14.32 Recitation 4

Paul Schrimpf

February 26, 2009

- Review of CLT
- F-test
- Code for confidence interval simulations

1 Central Limit Theorem

Basically, the central limit theorem says that in large samples, the sample mean is normally distributed. More precisely[1], the simplest version of the central limit theorem is something like:

Theorem 1. *If $x_1, ..., x_n$ are i.i.d. with finite mean, μ, and variance, σ^2, then*

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \mu^2)$$

Definition 2. Convergence in distribution, denoted by $x_n \xrightarrow{d} x$, (aka "weak convergence"), means that $\lim_{n\to\infty} F_n(a) = F(a)$ at all points where $F(\cdot)$ is continuous, where $F(\cdot)$ is the cdf of x and $F_n(\cdot)$ is the cdf of x_n

We already saw a demonstration of the CLT by simulation. We can do a similar demonstration by exact calculation for the Bernoulli random variables. The mean of a sample of n Bernoulli(p) variables has to $\frac{k}{n}$ for $k \in \{0, 1, ..., n\}$. This happens when we get k successes, each of which occurs with probability p and $n-k$ failures, each of which occurs with probability $1-p$. Then, there n choose k (written $\binom{n}{k} = \frac{n!}{(n-k)!k!}$) ways of picking the k successes, so the the distribution of the mean is

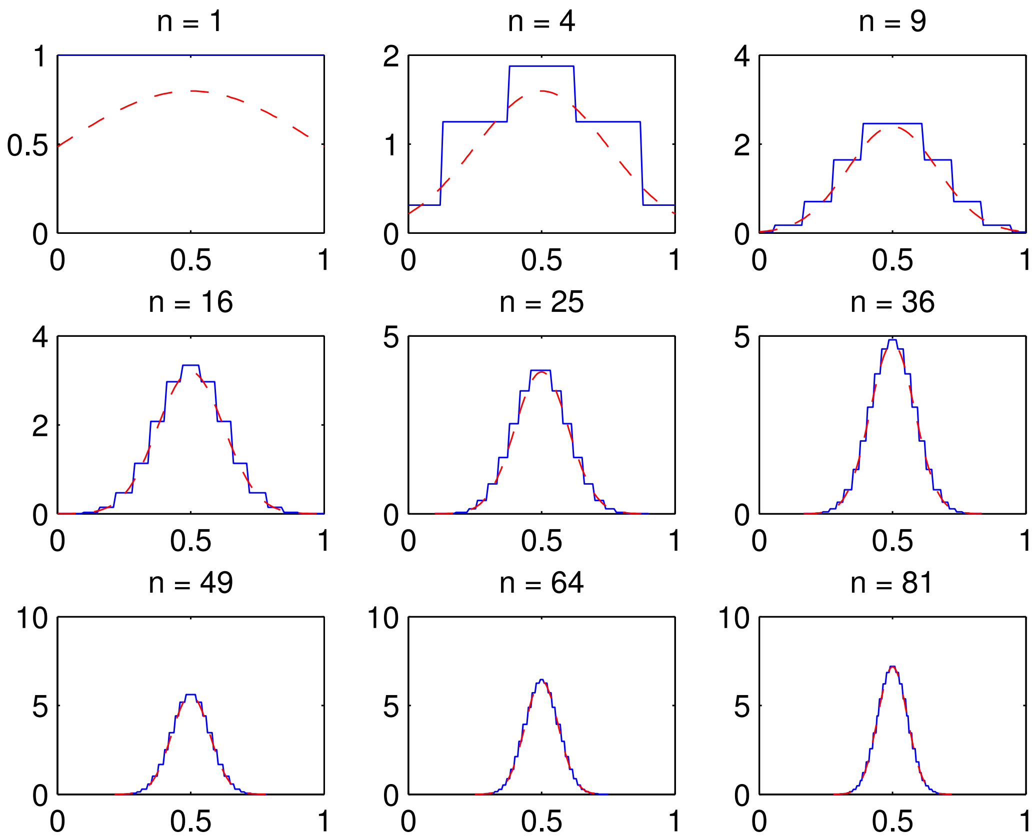
$$P\left(\bar{x}_n = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

Figure 1 plots this distribution alone with the normal approximation suggested by the CLT for various values of n.

