

Statistics and Inference

Paul Schrimpf

January 23, 2018

Contents

1	Properties of estimators	2
1.1	Unbiasedness	2
1.2	Variance of estimators	3
2	Finite sample inference	4
3	Asymptotics	5
3.1	Convergence in probability	5
3.2	Convergence in distribution	8
4	Hypothesis testing	11
4.1	Examples	12
4.2	Asymptotic interpretation of p-values	17

References

- Wooldridge (2013) appendix C
 - Stock and Watson (2009) chapter 3
 - Angrist and Pischke (2014) chapter 1 appendix
 - Diez, Barr, and Cetinkaya-Rundel (2012) chapters 4, 5, & 6
 - Linton (2017) Part II
-

Wooldridge (2013) and Angrist and Pischke (2014) provide brief overviews. Stock and Watson (2009) chapter 3 is more detailed, and is likely the best reference for this topic. Diez, Barr, and Cetinkaya-Rundel

(2012) is more introductory with many examples and not much mathematical rigor. Linton (2017) is available online through UBC library. It is very nice, but somewhat more mathematically technical than what we will cover.

Statistics

- Interested in *parameters* of population, e.g.
 - Moments, functions of moments
 - Conditional expectation functions
 - Distribution functions
 - Learn about parameters by observing sample from the population
 - Use sample to form estimates of parameters
-

1 Properties of estimators

1.1 Unbiasedness

Unbiasedness

- Setup:
 - Sample $\{y_1, \dots, y_n\}$
 - Parameter θ
 - Estimator W , some function of sample
 - The *bias* of W is $\text{Bias}(W) = E[W] - \theta$
 - W is *unbiased* if $E[W] = \theta$
 - Examples:
 - Sample mean: $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is unbiased
 - Sample variance: $s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ is biased
-

An example of a parameter is the population mean, $E[Y]$. An estimator for the population mean is the sample mean, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Another (less useful) estimator of the population mean is just the first observation, y_1 . Samples are random, and estimators are functions of a sample, so estimators are random variables. Therefore, it makes sense to think about the distribution of an estimator, the mean of an estimator, etc.

For an estimator to be useful, it should be related to a parameter in some nice way. One thing we might want to know about an estimator is whether on average, it will equal the parameter we want. If it does, the estimator is unbiased.

Here we show that the sample mean is unbiased:

$$\begin{aligned}
 E[\bar{y}] &= E\left[\frac{1}{n} \sum_{i=1}^n y_i\right] && \text{definition of } \bar{y} \\
 &= \frac{1}{n} \sum_{i=1}^n E[y_i] && \text{linearity of expectation} \\
 &= \frac{1}{n} \sum_{i=1}^n E[Y] && \text{assume expectation of } y \text{ does not depend on } i \\
 &= E[Y]
 \end{aligned}$$

Calculating the bias of the sample variance:

$$\begin{aligned}
 E\left[\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2\right] &= E\left[\frac{1}{n} \sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2)\right] \\
 &= \frac{1}{n} \sum_{i=1}^n E[y_i^2] - 2E\left[\left(\frac{1}{n} \sum_{i=1}^n y_i\right)\left(\frac{1}{n} \sum_{j=1}^n y_j\right)\right] + E\left[\left(\frac{1}{n} \sum_{i=1}^n y_i\right)\left(\frac{1}{n} \sum_{j=1}^n y_j\right)\right] \\
 &= E[Y^2] - E\left[\left(\frac{1}{n} \sum_{i=1}^n y_i\right)\left(\frac{1}{n} \sum_{j=1}^n y_j\right)\right] \\
 &= E[Y^2] - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E[y_i y_j] \\
 &= E[Y^2] - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \begin{cases} E[Y]E[Y] & \text{if } i \neq j \\ E[Y^2] & \text{if } i = j \end{cases} \\
 &= E[Y^2] - \frac{1}{n^2} (nE[Y^2] + n(n-1)E[Y]^2) \\
 &= \frac{(n-1)}{n} (E[Y^2] - E[Y]^2)
 \end{aligned}$$

where we assumed independent observations in the fifth line above. So the sample variance (if we divide by n) is biased. An unbiased estimate of the sample variance is given by dividing by $n - 1$ instead of n , i.e. $\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$

