Introductory Econometrics

Probability and Statistics Refresher

Monash Econometrics and Business Statistics

Semester 2, 2018

Outline

- Reference: Appendices B-1, B-2, B-3, B-4 of the textbook
  - Random variables (discrete, continuous) and their probability distribution
  - Mean, variance, standard deviation
  - Properties of expectation
  - Covariance and correlation
  - Joint and conditional distributions
  - Conditional expectation function as the fundamental target of modelling

Random variables

- Economic and financial variables are by nature random. We do not know what their values will be until we observe them.
- A random variable is a rule that assigns a numerical outcome to an event in each possible state of the world.
- For example, the first wage offer that a BCom graduate receives in the job market is a random variable. The value of ASX200 index tomorrow is a random variable. Other examples are ...
- A discrete random variable has a finite number of distinct outcomes. For example, rolling a die is a random variable with 6 distinct outcomes.
- A continuous random variable can take a continuum of values within some interval. For example, rainfall in Melbourne in May can be any number in the range from 0.00 to 200.00 mm.
- While the outcomes are uncertain, they are not haphazard. The rule assigns each outcome to an event according to a probability.

A random variable and its probability distribution

- A discrete random variable is fully described by

its possible values	$x_1, x_2, \ldots, x_m$
probability corresponding to each value	$p_1, p_2, \ldots, p_m$

  with the interpretation that $P(X = x_1) = p_1$, $P(X = x_2) = p_2$, ..., $P(X = x_m) = p_m$.

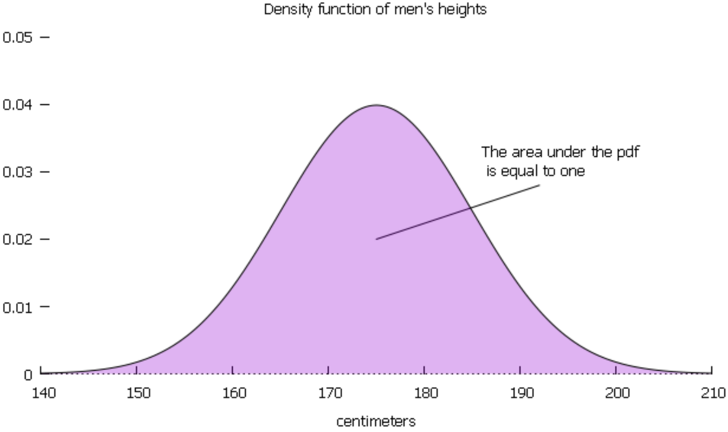
- The probability density function (pdf) for a discrete random variable X is a function f with $f(x_i) = p_i$, $i = 1, 2, \ldots, m$ and $f(x) = 0$ for all other x.

- Probabilities of all possible outcomes of a random variable must sum to 1.

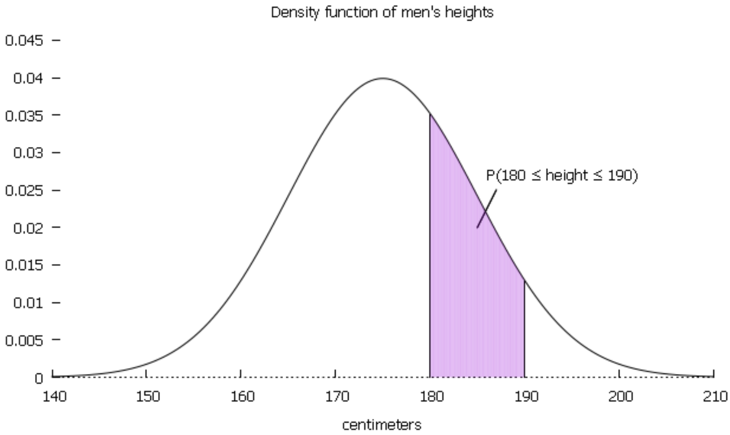
$$\sum_{i=1}^{m} p_i = p_1 + p_2 + \cdots + p_m = 1$$

- Examples:
    1. Rolling a die
    2. Sex of a baby who is not yet born. Is it a random variable?
    3. The starting wage offer to a BCom graduate
- The probability density function (pdf) for a continuous random variable X is a function f such that $P(a \leq X \leq b)$ is the area under the pdf between a and b.
- The total area under the pdf is equal to 1.

