ECON 3150/4150, Spring term 2014. Lecture 1

Ragnar Nymoen

University of Oslo

13 January 2014

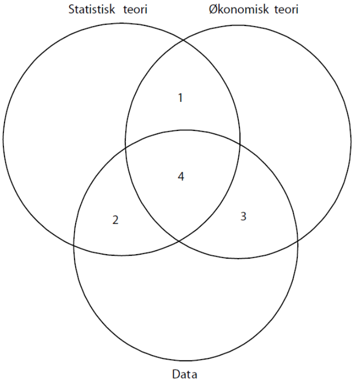
References to Lecture 1 and 2

- Stock and Watson (SW)
  - Ch 1-2;
- Bårdsen and Nymoen (BN)
  - Kap 4-4.5

The goal of econometrics

- The Econometric Project: Use real world data and statistical theory to obtain empirical knowledge about relationships that hold outside the given sample.
- Statistical inference is a main concept: Generalization of empirical evidence from a concrete and limited data set "to the population".
- Inference can be about a **parameter** in a economic relationship (the marginal propensity to consume in the consumption function) or about the (treatment) effect or a policy reform—a question about a **causal effect** as noted on page 48 in S&W.
- *Econometric models* are the hallmark of econometrics

Econometrics is a "combined discipline" I



- Ects. combine knowledge and skill from three main areas
- Several of the intersections are of interest, but
- Area 4 represents *Econometric Models*

Historical reference

A defining contribution to modern econometrics is *The Probability Approach to Econometrics* by Trygve Haavelmo from 1944.

PREFACE

This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon modern theory of probability and statistical inference. A few words may be said to justify such a study.

The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been

Historical reference

A defining contribution to modern econometrics is *The Probability Approach to Econometrics* by Trygve Haavelmo from 1944.

PREFACE

This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon modern theory of probability and statistical inference. A few words may be said to justify such a study.

The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been

Data types I

- Cross section: A data set where the variables vary across n individuals $i = 1, 2, \ldots, n$
- Time series: A data set where the variables vary over T time periods: $t = 1, \ldots, T$
- Panel data: Variation both across individuals and over time.

There are other distinctions between data types as well'

- Micro/macro
- Experimental/non-experimental

Data types II

- In this course we will concentrate on the "common ground" between models for cross-section and for time series data
- Will use notation like (Y_i, X_i) $i = 1, 2, \ldots, n$ for the most
- But will use (Y_t, X_t) $t = 1, 2, \ldots, T$ when it is relevant to be precise about time-series data for example
- For panel data we need both subscripts as in (Y_{it}, X_{it}) $i = 1, 2, \ldots, n, t = 1, 2, \ldots, T$

Random events and random variables I

- A random variable attaches a value to each single event of an experiment.
- The "experiment" can be literal (like in coin tossing).
- In econometrics we regard the real world data as if they were generated by a large experiment, that can be analysed with the language of mathematical statistics.
- An event represented by a random variable can be a simple result (boy/girl child) of an experiment, or a composite result (number of girls from 3 births).
- A random variable is therefore a function of the simple events of an experiment.

Random events and random variables II

Definition

A random variable is a function with numerical values defined over a value set ("utfallsrom")

- For the variable "Number of girls" the value set is $\{0, 1, 2, 3\}$.
- A *discrete random variable* can take a finite number of values.
- A *continuous random variable* can take an infinite number of values.
- The terms *random variable* and *stochastic variable* are synonyms

Cumulative distribution functions I

Let X denote a random variable (discrete or continuous) and x a value of that variable.

Definition (Cumulative distribution)

The cumulative distribution function (cdf) $F_X(x)$ gives the probability Pr that a random variable X is less than or equal the outcome x:

$$F_X(x) = \Pr(X \leq x).$$

- For a discrete variable, the *cdf* has a characteristic step-function shape, starting in 0 and increasing in steps until it reaches the value 1 for the highest value in the value set.

Cumulative distribution functions II

- The change in the discrete *cdf* takes place at the point where the discrete variable goes from a lower to a higher level. The change in the *cdf* corresponds to the probability distribution which gives the probability p_x for a value x:

$$p_x = f_X(x)$$

  with properties

  - $0 \leq p_x \leq 1$
  - $\sum_x p_x = 1$

- For a continuous variable, the *cdf* is continuous from the right.