1.2 Variance of estimators

Bias is not the only property we should want our estimators to have. Both the sample mean and just the first observation are unbiased estimators of the population mean. However, we should expect (and it is true) that the sample mean is more likely to be close to the population mean than just the first observation. The variance and mean square error of an estimator are two ways of quantifying how variable is an estimator.

- The *variance* of W is $\text{Var}(W)$
 - The *standard error* of W is $\sqrt{\text{Var}(W)}$
 - If W_1 and W_2 are two unbiased estimators, then W_1 is relatively efficient if $\text{Var}(W_1) \leq \text{Var}(W_2)$
 - The *mean square error* of W is $\text{MSE}(W) = E[(W - \theta)^2]$
 - $\text{MSE}(W) = \text{Bias}(W)^2 + \text{Var}(W)$
 - Prefer estimators with that are relatively efficient and/or have low MSE
-
-

Variance of sample mean

- Sample mean: $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
- Variance of sample mean:

$$\begin{aligned}
 \text{Var}(\bar{y}) &= \text{Var} \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \\
 &= \text{Cov} \left(\frac{1}{n} \sum_{i=1}^n y_i, \frac{1}{n} \sum_{i=1}^n y_i \right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(y_i, y_j) \\
 &\quad \text{assume independent observations} \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(y_i) \\
 &= \frac{1}{n} \text{Var}(Y)
 \end{aligned}$$

- Standard error of sample mean: $\text{SE}(\bar{y}) = \sqrt{\text{Var}(Y)/n}$
-

2 Finite sample inference

Knowing estimators' variances and MSEs is useful for choosing which estimator to use. Once we have chosen an estimator and calculated it for a given sample, we would like to know something about how "close" our estimate is to the parameter we're interested. Statistical inference is the tool for doing this. Estimators are functions of a sample of random variables, so estimators themselves are random variables. If we draw different samples, or repeat our experiment, our estimator will take on different values. Therefore, we can think about describing the distribution of an estimator.

Inference

- Inference: (roughly) how close are our sample estimates to the population parameters
 - Example: N independent observations distributed Bernoulli(p), $X_i = 1$ if success, else 0
 - Joint distribution $f(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{N-\sum x_i}$
 - Could use to compute how surprising sample \bar{x} that we observe is if the true $p = p_0$, e.g. compute $P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - p_0\right| \geq |\bar{x} - p_0|\right)$
 - Example: x_i normally distributed, can do inference on sample mean using t-tests
 - Problem: usually do not know distribution of sample and/or distribution of estimator intractable
-

In a few cases, such as when our sample consists of Bernoulli or normal random variables, we can calculate the exact distribution of some estimators. Unfortunately, most of the time, calculating the exact distribution of an estimator is either very complicated or requires too many strong assumptions.

3 Asymptotics

When we cannot calculate the exact finite sample distribution of an estimator, we must try to approximate the distribution somehow. Fortunately, in large samples, most estimators have a distribution that can be calculated. We can use this large sample distribution to approximate the finite sample distribution of an estimator.

Asymptotic inference

- Idea: use limit of distribution of estimator as $N \rightarrow \infty$ to approximate finite sample distribution of estimator
 - Notation:
 - Sequence of samples of increasing size n , $S_n = \{y_1, \dots, y_n\}$
 - Estimator for each sample W_n
 - References:
 - Wooldridge (2013) appendix C
 - Menzel (2009) especially VI-IX
-

3.1 Convergence in probability

Loosely speaking you can think of probability limits as the large sample version of expectations.