- Example: Distribution of men's height
- The area under the pdf is equal to 1



- The probability of that the height of a randomly selected man lies in a certain interval is the area under the pdf over that interval



Features of probability distributions: 1. Measures of Central Tendency

Textbook reference B-3

- Expected value or mean of a discrete random variable is given by

$$E(X) = p_1x_1 + p_2x_2 + \cdots + p_mx_m = \sum_{i=1}^{m} p_ix_i$$

  and for a continuous random variable is given by

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- Intuitively, expected value is the long-run average if we observe X many, many, many times.
- It is convention to use the Greek letter μ to denote expected value:

$$\mu_X = E(X)$$

- Another measure of central tendency is the median of X, which is the "middle-most" outcome of X, i.e. x_{med} such that $P(X \leq x_{med}) = 0.5$. Median is preferred to the mean when the distribution is heavily skewed, e.g. income or house prices.
- Finally, there is the mode which is the most likely value, i.e. the outcome with the highest probability. It is not a widely used measure of central tendency.

2. Measures of dispersion

Textbook reference B-3

- Variance of a random variable:

$$\sigma_X^2 = Var(X) = E(X - \mu_X)^2$$

- Variance is a measure of spread of the distribution of X around its mean.
- If X is an action with different possible outcomes, then $Var(X)$ gives an indication of riskiness of that action.
- Standard deviation is the square root of the variance. In finance, standard deviation is called the volatility in X.

$$\sigma_X = sd(X) = \sqrt{E(X - \mu_X)^2}$$

- The advantage of standard deviation over variance is that it has the same units as X.

Properties of the Expected Value

Textbook reference B-3

1. For any constant c, $E(c) = c$.
2. For any constants a and b,

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

3. Expected value is a linear operator, meaning that expected value of sum of several variables is the sum of their expected values:

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

- The above three properties imply that for any constants a, b, c and d and random variables X, Y and Z,

$$E(a + bX + cY + dZ) = a + bE(X) + cE(Y) + dE(Z)$$

Question - http://mars.mu - Feed Code: Z3DKLD

We have put one quarter of our savings in a term deposit (a risk free investment) with annual return of 2% and invested the other three quarters in an investment fund with expected annual return of 3% and variance of 4.
The expected value of the annual return of our portfolio is

A. $\frac{2+3}{2} = 1.5\%$

B. $\frac{1}{4} \times 2 + \frac{3}{4} \times 3 = 2.75\%$

C. $\frac{3}{4} \times 3 = 2.25\%$

D. $\frac{1}{4} \times 2 + \frac{3}{4} \times 3 \times \sqrt{4} = 5\%$

E. $\frac{1}{4} \times 0 + \frac{3}{4} \times 4 = 3\%$

- It is important to have in mind that E is a linear operator, so it "goes through" sums of random variables, but it does not go through non-linear transformations of random variables. For example:

$$E(X^2) \neq (E(X))^2$$

$$E(\log X) \neq \log(E(X))$$

$$E(XY) \neq E(X)E(Y) \text{ unless } X \text{ and } Y \text{ are statistically independent}$$

- Using properties of expectations, we can now show that $Var(X) = E(X^2) - \mu^2$

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

Properties of the Variance

Textbook reference B-3

1. For any constant c, $Var(c) = 0$.
2. For any constants a and b,

$$Var(aX + b) = a^2 Var(X)$$

3. There is a third property related to the variance of linear combinations of random variables that is very important and we will see later after we introduce the covariance.

Question - http://mars.mu - Feed Code: Z3DKLD

We have put one quarter of our savings in a term deposit (a risk free investment) with annual return of 2% and invested the other three quarters in an investment fund with expected annual return of 3% and variance of 4.
The variance of the annual return of our portfolio is