Cumulative distribution functions III

- The probability of a given value is zero. To represent the change in a continuous *cdf* the *probability density function* (*pdf*) is used:

$$f_X(x) \geq 0 \quad \forall \text{ ("for all") } x,$$

The *pdf* is scaled in such a way that the area under the function is 1:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1.$$

The probability for $P(X \leq a)$ is then

$$\Pr(X \leq a) = F_X(a) = \int_{-\infty}^{a} f_X(x)\, dx.$$

Cumulative distribution functions IV

Conversely:

$$f_X(x) = \frac{d}{dx} P(X \leq x)$$

confirming that the *pdf* $f_X(x)$ represents the change in the *cdf*.

Expectation

If X a discrete random variable, the expectation is

$$\mu_X = E\left(X\right) = \sum_{i=1}^{k} x_i f_X\left(x_i\right)$$

If X is a continuous random variable:

$$\mu_X = E\left(X\right) = \int_{-\infty}^{\infty} x f_X\left(x\right) dx$$

Rules for the expectation

1. $E\left(a\right) = a$, for a constant a
2. $E\left(bX\right) = bE\left(X\right) = b\mu_X$, for a constant b
3. $E\left(a + bX\right) = E\left(a\right) + E\left(bX\right) = a + b\mu_X$
4. $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E\left(X_i\right) = \sum_{i=1}^{n} \mu_{X_i}$ for n random variables

See *Key Concept 2.3* in SW

Variance and standard deviation

$$\sigma_X^2 = var(X) = E\left[(X-\mu_X)^2\right]$$
$$\sigma_X = sd(X) = \sqrt{var(X)}$$

Note $var(X)$ can be written as

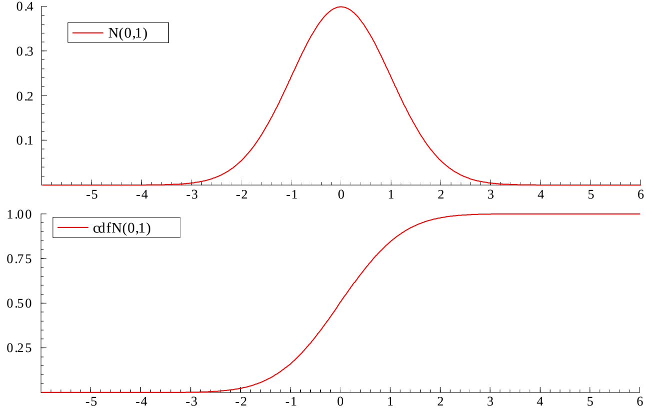
$$var(X) = E\left[(X-\mu_X)^2\right] = E\left(X^2 - 2X\mu_X + \mu_X^2\right) = E(X^2) - \mu_X^2$$

This seem to imply that $var(X)$ can be negative! Is that correct?
Rules for the variance:

1. $var(a) = 0$
2. $var(bX) = b^2 Var(X) = b^2\sigma_X^2$
3. $var(a+bX) = b^2 Var(X) = b^2\sigma_X^2$

- These results (and generalizations, see below, and *Key Concept 2.3* in SW) are particularly useful because we often

Standard normal distribution



- Compare figure 2.5 and figure 2.6 in SW

Skewness and kurtosis

- The degree of non-symmetry in a distribution, skewness, is measured by the third moment:

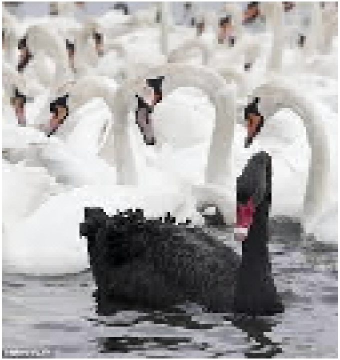
$$Skewness = \frac{E\left[(X-\mu_X)^3\right]}{\sigma_X^3}$$