Convergence in probability

- W_n converges in probability to θ if for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|W_n - \theta| > \epsilon) = 0$$

denote by $\text{plim } W_n = \theta$ or $W_n \xrightarrow{p} \theta$

- Show using *law of large numbers*: if y_1, \dots, y_n are i.i.d. with mean μ , or if y_1, \dots, y_n have finite expectations and $\sum_{i=1}^{\infty} \frac{\text{Var}(y_i)}{i^2}$ is finite, then $\bar{y} \xrightarrow{p} \mu$
 - Properties:
 - $\text{plim } g(W_n) = g(\text{plim } W_n)$ if g is continuous (*continuous mapping theorem (CMT)*)
 - If $W_n \xrightarrow{p} \omega$ and $Z_n \xrightarrow{p} \zeta$, then (*Slutsky's lemma*)
 - * $W_n + Z_n \xrightarrow{p} \omega + \zeta$
 - * $W_n Z_n \xrightarrow{p} \omega \zeta$
 - * $\frac{W_n}{Z_n} \xrightarrow{p} \frac{\omega}{\zeta}$
 - W_n is a *consistent* estimate of θ if $W_n \xrightarrow{p} \theta$
-

We will not need to directly work with the definition of convergence in probability. It is stated on the slide mostly for completeness. We will use the law of large numbers frequently. During this course, we will not worry about the precise assumptions needed for the law of large numbers to hold. We will just take as given that if $E[y_i] = E[Y]$ for all i , then $\frac{1}{n} \sum_{i=1}^n y_i \xrightarrow{p} E[Y]$, or in other words, sample averages converge in probability to population expectations.

Probability limits are much easier to work with than expectations, mostly because of the continuous mapping theorem and the second and third parts of Slutsky's lemma. These properties of probability limits do not apply to expectations.

Demonstration of LLN

```
1 n <- seq(1,200) ## sample sizes
2 sims <- 10 ## number of simulations for each sample size
3 rand <- function(N) { return(runif(N)) }
4 ## simulation
5 sampleMean <- matrix(0, nrow=length(n), ncol=sims)
6 for(i in 1:length(n)) {
7   N = n[i];
8   x <- matrix(rand(N*sims), nrow=N, ncol=sims)
9   sampleMean[i,] = colMeans(x)
10 }
11 ## plot sample means
12 library(ggplot2)
13 library(reshape)
14 df <- data.frame(sampleMean)
```

```

15 df$n <- n
16 df <- melt(df, id="n")
17 llnPlot <- ggplot(data=df, aes(x=n, y=value, colour=variable)) +
18   geom_line() + scale_colour_brewer(palette="Paired") +
19   scale_y_continuous(name="sample mean") +
20   guides(colour=FALSE)
21 llnPlot

```

Demonstration of LLN



Convergence in probability: examples

- $\bar{y} \xrightarrow{p} E[y]$ by LLN
- Sample variance:

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= \underbrace{\frac{1}{n} \sum_{i=1}^n y_i^2}_{\xrightarrow{p} E[y^2] \text{ by LLN}} - \underbrace{\bar{y}^2}_{\xrightarrow{p} E[y]^2 \text{ by LLN and CMT}} \\
 &\xrightarrow{p} E[y^2] - E[y]^2 = \text{Var}(y)
 \end{aligned}$$

- Mean divided by variance:

$$\begin{aligned} \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n (y_i - \bar{y})^2} &= \frac{\frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{\bar{y}}{\hat{\sigma}^2} \\ &\xrightarrow{p} \frac{E[y]}{\text{Var}y} \text{ by above and Slutsky's lemma} \end{aligned}$$

Convergence in probability: examples

- Sample correlation:

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

- $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \xrightarrow{p} \text{Var}(X)$ and $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \xrightarrow{p} \text{Var}(Y)$ as above
 - $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \xrightarrow{p} \text{Var}(X)\text{Var}(Y)$ by Slutsky's lemma
 - $\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \xrightarrow{p} \sqrt{\text{Var}(X)\text{Var}(Y)}$ by CMT
 - $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \xrightarrow{p} \text{Cov}(X, Y)$ by similar reasoning as for sample variance
 - $\hat{\rho} \xrightarrow{p} \text{Corr}(X, Y)$ by Slutsky's lemma
-

3.2 Convergence in distribution

Convergence in probability tells us that an estimator will eventually be very close to some number (hopefully the parameter we're trying to estimate), but it does not tell us about the distribution of an estimator.