A. $\left(\frac{3}{4}\right)^2 \times 4 = 2.25$

B. $\frac{1}{4} \times 0 + \frac{3}{4} \times 4 = 3$

C. $\frac{1}{4} \times 2 + \frac{3}{4} \times 4 = 3.25$

D. $\left(\frac{1}{4}\right)^2 \times 2 + \left(\frac{3}{4}\right)^2 \times 3 = 1.8125$

E. $\left(\frac{1}{4}\right)^2 \times 2 + \left(\frac{3}{4}\right)^2 \times 4 = 2.375$

Properties of the Conditional Expectation

- Conditional expectation of Y given X is generally a function of X.
- Property 1: Conditional on X, any function of X is no longer a random variable and can be treated as a known constant, and then the usual properties of expectations apply. For example, if X, Y and Z are random variables and a, b and c are constants, then

  $$E(XY \mid X) = XE(Y \mid X)$$

  or

  $$E((a + bX + cXY + X^2Z) \mid X) = a + bX + cXE(Y \mid X) + X^2E(Z \mid X)$$

- Property 2: If $E(Y \mid X) = c$ where c is a constant that does not depend on X, then $E(Y) = c$. This is intuitive: if no matter what X happens to be, we always expect Y to be c, then the expected value of Y must be c regardless of X, i.e. the unconditional expectation of Y must be c.

Important features of joint probability distribution of two random variables: Measures of Association

Textbook reference B-4

- Statistical dependence tells us that knowing the outcome of one variable is informative about probability distribution of another.
- To analyse the nature of dependence, we can look at the joint probability distribution of random variables
- This is too complicated when random variables have many possible outcomes (e.g. per capita income and life span, or returns on Telstra and BHP stocks)
- We simplify the question to: when X is above its mean, is Y more likely to be below or above its mean?
- This corresponds to the popular notion of X and Y being "positively or negatively correlated"

Covariance

- Question: "when X is above its mean, is Y more likely to be below or above its mean?"
- We can answer this by looking at the sign of the covariance between X and Y defined as:

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X \mu_Y$$

- If X and Y are independent $Cov(X, Y) = 0$.
- For any constants a_1, b_1, a_2 and b_2

$$Cov(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 Cov(X, Y)$$

Correlation

- Only the sign of covariance is informative. Its magnitude changes when we scale variables.

$$Cov(aX, bY) = a\ b\ Cov(X, Y)$$

- A better and unit free measure of association is correlation which is defined as:

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}$$

- Correlation is always between -1 and $+1$, and its magnitude, as well as its sign, is meaningful.
- Correlation does not change if we change the units of measurement

$$Corr(a_1X + b_1, a_2Y + b_2) = Corr(X, Y)$$

Variance of sums of random variables: Diversification

Textbook reference B4

- One of the important principles of risk management is "Don't put all your eggs in one basket."
- The scientific basis of this is that:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

- Example: You have the choice of buying shares of company A with mean return of 10 percent and standard deviation of 5 percent, or shares of company B with mean return of 10 percent and standard deviation of 10 percent. Which would you prefer?
- Obviously A is less risky, and you prefer A to B.

Variance of sums of random variables: Diversification

- Now consider a portfolio of investing 0.8 of your capital in company A and the rest in B, where, as before, A has mean return of 10 percent and standard deviation of 5 percent, and B has mean return of 10 percent and standard deviation of 10 percent. What are the return and the risk of this position with the added assumption that the returns to A and B are independent.

- Denoting the portfolio return by Z, we have

$$
\begin{aligned}
Z &= 0.8A + 0.2B \\
E(Z) &= E(0.8A + 0.2B) = \\
Var(Z) &= Var(0.8A + 0.2B) =
\end{aligned}
$$

- We can see that this portfolio has the same expected return as A, and is safer than A.

Diversification in econometrics - Averaging

- Suppose we are interested in starting salaries of BCom graduates. This is a random variable with many possibilities and a probability distribution.
- Let's denote this random variable by Y. We also denote its population mean and variance by μ and σ^2. We are interested in estimating μ, which is the expected wage of a BCom graduate.
- Suppose we choose one BCom graduate at random and denote his/her starting salary by Y_1. Certainly Y_1 is also a random variable with the same possible outcomes and probabilities as Y. Therefore $E(Y_1) = \mu$. So it is OK to take the value of Y_1 as an estimate of μ, and the variance of this estimator is σ^2.
- But if we take 2 independent observations and use their average as our estimator of μ, we have:

$$
\begin{aligned}
E(\frac{1}{2}(Y_1 + Y_2)) &= \frac{1}{2}(\mu + \mu) = \mu \\
Var(\frac{1}{2}(Y_1 + Y_2)) &= \frac{1}{4}Var(Y_1) + \frac{1}{4}Var(Y_2) \\
&= \frac{1}{4}(\sigma^2 + \sigma^2) = \sigma^2/2
\end{aligned}
$$