- Kurtosis (the fourth moment)

$$Kurtosis = \frac{E\left[(X-\mu_X)^4\right]}{\sigma_X^4}$$

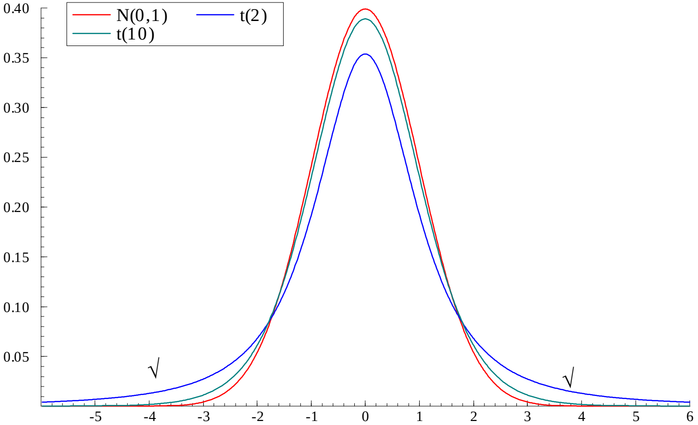
- The *Normal distribution* is often used as a reference. It has $Skewness = 0$ and $Kurtosis = 3$.
- $Kurtosis > 3$ implies*fat-tails*, or *heavy-tails*
- *Investment strategies* or *prediction models* that assume a normal distribution when the distribution is in fact, heavy-tailed, can lead to financial losses and forecast failures.

A black swan I



- A black swan may be rare
- but less rare than once believed
- Excess kurtosis

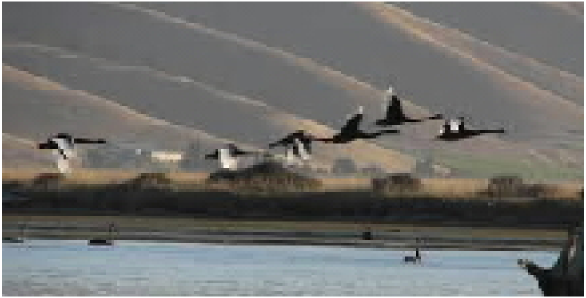
Normal and fatter tails I



Normal and fatter tails II

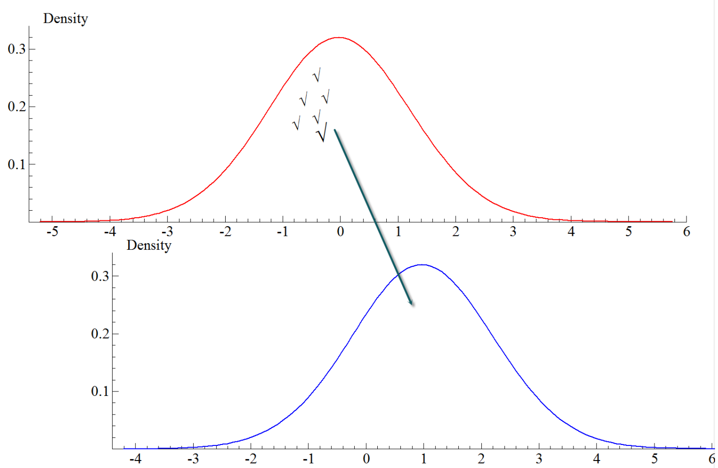
- Normal pdf cf Figure 2.5
- Together with pdf for t(2) and t(10) with two black swans indicated

A flock of black swans



- A flock of black swans
- Is more like a *location-shift* the distribution—a shift in the expectation
- Can be a more useful metaphor in economics

A flock of black swans I



The distribution of a function of a variable I

We are often interested in the distribution of a function $g(X)$ of the random variable X.

Table: Discrete probability distribution for X

Values for $X = x_i$:	0	1	2	3
Probability $f_X(x_i)$:	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

If $g(X) = X^2$ what is the distribution of $Y = (X-2)^2$?