From the law of large numbers, we know that $\bar{y}_n - \mu$ converges to the constant 0. The central limit theorem tells us that if we scale up the difference by \sqrt{n} , then $\bar{y}_n - \mu$ converges to normal distribution with variance $\text{Var}(y) = \sigma^2$.

Convergence in distribution

- Let F_n be the CDF of W_n and W be a random variable with CDF F
- W_n converges in distribution to W , written $W_n \xrightarrow{d} W$, if $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ for all x where F is continuous
- *Central limit theorem:* Let $\{y_1, \dots, y_n\}$ be i.i.d. with mean μ and variance σ^2 then $Z_n = \sqrt{n} \frac{\bar{y}_n - \mu}{\sigma}$ converges in distribution to a standard normal random variable

- As with the LLN, the i.i.d. condition can be relaxed if additional moment conditions are added

- **Demonstration**

- Properties:

- If $W_n \xrightarrow{d} W$, then $g(W_n) \xrightarrow{d} g(W)$ for continuous g (*continuous mapping theorem (CMT)*)
- Slutsky's theorem: If $W_n \xrightarrow{d} W$ and $Z_n \xrightarrow{p} c$, then (i) $W_n + Z_n \xrightarrow{d} W + c$, (ii) $W_n Z_n \xrightarrow{d} cW$, and (iii) $W_n/Z_n \xrightarrow{d} W/c$

The central limit theorem can be paraphrased as “scaled and centered sample means converge in distribution to normal random variables.” “Scaling” refers to the fact that we multiply by \sqrt{n} . “Centering” refers to the subtraction of the population mean. If an estimator converges in distribution to a normal random variable, then we often say that the estimator is asymptotically normal.

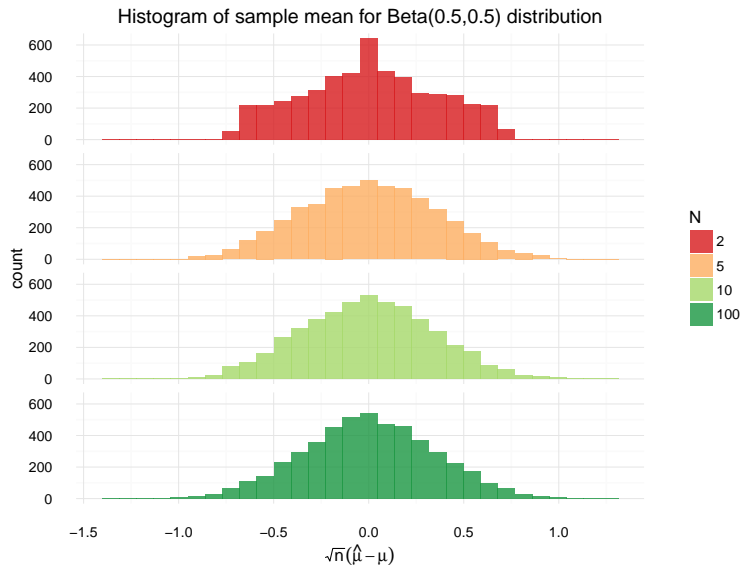
Demonstration of CLT

```

1 N <- c(2,5,10,100)
2 simulations <- 5000
3 means <- matrix(0,nrow=simulations,ncol=length(N))
4 for(i in 1:length(N)) {
5   n <- N[i]
6   dat <- matrix(rbeta(n*simulations,
7                     shape1=0.5,shape2=0.5),
8                 nrow=simulations,ncol=n)
9   means[,i] <- (apply(dat,1,mean) - 0.5)*sqrt(n)
10 }
11
12 # Plotting
13 df <- data.frame(means)
14 df$n <- N
15 df <- melt(df)
16 cltPlot <- ggplot(data=df, aes(x=value, fill=variable)) +
17   geom_histogram(alpha=0.8, position="identity") +
18   scale_x_continuous(name=expression(sqrt(n)(hat(mu) - mu))) +
19   scale_fill_brewer(type="div", palette="RdYlGn",
20                     name="N",label=N) +
21   labs(title="Histogram of sample mean for Beta(0.5,0.5) distribution") +
22   facet_grid(variable ~., label=function(x) {""}) + theme_minimal()