Diversification in econometrics - Averaging

- Now consider a sample of n independent observations of starting salaries of BCom graduate $\{Y_1, Y_2, \ldots, Y_n\}$
- Y_1 to Y_n are $i.i.d.$ (independent and identically distributed) with mean μ and variance σ^2.
- Their average is a portfolio of that gives each of these n random variables the same weight of $\frac{1}{n}$. So

$$E\left(\bar{Y}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

$$\begin{aligned} Var\left(\bar{Y}\right) &= Var\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} Y_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

- The sample average has the same expected value as Y but a lot less risk. In this way, we use the scientific concept of diversification in econometrics to find better estimators!

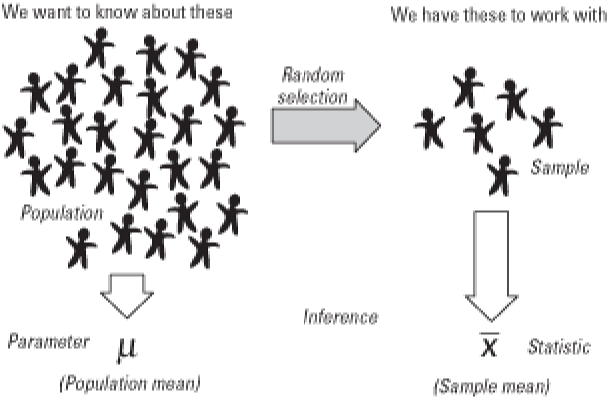
Key concepts and their importance

- Business and economic variables can be thought of as random variables whose outcomes are determined by their probability distribution.
  - A measure of central tendency of the distribution of a random variable is its expected value.
  - Important measures of dispersion of a random variable are variance and standard deviation. These are used as measures of risk.
- Covariance and correlation are measures of linear statistical dependence between two random variables.
  - Correlation is unit free and measures the strength and direction of association.
  - Statistical dependence or association does not imply causality.
  - Two random variables that have non-zero covariance or correlation are statistically dependent, meaning that knowing the outcome of one of the two random variables gives us useful information about the other.
  - Averaging is a form of diversification and it reduces risk.

Let's Play

http://www.onlinestatbook.com/stat_sim/sampling_dist/

Modelling mean



$$E(y \mid x) = \beta_0 + \beta_1 x$$
(Conditional expectation function)

$$\hat{y} = \widehat{E(y \mid x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$
(Sample regression function)

Laws of Probability

- To understand what a conditional expectation is, we start with laws of probability

1. Probability of any event is a number between 0 and 1. The probabilities of all possible outcomes of a random variable add up to 1
2. If A and B are mutually exclusive events then

$$P(A \text{ or } B) = P(A) + P(B)$$

   Example: You role a fair die. What is the probability of the die showing a number less than or equal to 2?

3. If A and B are two events, then

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

   Example: You role a fair die. What is the probability that it shows 1 given that it is showing a number less than or equal to 2?

Joint probability density function

- The ultimate goal of this unit is to study the relationship between one variable y with many variables x_1 to x_k. But let's start with only one x, and with the discrete case.
- Suppose y is the number of bathrooms and x is the number of bedrooms in an apartment in Melbourne. Assume y has two possible values, 1 and 2, and x has three possible values 1, 2 and 3. The joint pdf is gives us the probabilities of every possible outcome of (x, y).

$y \downarrow, x \rightarrow$	1	2	3	marginal f_y
1	0.40	0.24	0.04	
2	0.00	0.16	0.16	
marginal f_x				

- The entries show the probabilities of different possible combination of bedrooms and bathrooms (x, y). For example the top left cell shows $P(x = 1 \& y = 1) = 0.40$.
- Check the first law of probability: all probabilities are between 0 and 1 and the probabilities of all possible outcomes sum to 1
- Using the second law of probability, we can use the joint pdf to deduce the pdf of x by itself (called the marginal density of x), and also the marginal density of y

Conditional density

- Using the third law of probability, we can also deduce the conditional distribution of number of bathrooms in an apartment given that it has 1 bedroom.