The distribution of a function of a variable II

- First: Find the values set of Y: $\{4, 1, 0, 1\}$
- The value 1 occurs twice (for $x_i = 1$ and $x_i = 3$)

Values for $X = x_i$:	0	1	2	3
Probability $f_X(x_i)$:	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{1}{8}$
Values for $Y = (x_i - 2)^2$:	0	1	4	
Probability $f_Y(y_i)$	$\frac{2}{8}$	$\frac{3+1}{8}$	$\frac{2}{8}$	

The distribution of a function of a variable III

- When $Y = g(X)$ is continuous, and the inverse function $X = g^{-1}(Y)$ exists, we have the important result:

$$f_Y(y) = f_X(x)\left|\frac{dg^{-1}(y)}{dy}\right| = f_{X(y)}[X(y)]\left|\frac{dg^{-1}(y)}{dy}\right| \quad (1)$$

- The reason we use the absolute value $\left|dg^{-1}(y)/dy\right|$ is that $dg^{-1}(y)$ can be a declining function, while a *pdf* by definition is non-negative for all values of the random variable
- Below we use (1) to find the *pdf* of a linear function of a normally distributed X.

Normal distribution I

The random variable X has a normal distribution if the *pdf* is:

$$f_X(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right], \quad \sigma_X > 0 \tag{2}$$

which we write $X \sim N\left(\mu_x, \sigma_x^2\right)$.
X is fully characterised by the 1st and 2nd order moments μ_X and σ_x^2.
The *standardized normal variable* Z is defined by the function

$$Z = g(X) = \frac{X-\mu_x}{\sigma_x} \tag{3}$$

We can use (1) to find the *pdf* of Z:

Normal distribution II

$$f_Y(z) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\sigma_x z + \mu_x - \mu_x}{\sigma_x}\right)^2\right] \cdot |\sigma_x| \quad (4)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} z^2\right]$$

since

$$X = g^{-1}(X) = \sigma_x Z + \mu_x$$

$$\left|\frac{dg^{-1}(x)}{dz}\right| = |\sigma_x|$$

Normal distribution III

Since the pdf is a normal pdf with moments 0 and 1, we write:

$$Z \sim N(0, 1) \tag{5}$$

A linear combination of n independent normal variables $\{X_i; i = 1, 2, \ldots, n\}$

$$Y = a + \sum_{i=1}^{n} b_i X_i$$

has a normal distribution:

$$Y \sim N\left(a + \sum_{i=1}^{n} b_i \mu_{X_i}, \sum_{i=1}^{n} b_i^2 \sigma_{X_i}^2\right)$$

Normal distribution IV

It follows that the average of n identically distributed independent normal variables (IIN) is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \tag{6}$$

which can be standardized:

$$\frac{\bar{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} \sim N\left(0, 1\right), \text{ when } \left\{X_i; i = 1, 2, \ldots, n\right\} \; IIN \tag{7}$$

(7) is central in the inference theory for regression models that we will use.

Chi-square distribution I

Definition (Chi square)

If $Z_i \sim N(0,1)$, $i = 1, \ldots, n$. $U = \sum_{i=1}^{n} Z_i^2$ is Chi-square distributed (χ^2) with n degrees of freedom:

$$U \sim \chi^2(n)$$

- SW uses m at this point
- It follows from this definitions that: $Z^2 \sim \chi^2(1)$ for $Z \sim N(0,1)$
- The number of *degrees of freedom* (df) is the number of variables minus the number of restrictions between the variables

Chi-square distribution II

- For $Z_i \sim N(0,1)$, $i = 1, \ldots n$ in the definition, there are no restrictions, so $df = n$.
- If we instead consider $X_i \sim N\left(0, \sigma_x^2\right)$ $i = 1, 2, \ldots, n$ and the squares of $\frac{(X_i - \bar{X})}{\sigma_X}$ there will be one restriction between the n variables, since $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$. It can be shown (BN page 83) that:

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma_X^2} \sim \chi^2(n-1)$$

- It is not uncommon to use the notation for the sum of squares for normal variables by:

$$S^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2. \qquad (8)$$

Chi-square distribution III

- With this notation we have:

$$\frac{S^2}{\sigma_X^2} \sim \chi^2 \left(n-1\right). \tag{9}$$

which will be used in the construction of t-distributed random variables that we use to make inference.

