

  Example: $\text{Cov}(X, 3W) = 3\text{Cov}(X, W)$

- If b is a constant,

  $\text{Cov}(X, b) = 0$

  Example: $\text{Cov}(X, 10) = 0$

Exercise (Homework)

- Q is normally distributed with mean 2 and variance 6. Let P=3 + 5Q.
  - Find E(5P) (Ans: 65)
  - Find the variance and standard deviation of P. (Ans: Var = 150)
  - Find the covariance between P and Q. Interpret your result. (Ans: 30)

$$E(5P) = E[5 \times (3 + 5Q)] = E(15 + 25Q)$$
$$= 15 + [25 \times E(Q)] = 15 + (25 \times 2)$$
$$= 65$$

Correlation coefficient

- Interpreting the actual value of Cov(X, Y) or σ_{XY} is difficult because X and Y may have different units of measurement
  - Scaling the covariance by the standard deviations of the variables eliminates the units of measurement, and defines the **correlation** between X and Y:

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- the correlation ρ measures the degree of linear association between two random variables ($-1 \leq \rho \leq +1$)

Correlation coefficient

- For our example:

$$E(XY) = \sum_{x=0.4}^{1} \sum_{y=35}^{85} xy\, f(x,y)$$

$$= (1 \times 65 \times 0.1) + (0.4 \times 35 \times 0.1) + (1 \times 85 \times 0.1)$$
$$+(0.8 \times 65 \times 0.1) + (0.6 \times 65 \times 0.1) + (1 \times 75 \times 0.1)$$
$$+(0.8 \times 75 \times 0.1) + (0.8 \times 55 \times 0.2) + (0.6 \times 45 \times 0.1)$$
$$= 6.5 + 1.4 + 8.5 + 5.2 + 3.9 + 7.5 + 6 + 8.8 + 2.7 = 50.5$$

Attendance (*X*)	Exam mark (*Y*)
1.00	65
0.40	35
1.00	85
0.80	65
0.60	65
1.00	75
0.80	75
0.80	55
0.80	55
0.60	45

$$Cov(X,Y) = \sigma_{XY} = E(XY) - \mu_X\mu_Y = 50.5 - (0.78 \times 62) = 2.14$$

$$\rho = \frac{Cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{2.14}{\sqrt{0.0356}\sqrt{201}} = 0.8$$

Probability Distributions of a Binary (Indicator) variable

- An **indicator (binary) variable** taking the values one if yes, or zero if no to represent qualitative characteristics such as gender (male or female)
- Since there is two possible outcome, if the probability of the variable taking the value of 1 is p, the probability of the variable taking the value of zero is $(1 – p)$.

Gender (X)	$f(X)$	
1	0.6	
0	?	
$E(X) = \sum x_i f(x_i)$		

- Homework:
- Find the expected value and variance of X.

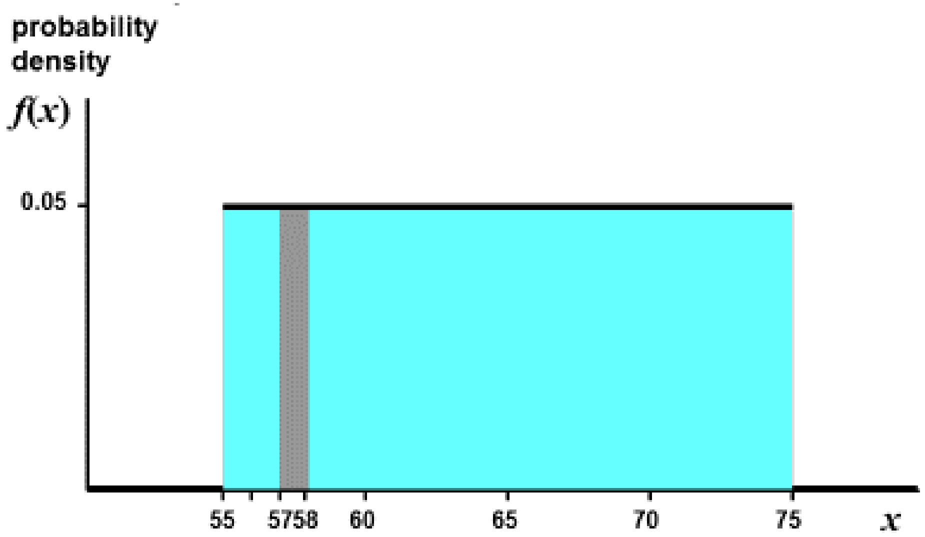
(X^2)	$f(X)$	
1	0.6	0.6
0	0.4	0
$E(X^2) = \sum x_i^2 f(x_i)$		0.6