The simple version of the CLT given above assume that the data is iid. There are many situations were we do not think our data is identically distributed, and there are also many situations where we do not think the data is independent. Fortunately, there are versions of the CLT that apply in these cases as well. The exact conditions of these CLTs are a bit involved, but basically as long as the data is not too dependent or too heterogenous, then some CLT applies.

[1] I like precision, but if this confuses you, you probably don't need to worry about

Figure 1: CLT for Bernoulli(0.5)



2 F-test

In class yesterday, we saw how you can get a slightly different t-statistic for testing equality of two means depending on whether you are willing to assume the two groups have the same variance. In particular if you are willing to assume equal variances, then you estimate the variance more accurately by pooling the groups together. In practice, it is unlikely to make much of a difference for your t-tests whether you assume equal or unequal variances. Nonetheless, there are situations where you might be interested in testing whether the variances for two groups are equal. For example, we might be interested in comparing the riskiness of two types of assets, and variance can be a measure of risk. Similarly, if we want to evaluate the effectiveness of some policy meant to provide insurance, we might want to know whether the policy reduced variance. The same could be said for policies that target inequality.

In any case, the usual way to test $H_0 : \sigma_1^2 = \sigma_2^2$ is with an F-test. Our F-test statistic is simply the ratio of the sample variances

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\frac{1}{n_1}\sum(x_{1,i} - \bar{x}_1)^2}{\frac{1}{n_2}\sum(x_{2,i} - \bar{x}_2)^2} \tag{2}$$

The F-distribution is defined as the ratio of two χ^2 random variables. More specifically, we say that $y \sim F(d_1, d_2)$ if $y = \frac{w_1/d_1}{w_2/d_2}$ where $w_1 \sim \chi^2(d_1)$ and $w_2 \sim \chi^2(d_2)$. When $x_i \sim N(\mu, \sigma^2)$, we know that $\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-1)$. Therefore, the F-stat is distributed as $F(n_1 - 1, n_2 - 1)$. You can do two-tailed hypothesis tests at level α by checking whether the F-stat is between the $\alpha/2$ and $1 - \alpha/2$ quantiles of $F(n_1 - 1, n_2 - 1)$. One-tailed tests mean checking whether the F-stat is greater than the α quantile (or less thean the $1 - \alpha$ quantile).

You might be wondering why we test for differences in variances by looking at their ratio and test for differences in means by looking at the difference. That's a good question. A nice intuitive explanation is that means are measures of location, while variances are measures of scale. It sort of makes sense to measure the difference in location by the arithmetic difference, and it makes sense to measure the difference in scale by ratios. I'm fairly certain that there is also a deeper answer – likely the F-test is the uniformly most powerful invariant test when the data is normally distributed. By invariant, we mean that the test gives the exact same answer after we apply a linear transformation to our data. By most powerful, we mean that the test maximizes the probability of rejecting the null when the null is false. In general, the Neyman-Pearson lemma says that likelihood ratio tests are most powerful. A bit of algebra could probably show that the F-test is equivalent to a likelihood ratio test. I'll add the algebra if I have time, if not it's left as an exercise.

In Stata, the command sdtest var, by(group) does F-tests. The help for this command notes that F-tests are quite sensitive to departures from normality and recommends using robtest instead. In economics, we usually have such large samples that we do not worry too much about departures from normality. However, small samples are more common in the experimental sciences, and so there is a large literature on robust testing. The idea is to come up with a test that has the right size for many different distributions even in small samples and is nearly as powerful as the most powerful test for any particular distribution.

3 Code for Confidence Interval Simulations

See do files.