- Expectation and variance:

$$E(\frac{S^2}{\sigma_X^2}) = (n-1) \quad Var(\frac{S^2}{\sigma_X^2}) = 2(n-1)$$

t-distribution I

Assume $Y \sim N(0,1)$, $U \sim \chi^2(v)$, and that Y and U are independent. Then

$$t = \frac{Y}{\sqrt{\frac{U}{v}}} \sim t(v) \tag{10}$$

is t-distributed with v degrees of freedom. $var(t) = v/(v-1)$ for $v \geq 3$

If $X_i \sim N\left(\mu_X, \sigma_X^2\right)$, $i = 1, 2, ..., n$ and independent: Then

$$Y = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1)$$

$$\frac{S^2}{\sigma_X^2} \sim \chi^2(n-1)$$

t-distribution II

as above and

$$\frac{\frac{\bar{X}-\mu_X}{\frac{\sigma_X}{\sqrt{n}}}}{\sqrt{\frac{S^2}{\sigma_X^2}}}\sqrt{n-1} = \frac{\bar{X}-\mu_X}{\sqrt{\frac{S^2}{n}}}\sqrt{n-1} \sim t\left(n-1\right). \qquad (11)$$

which is the basis for testing the hypothesis about the mean of normally distributed variables.

F-distribution I

- The last distribution we will need in this course is the F-distribution.
- If $U \sim \chi^2(v_1)$ and $V \sim \chi^2(v_2)$ are independent:

$$F = \frac{\frac{U}{v_1}}{\frac{V}{v_2}} = \frac{U}{V}\frac{v_2}{v_1} \sim F(v_1, v_2) \tag{12}$$

F-distribution II

- If $X_i \sim N\left(\mu_X, \sigma_X^2\right)$ $i = 1, 2, ..., n$ are independent, we have found that

$$Y = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1),$$

$$Y^2 \sim \chi^2(1), \text{ and}$$

$$\frac{S^2}{\sigma_X^2} \sim \chi^2(n-1).$$

combined with (12), we can conclude that

$$F = \frac{\frac{\left(\bar{X}-\mu_X\right)^2}{\frac{\sigma_X^2}{n}}}{\frac{S^2}{\sigma_X^2}}(n-1) \sim F(1, n-1).$$

F-distribution III

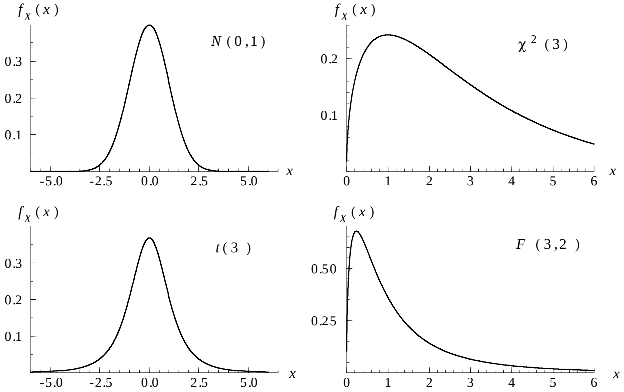
If you compare with the definition of $t(-1)$ in (11) you will see that:

$$t(n-1)^2 = F\left(1, n-1\right) \tag{13}$$

Generally, the square of a t-distributed variable with v degrees of freedom is F-distributed with 1 and v degrees of freedom.

$$t(v)^2 = F(1, v). \tag{14}$$

The F distribution is indispensable when we work with multiple regression models.



Four density functions: a) standard normal, b) χ^2 pdf with 3 degress of freedom, c) t- pdf with 3 df, d) F-pdf with 3 and 2 df.

Joint probability distribution

- In the probability approach to econometrics, the joint probability function is the main tool for modelling interactions between variables
- Read Ch2.3 in SW to discrete joint distributions. One of the exercises to Seminar 1 invites you to review discrete joint probability functions
- For two *continuous* random variables Y and X, the *joint pdf* is

$$f_{Y,X}(y,x) = \frac{\partial^2}{\partial y \partial x} P(Y \leq Y \text{ and } X \leq x).$$

$f_{Y,X}(Y,X)$ has the properties:

1. $f_{Y,X}(Y,X) \geq 0, \forall Y, X$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y,X}(Y,X)\, dydx = 1$

If and only if Y and X are independent:

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

Marginal and conditional pdf

The *pdf* for X can always be retrieved from the joint *pdf*:

$$f_X(x) = \int_{-\infty}^{\infty} f_{Y,X}(y,x)\,dy \equiv f_X(x) = \frac{d}{dx} P(X \leq x).$$