Answer: $E(X)$=0.6; Var(X)=0.24

$$Var(X) = E(X^2) - \mu_X^2 = 0.6 - 0.36 = 0.24$$

Probability Distributions of a Continuous Variable

- Most economic variables are continuous.
- Example, take the potential costs of an investment. Assume that it can be anywhere from 55 to 75 million dollars with equal probability within the range.
- In the case of a continuous random variable, the probability of it being equal to a given finite value (for example, cost equal to exactly 57 million) is always infinitesimal (almost zero). Therefore, we have to describe probability of a range of the values of a continuous variable.



$f(x) = 0.05$ for $(57 < x < 58)$

$f(x) = 0$ for $(x = 57)$

$f(x) = 0.05$ for $(57 \leq x \leq 58)$

$f(x) = 0.00$ for $(x < 55$ *or* $x > 75)$

$f(x) =?$ for $(60 < x < 65)$

Probability Distributions of a Continuous variable

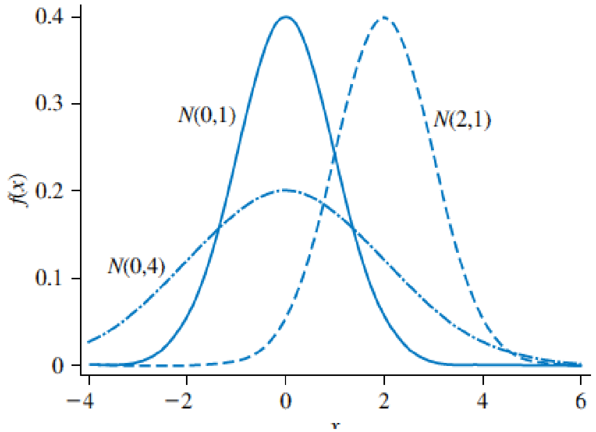
Normal Distribution

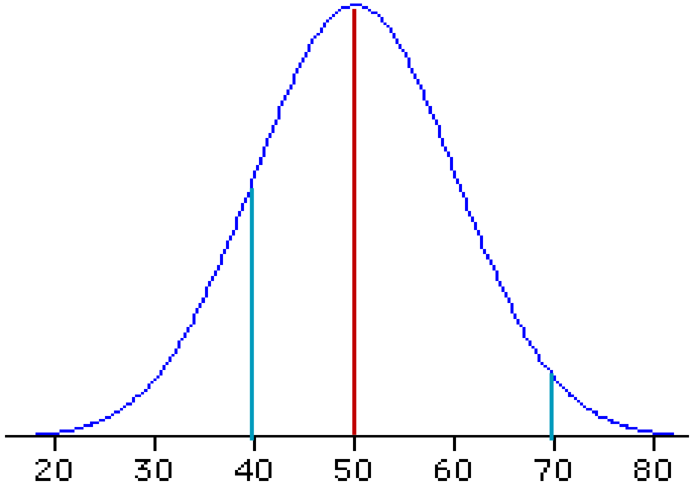
- If X is a normally distributed random variable with mean μ and variance σ^2 , it can be symbolized as $X \sim N(\mu, \sigma^2)$
- The *pdf* of X is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty$$

- The cumulative probability can be calculated for any x between -∞ and +∞
- For a continuous random variable, areas under a curve represent probabilities that X falls in an interval, such as:

$$P(40 \leq X \leq 70) = F(70) - F(40)$$
$$= 0.9772 - 0.1587 = 0.8185$$





$X \sim N(50,10)$

Lecture Summary

- After this lecture, you should be able to:
  - understand Random Variables & Probability Distributions of discrete and continuous random variable
  - calculate expected value, variance, covariance and correlation coefficient of random variables
  - apply rules and properties of expected value, variance and covariance
  - understand the concept of Normal Distribution

University of Nottingham
UK | CHINA | MALAYSIA

Thank you and any question?