# Supplementary Notes: 'Noise'*

Economics of Education (ECO383)

October 2011

## Overview

This note provides a brief introduction to the way randomness is treated in econometrics. It contains some technical parts, which we will need in order to carry out inference in the regression model.

Building on our earlier discussions, we will focus on the error terms that we have been tacking on to equations like

$$T_i = \alpha + \beta C_i + u_i \tag{1}$$

in the class size-student performance case, or

$$S_i = A_i + u_i, \tag{2}$$

in the case of student scores and student ability. In each case, think of the error term as capturing (in a single number) the random unsystematic factors affecting student performance. If one plots data describing test scores and class size, for example, one will get a cloud of points. And the only way to rationalize such a cloud, given that we postulate a *linear* relation, is to invoke some noise in the education process. Simply put, we do not observe all the data points on a single line, so the error helps rationalize the data.

Statistics has developed ways of describing the varieties of underlying randomness using a very useful mathematical apparatus that begins with probability theory. It's amazing stuff – very elegant indeed. So we will take an introductory look, assembling some of the parts we will need for the next steps.

*Note by Robert McMillan, email: mcmillan@chass.utoronto.ca.

## Introduction to Random Variables

Coin-tossing provides a nice example of randomness in action, in ways that we will formalize shortly.

If we start with a 'fair' coin – one that could equally come up heads ('H' for short) or tails ('T') – then prior to tossing it, we do not know what the outcome will be. In advance, the precise outcome is uncertain. We can say, though, that there are two possible, mutually exclusive outcomes; the random outcome must be one or the other; and the chance that the coin will come up heads (or tails) is equal to one half. How so? Well, there are two possible outcomes, and an equal chance of each.[1]

Let the outcome of a single trial be denoted by the random variable $X$. (It is random in the above sense – that we cannot say for certain what the realized outcome will be prior to the coin tossing.) We will say that the probability that $X = `H'$ is equal to one half, and write this as $Pr(X = `H') = 1/2$. Analogously, we have that $Pr(X = `T') = 1/2$ also.

**Question:**

a) Suppose we toss the same fair coin three times, and the outcomes across the trials are independent – that is, the outcome of one trial does not have any bearing on the outcome of any others. What is the probability that we end up with three heads?

Solution: We have three trials. The probability that the outcome is heads is equal to a half for each trial. Given the independence assumption, the probability of three heads is $1/2 \times 1/2 \times 1/2 = (1/2)^3 = 1/8$.

b) What is the probability that we end up with just two heads?

Solution: Here, let us write down all possible sequences. There are eight (two on the first, two on the second, and two on the third trial, or $2 \times 2 \times 2 = 2^3 = 8$. We can write the eight possible outcomes as $HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$.

How many of these involve just two heads? By inspection, they are $HHT, HTH, THH$. That is, three possible outcomes out of a total of 8 involve just two heads. Thus the relevant probability is 3/8.

[1] Is a given coin *fair* in practice? We can assess this experimentally: toss the coin a very large number of times (trials). After each trial, record whether the answer was 'H' or 'T'. Then, after completing all $N$ trials, tally the overall proportion of heads and tails. With large enough $N$, the overall proportion of each should be arbitrarily close to 1/2.

# Types of Random Variable

We can divide random variables into two types: discrete and continuous random variables. Let us start with the former.

## Discrete Random Variables

The variable $X$ is a discrete if it takes on $J$ possible values, written $X = x_j$, each with probability $P(X = x_j) \equiv p_j$, such that $\sum_{j=1}^{J} p_j = 1$.

**Example:**

Let $X$ be a discrete random variable, distributed as follows:

$X = x_1 = 0$ with $p_1 = 0.5$; and $X = x_2 = 1$ with $p_2 = 0.5$.

Here, X takes on two values, and the sum of the discrete probabilities equals 1. (Note this is analogous to the coin tossing case, where (say) the heads outcome is designated "0" and the tails outcome is designated "1," and we have a fair coin.)

**Example – a Bernoulli random variable:**

$X = x_1 = 0$ with $p_1 = a$; and $X = x_2 = 1$ with $p_2 = 1 - a$, where $0 < a < 1$.

Here also, X takes on two values, and the sum of the discrete probabilities equals 1, though we are now generalizing the previous case to allow the probability of taking a value of zero to be any number $a$ between zero and one. A Bernoulli random variable is perhaps the simplest discrete random variable we can consider.

**Example:**

Let $X$ be a discrete random variable taking on three values:

$X = x_1 = -a$ with $p_1 = 0.25$; $X = x_2 = 0$ with $p_2 = 0.5$; $X = x_3 = a$ with $p_3 = 0.25$.

## Expectation

Generally speaking, the expectation of a discrete random variable $X$ is defined by

$$E(X) = \sum_{j=1}^{J} x_j P(X = x_j) = \sum_{j=1}^{J} x_j p_j.$$

**Properties of the Expectation Operator** $E(.)$

The following properties are easy to prove, whether $X$ is a discrete or continuous random variable.

- Let $X = \alpha$, a constant. Then $E(\alpha) = \alpha$. (Intuitively, the expected value of a quantity that never changes is the value of the quantity itself.)
- Let $Y = \alpha X$. Then $E(Y) = E(\alpha X) = \alpha E(X)$. (Intuitively, if all (random) values of $X$ are multiplied by a constant factor, then we would expect the mean of the random variable to be multiplied by that factor also.
- $E(X+Y) = E(X) + E(Y)$. This property is implied by the expectation operator being linear. It is easy to prove.

**Question:** Prove the last property (above) in the case of a discrete random variable.

**Question:** What is the variance of the Bernoulli random variable whose distribution is given above?

## Variance

Generally, the variance of a random variable $X$ is given by $Var(X) = E(X - E(X))^2$. Let $E(X) = \mu$, a constant, for simplicity. Then $Var(X) = E(X - \mu)^2$.

We can re-write this latter expression, which may be convenient in some cases. Expanding the bracket, we have $Var(X) = E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2)$. Now, $\mu$ is a constant, so the only random variable in the previous expression is $X$. The expectation operator is linear (see properties above), so we can break the last expression in the previous line into parts, and take the expectation of each, namely: $E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$.

But we know that $E(X) = \mu$, so $E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$.

In the case of a discrete random variable, the variance is given by

$$Var(X) = \sum_{j=1}^{J} (x_j - \mu)^2 P(X = x_j) = \sum_{j=1}^{J} (x_j - \mu)^2 p_j.$$

The standard deviation is simply the square root of the variance.

**Question: What is the variance of a Bernoulli random variable?**

**[Next up: continuous random variables...]**