# Economics 120A

# 9: Confidence Intervals

Graham Elliott

# Confidence Intervals

We have seen

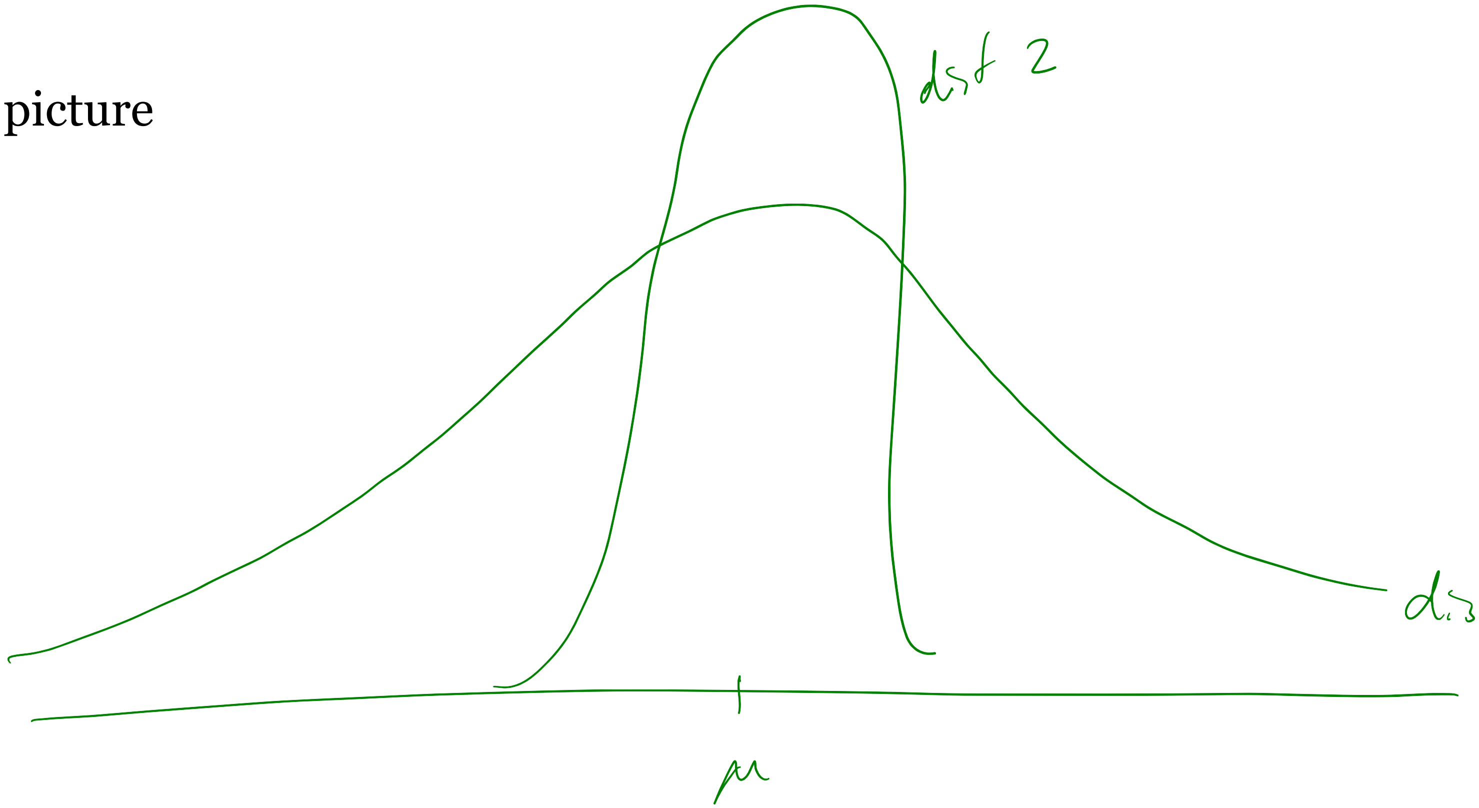
(a) We can estimate μ using the sample mean $\bar{X}$ or
(b) We can test if the data (based on $\bar{X}$) suggests that some theoretical μ is reasonable.

Suppose we do not have any idea what a reasonable μ might be, we could use the sample mean $\bar{X}$ as an estimator but we have no idea if it is really close to the true mean or not.

Example: Poll: We might have "59% for" and "41% against" but how accurate are these numbers?