$$P(y = 1 \mid x = 1) = \frac{P(y = 1 \;\&\; x = 1)}{P(x = 1)} = \frac{0.40}{0.40} = 1.00$$

$$P(y = 2 \mid x = 1) =$$

- Similarly, we can deduce the conditional distribution of y given $x = 2$

$$P(y = 1 \mid x = 2) = \frac{P(y = 1 \;\&\; x = 2)}{P(x = 2)} =$$

$$P(y = 2 \mid x = 2) =$$

- And y given $x = 3$

$$P(y = 1 \mid x = 3) =$$

$$P(y = 2 \mid x = 3) =$$

Conditional expectation function

- Each of these conditional densities has an expected value

$$\begin{array}{cc} y \mid x=1 & f_{y|x=1} \\ 1 & 1.00 \\ 2 & 0.00 \end{array} \quad \Rightarrow \quad E(y \mid x=1) = 1 \times 1.00 + 2 \times 0.00 = 1.00$$

$$\begin{array}{cc} y \mid x=2 & f_{y|x=2} \\ 1 & 0.60 \\ 2 & 0.40 \end{array} \quad \Rightarrow \quad E(y \mid x=2) = 1 \times 0.60 + 2 \times 0.40 = 1.40$$

$$\begin{array}{cc} y \mid x=3 & f_{y|x=3} \\ 1 & 0.20 \\ 2 & 0.80 \end{array} \quad \Rightarrow \quad E(y \mid x=3) = 1 \times 0.20 + 2 \times 0.80 = 1.80$$

- Plot the expected values for different values of x. Do they fit on a straight line? What is the equation of that line?

- When y and x have many possible outcomes or when they are continuous random variables (like birth weight and number of cigarettes during pregnancy, or price of a house and its land size) we cannot enumerate the joint density and perform the same exercise
- Therefore we go after the conditional expectation function directly

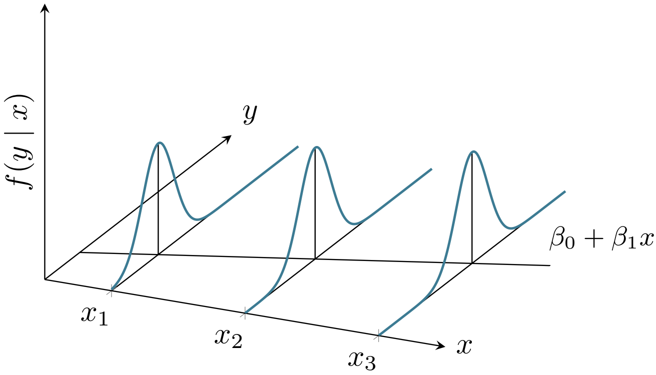
$$E(y \mid x) = \beta_0 + \beta_1 x \tag{PRF}$$

- For example, if y is the price of a house and x is its area, the mean of the price of houses with area x is given by $\beta_0 + \beta_1 x$
- The price of each house can be written as random variations around this central value

$$y = \beta_0 + \beta_1 x + u$$

where u is a random variable with $E(u \mid x) = 0$, which implies that $E(u) = 0$ also

The simple linear regression model in a picture



The simple linear regression model in equation form

- The following equation specifies the conditional mean of y given x

$$y = \beta_0 + \beta_1 x + u \text{ with } E(u \mid x) = 0$$

- It is an incomplete model because it does not specify the probability distribution of y conditional of x
- If we add the assumptions that $Var(u \mid x) = \sigma^2$ and that the conditional distribution of u given x is normal, then we have a complete model, which is called the "classical linear model" and was shown in the picture on the previous slide.
- In summary: we make the assumption that in the big scheme of things, data are generated by this model, and we want to use observed data to learn the unknowns β_0, β_1 and σ^2 in order to predict y using x.

Summary

- We reviewed some fundamentals of probability:
  - Random variables (discrete, continuous) and their probability distribution
  - Mean, variance, standard deviation
  - Expectation, covariance and correlation
  - Joint and conditional probability distributions
- We established that the problem of finding good estimators and the problem of forming optimal portfolios are based on the same principles.
- We concluded by establishing that the conditional expectation function of y given x as the vehicle for predicting y using x.
- Next week we learn how to estimate the parameters of the conditional expectation function.