In this interpretation, $f_X(x)$ is called the *marginal* pdf.
Likewise, for Y we obtain the marginal pdf as:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y,X}(y,x)\,dx$$

The conditional probability density function for X given $Y = y$ is

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

$f_{X|Y}(x \mid y)$ is a valid *pdf*, meaning that the *conditional probability function* is

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(x \mid y)\,dx.$$

Definition of conditional expectation I

- To save some time and space we concentrate on the continuous random variable case.
- There is seminar exercise about the discrete variable case.
- Using the concepts that we have reviewed: The *conditional probability density function* (pdf) for Y given $X = x$ is

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} \tag{15}$$

where $f_{XY}(x, y)$ is the *joint pdf* for the two random variables X and Y, and $f_X(x)$ is the *marginal pdf* for X.

Definition of conditional expectation II

Definition (Conditional expectation)

Let Y be the random variable with conditional pdf $f_{Y|X}(y \mid x)$. The conditional expectation of Y is

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy = \mu_{Y|x}.$$

Conditional expectation function I

- For a given value of $X = x$ the conditional expectation $E(Y \mid X = x)$ is deterministic, it is a number.
- We can however consider the expectation of Y for the whole value set of X. In this interpretation, $E(Y \mid X)$ is a random variable with $E(Y \mid x)$ as a value for $X = x$.
- This line of reasoning motivates that the *conditional expectation function* $E(Y \mid X)$ is a function of the random variable X:

$$E(Y \mid X) = g_X(X)$$

Conditional expectation function II

- If the *conditional expectation function* is linear, we can write it as

$$E(Y \mid X) = \mu_{Y|X} = \beta_0 + \beta_1 X \quad (16)$$

which we shall see later is the essential part of a the *linear regression model* with Y as the *regressand* and X as the *regressor*.

- We next discuss three important aspects of conditioning:
  - Conditional variance
  - The law of iterated expectations
  - Linear independence of X and Y when $E(Y \mid X) = \textit{constant}$

Conditional variance I

- Since the conditional probability function is a valid probability function, we can also define higher order moments for the condtional pdf.
- Specifically, the conditional variance: $var(Y \mid X = x)$
- In a regression model, the variance of the *disturbance term*, the non-explained part, is interpretable as $var(Y \mid X = x)$
- For a relevant regression model we should expect that

$$var(Y \mid X = x) < var(Y)$$

Why?

Independence I

- Two events A and B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
- By analogy, two continuous random variables Y and X are independent if

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

- If Y and X are independent, the conditional expectation of Y is constant and equal to the marginal expectation

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy = \\ = \int_{-\infty}^{\infty} y \frac{f_{XY}(x,y)}{f_X(x)} dy = \int_{-\infty}^{\infty} y f_Y(y) = \mu_Y.$$

The law of iterated expectations I

- Let Y and X be two random variables and let $E(Y \mid X)$ be a conditional expectation function (not necessarily linear)
- the *Law of iterated (or double) expectations* says that:

$$E[E(Y \mid X)] = E(Y). \tag{17}$$

  as in equation (2.19) in SW.
- The law says that if we take the expectation over all the values that we first condition on, we obtain the unconditional expectation.
- Interpretation (i): If we use the probabilities of all the values that X can take, it does not matter what the value of X is

The law of iterated expectations II

- Interpretation (ii): $E(Y)$ is the probability weighted average of all the conditional means of Y given x.
- For discrete variables, the proof is by direct inspection, as on page 71 in SW.
- For continuous variables: Proof is by conditional pdf as in, for example BN exercise 4.12

Note the extension of (17) to "multivariate conditioning" (17) on page 72 in SW:

$$E\left[E\left(Y \mid X, Z\right)\right] = E\left(Y\right)$$

Covariance I

- The covariance between Y and X is defined as

$$cov\,(Y, X) = E\left[(Y - \mu_Y)\,(X - \mu_X)\right] \equiv \sigma_{Y,X} \equiv \sigma_{X,Y}$$

- $cov\,(Y, X) = 0$,if Y and X are independent. Proof is by way of using $f(Y, X) = f(Y)f(X)$ (as in exercise 4.6 in BN)
- But $cov\,(Y, X) = 0$ does not logically imply independence!!!!
- Often we will use one of the shorter forms:

$$cov\,(Y, X) = E\left[(Y - \mu_Y)\,X\right] = E\left[Y\,(X - \mu_X)\right]$$

Show!