# Confidence Intervals

## In a picture



# Basic Idea

The idea of hypothesis tests is to present a range of possible values for the parameter with the plan to be able to be precise as to how likely it is that the true parameter is included in the interval.

So we want to report functions of the data $\{L(X_1, X_2, \dots, X_n), U(X_1, X_2, \dots, X_n)\}$ so that we can know

$$P[L(X_1, X_2, \dots, X_n) < \mu < U(X_1, X_2, \dots, X_n)]$$

is known (presumably set to a high number).

$\{L(\cdot), U(\cdot)\}$

So the question is how to do this.

# Inverting Hypothesis tests

Recall that we fail to reject that a particular μ is compatible with the data in a two tail test whenever $|t| \leq cv$.

But this is

$$P\left[\,|t| < cv\,\right] = (1-\alpha)$$

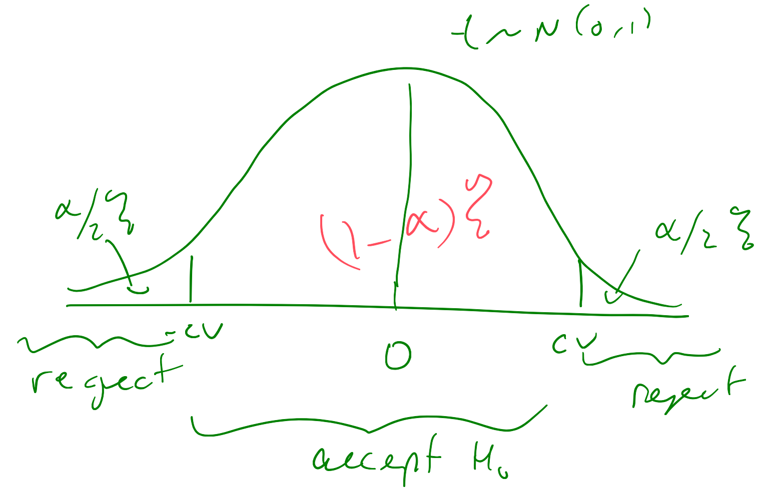
$$= P\left[-cv < t < cv\right]$$

$$= P\left[-cv < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < cv\right]$$

$$= P\left[-cv\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < cv\frac{\sigma}{\sqrt{n}}\right]$$

$$= P\left[-\bar{x} - cv\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + cv\,\sigma/\sqrt{n}\right]$$

$$= P\left[\underbrace{\bar{x} - cv\frac{\sigma}{\sqrt{n}}}_{L(X_1 \dots X_n)} < \mu < \underset{U(X_1 \dots X_n)}{\bar{x} + cv\frac{\sigma}{\sqrt{n}}}\right] = (1-\alpha)$$



