The normal distribution – background*

The normal distribution is the most widely used distribution in statistics. It is convenient to use, for a variety of reasons. But there is a quite <u>remarkable</u> reason why it is particularly appealing. This is known as the Central Limit Theorem.**

This theorem states the following: Suppose we obtain data by sampling randomly from a population with mean μ and variance σ^2. Given the draws are random, let us write that the random variable X is distributed

$$X_i \sim \text{i.i.d.}\ (\mu, \sigma^2),$$

where X_i is a particular random draw from the given population and "i.i.d." means "identically and independently drawn." Then the sample mean $\bar{X}$, when appropriately standardized, follows a normal $(0,1)$ distribution, when N (the sample size) gets very large.

Specifically, construct $s^2 = \frac{1}{N-1} \sum (X_i - \bar{X})^2$,

---

* Not on any test in ECO 383!

** Actually, there are several such theorems.

an unbiased estimate of the sample variance.

Then the standardized quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{N}} \text{ tends to a } N(0,1) \text{ distribution}$$

as $N \to \infty$.

So what? What this central limit theorem says is that, regardless of the distribution the sample is drawn from, the sample average is approximately normally distributed for large samples.

[Proving this theorem would be interesting, not least because it would yield the normal p.d.f.]

In nature, the normal distribution is actually encountered a lot! Even if individual random influences are <u>not</u> normally distributed, the mean of those influences <u>is</u> normally distributed. And that is a truly remarkable thing...