Covariance II

- A third way of re-writing $cov\left(Y,X\right)$ is

$$\begin{aligned} cov\left(Y,X\right) &= E\left[\left(Y-\mu_Y\right)\left(X-\mu_X\right)\right] \\ &= E\left[YX - Y\mu_X - \mu_Y Y + \mu_Y\mu_X\right] \\ &= E\left(YX\right) - \mu_Y\mu_X. \end{aligned}$$

- We will need the variance of a sum of random variables

$$var(Y+X) = ?$$

Covariance III

$$
\begin{aligned}
var(Y+X) &= E\left[Y+X-E(Y+X)\right]^2 \\
&= E\left[(Y-E(Y))+(X-E(X))\right]^2 \\
&= var(Y)+var(X)+2cov(Y,X)
\end{aligned}
$$

which generalizes to

$$
\begin{aligned}
var(Y+X+Z) = {} & var(Y)+var(X)+var(Z) \\
& +2cov(Y,X)+2cov(Y,Z)+2cov(X,Z)
\end{aligned}
$$

and so on for longer sums
See Key concept 2.3 in SW

Correlation coefficient

$$\rho_{XY} = corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}}$$

$$= \frac{\sigma_{XY}}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- Using the different notations for *standard deviations* that we use.
- The bivariate normal distribution is completely described by the two expectations, the two variances and the correlation coefficient (alternatively the covariance)
- We say the bivariate normal distribution has five *parameters.*

- Above we saw that independence implies

$$E\left(Y \mid X = x\right) = \mu_Y \text{ (a constant)}$$

- On page 74 in SW it is shown that $E\left(Y \mid X = x\right) = \mu_Y \implies cov(X, Y) = 0$
- However, $cov(X, Y) = 0$ does not imply $E\left(Y \mid X = x\right) =$ a constant.
- See e.g. Exercise 2.23 in SW
- As we shall see, the implication follows if $E\left(Y \mid X = x\right)$ is a linear function ("the relationship between X and Y is linear").
- The case where $f\left(Y, X\right)$ is bivariate normal is an important case of this (we will study this when we get to regression)

Functions of random variables (again)

- As already noted, we often work with functions of random variables
- The main functions are
  - Sums and linear functions
  - Products
  - Ratios
- Above we gave the main tool, (1), for finding the distribution of $Y = g(X)$.
- But often it serves our purpose if we can find the 1st and 2nd moments of the new variable Y

Linear functions

- We have given the rules for E and Var of sums of random variables
- The rules for a linear function is a direct generalization/application:

$$Y = a + bX + cZ$$

$$\begin{aligned} E(Y) &= \mu_Y = a + b\mu_X + c\mu_Z \\ var(Y) &= \sigma_Y^2 = b^2\sigma_X^2 + c^2\sigma_Z^2 + 2bc\sigma_{X,Z} \end{aligned}$$

- See Key concept 2.3 in SW BN Kap 4.4 also gives more complete results for a system of linear equations.

Non-linear functions (the delta method)

- Consider

$$g(X, Y) = \frac{X}{Y}$$

- Since E and var are linear operators, we must first find a linear approximation to $g(X, Y)$.
- This is done by Taylor expansion (Sydsæter 2003, Kap 7).
- Many textboks, but not SW(?), now include it under the name *delta method*,
- In BN page 72-73 it is show that the following hold

$$E\left(\frac{X}{Y}\right) \approx \frac{\mu_X}{\mu_Y}, \tag{18}$$

$$var\left(\frac{X}{Y}\right) \approx \left(\frac{1}{\mu_Y}\right)^2 \left[\sigma_X^2 + \left(\frac{\mu_X}{\mu_Y}\right)^2 \sigma_Y^2 - 2\left(\frac{\mu_X}{\mu_Y}\right)\sigma_{X,Y}\right] \tag{19}$$

  under mild assumptions.
- Since many variables that interest us are ratios of observable variables (the "natural rate of unemployment" is one example that we shall look at), these equations are very relevant for applied work.
- We therefore keep this a (precious!) reference.