# Inverting Hypothesis tests

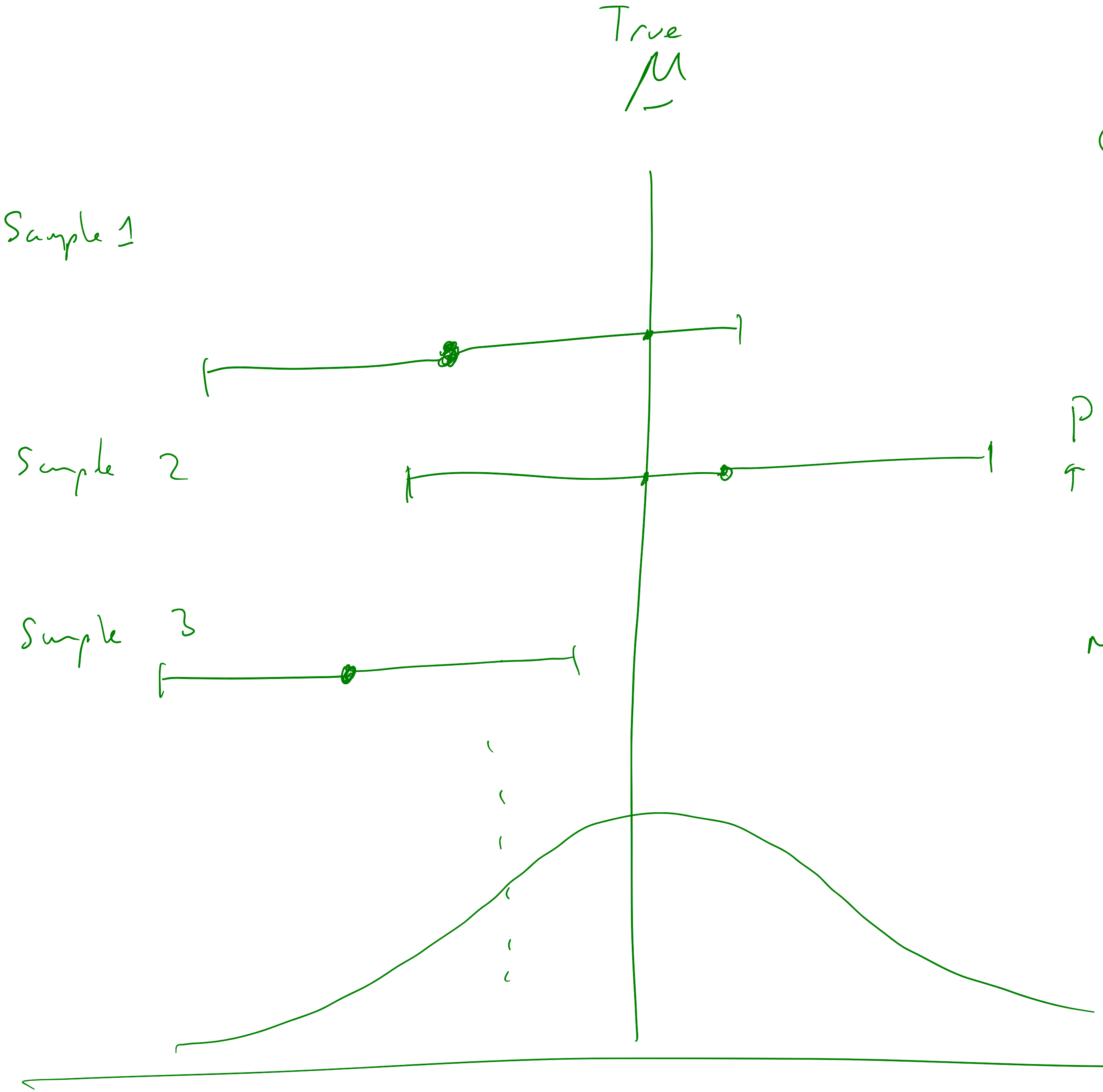
The general form of confidence intervals for statistics that satisfy the CLT is

estimate ± cv * standard error of the estimate.

You will see this over and over again in 120b and 120c.

Interpretation: The CI of level (1-α)% is the set of null hypothesis values for μ that we would fail to reject in a two sided test of size α%.

The 'cuttoff' is exactly that t statistic that would be on the border between rejection and acceptance.




Confidence interval either contains $\mu$ or it doesn't.

$$P\left[\bar{x} - c_v \sigma/\sqrt{n} < \mu < \bar{x} + c_v \frac{\sigma}{\sqrt{n}}\right] = (1-\alpha)\%$$

mean: is that $(1-\alpha)\%$ of the times CI includes $\mu$.

# Example: Cereal Box Example

$\bar{x} = 16.2$

$\bar{x} \pm cv \sqrt{\sigma^2/n}$

Each box had $X_i \sim N(\mu, (0.4)^2)$, and n=25.

cv is cv in 2-sided test of size $\alpha$% if CI coverage is $(1-\alpha)$%

$\alpha = 5\%$

cv is from normal at $\alpha/2 = 2.5\%$

$(\alpha = 5\%,\ cv = 1.96)$

So our confidence interval is

$$\left\{ \bar{x} \pm cv \cdot \sigma/\sqrt{n} \right\}$$

$$= \left\{ 16.2 \pm 1.96 \frac{0.4}{5} \right\}$$

$$= \left\{ 16.2 \pm 0.16 \right\}$$

$$= \left\{ 16.04,\ 16.36 \right\}$$

Notice that since we have that the sample mean is exactly normally distributed here, then the coverage for the confidence interval is exact.

# Example: Efficient Markets Example

$\bar{x} = 0.02$

Each trade had $X_i \sim (\mu, \sigma^2)$, and n=61. We estimated $\sigma^2$ using $s^2$, which was 0.05.

Here we approximate the distribution for t, i.e. $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim^a N(0,1)$

So our confidence interval is

$cv = 1.96 \quad (\alpha = 5\%)$

$\{\bar{x} \pm cv \cdot s/\sqrt{n}\}$

$= \{0.02 \pm 1.96 \sqrt{\frac{0.05}{61}}\}$

$= \{0.02 \pm 0.056\}$

$= \{-0.036, 0.076\}$

Approximate 95% CI.

# Example: Poll Example

closing $\pi = 0.54$

Here $X_i \sim (\pi, \pi(1-\pi))$, and n=507. ~~We estimated $\sigma^2$ using $s^2$, which was 0.05.~~

Here we approximate the distribution for t, i.e. $t = \frac{\bar{X}-\pi}{\sqrt{\pi(1-\pi)/n}} \sim^a N(0,1)$

Coverage = 95%

cv = 1.96

So our confidence interval is

$$\left\{ \bar{\pi} \pm cv \sqrt{\frac{\pi(1-\pi)}{n}} \right\}$$

(use $\bar{\pi}(1-\bar{\pi})$ for $\pi(1-\pi)$)

$$= \left\{ 0.54 \pm 1.96 \sqrt{\frac{0.54 \times 0.46}{507}} \right\}$$

$$= \left\{ 0.54 \pm 1.96 \times 0.022 \right\}$$

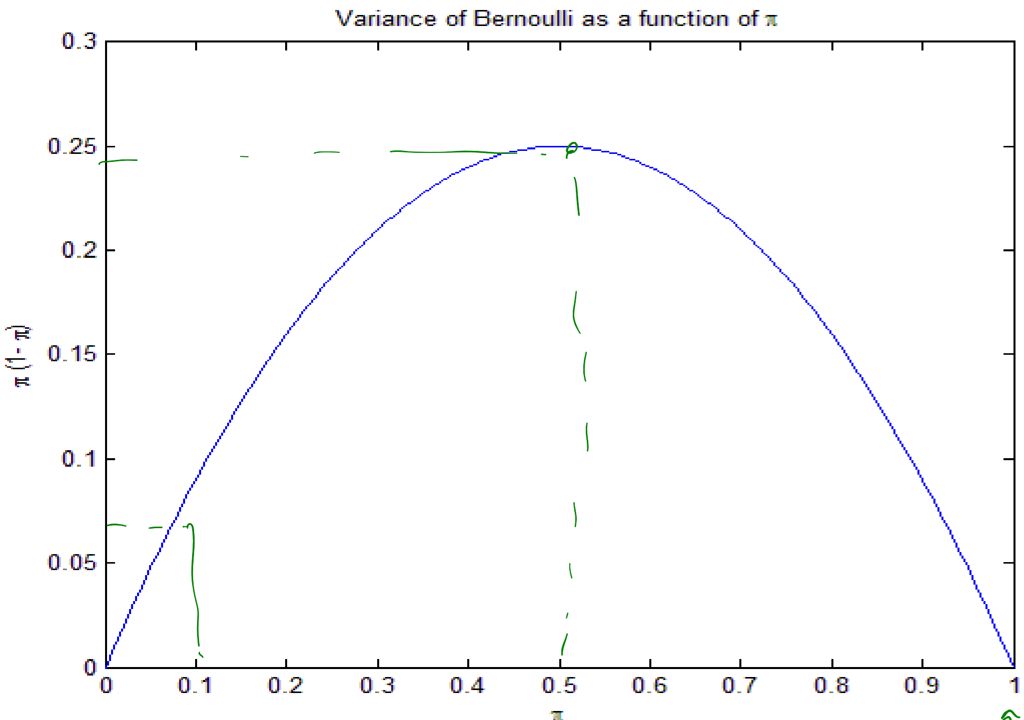
$$= \left\{ 0.497, 0.583 \right\}$$

# Proportions

There is one interesting issue that arises when we have a poll.

$\sigma^2$ ?

Since π ranges from zero to one, we can consider the range of π(1-π) as a function of π.




so why not use ½(1-½)?

It is a parabola starting at each end at zero and peaking at π=0.5 which gives π(1-π)=0.25. This is the worst case scenario for the standard error in a proportions problem.

# Proportions

Suppose that we just worked with the worst case scenario.

Our confidence interval for a level of 95% then would be

$$\bar{x} \pm 1.96\sqrt{\frac{0.5*0.5}{n}}$$

and so the amount we add and subtract now depends only on the sample size.

If n=1000 (a typical poll size), we have

$$1.96\sqrt{\frac{0.5*0.5}{n}} = 0.031$$

This is called the Margin of Error (MoE), and is the part you add and subtract to the estimate to get the CI.

# Proportions

$\{\bar{x}, n\}$ $\{x, M.E\}$