```

Demonstration of CLT



CLT examples

- Mean: $\sqrt{n} \frac{\bar{y} - \mu}{\sigma} \xrightarrow{d} N(0, 1)$ by CLT
- What if σ estimated instead of known?

$$\sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}}$$

- $\hat{\sigma} \xrightarrow{p} \sigma$ from before
- $\sqrt{n}(\bar{y} - \mu) \xrightarrow{d} N(0, \sigma^2)$ by CLT
- Slutsky's theorem with $W_n = \sqrt{n}(\bar{y} - \mu)$, $Z_n = \hat{\sigma}$ gives

$$\sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

CLT examples

- Variance:

-

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - \mu + \mu - \bar{y})^2 \\ &= \left[\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 \right] - (\bar{y} - \mu)^2\end{aligned}$$

- So

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) = \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 - \sigma^2 \right] - \sqrt{n}(\bar{y} - \mu)(\bar{y} - \mu)$$

- $\sqrt{n}(\bar{y} - \mu) \xrightarrow{d} N(0, \sigma^2)$, $(\bar{y} - \mu) \xrightarrow{p} 0$, so $\sqrt{n}(\bar{y} - \mu)(\bar{y} - \mu) \xrightarrow{d} 0$
- CLT implies $\sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2 - \sigma^2 \right] \xrightarrow{d} N(0, V)$

4 Hypothesis testing

The main way we will use the central limit theorem is to use the asymptotic distribution of an estimator to approximate the estimator's finite sample distribution. We do this so that we can make (approximate) statements about how uncertain we are about a given estimate.

Hypothesis testing is one way we make statements about how random is an estimate. The logic of hypothesis testing is slightly awkward. To understand it, keep in mind that we are trying to learn about a parameter. Parameters are fixed characteristics of a population. They are not random. Therefore, it makes no sense to talk about the probability that a parameter is a certain value or in a certain range. What is random, is the sample that we observe and the estimate that we calculate from the sample. Therefore, we can calculate the probability that our estimate is a certain value in some range. The probability that an estimate takes on a value depends on the true population parameter. Therefore, we generally quantify the randomness of an estimate by making statements like, assuming the true parameter is θ_0 , then the probability of getting an estimate as far or farther from θ_0 than the one we observe is p .

Hypothesis testing

- *Null hypothesis, H_0*
- *p-value*: if null hypothesis what is the probability of a sample as or more extreme than what is observed
- From 325: if $y_i \sim N(\mu, \sigma^2)$, then $\sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}^2} \sim t(n - 1)$

4.1 Examples

Example: testing mean

- From 325:

- If $y_i \sim N(\mu, \sigma^2)$, then $\sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}} \sim t$ distribution with $n - 1$ degrees of freedom
- p-value for $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ is

$$\begin{aligned} p &= P \left(\left| \frac{\frac{1}{n} \sum_{i=1}^n y_i - \mu_0}{s_y} \right| \geq \left| \frac{\bar{y} - \mu_0}{\hat{\sigma}} \right| \right) \\ &= 2 F_t \left(- \left| \frac{\bar{y} - \mu_0}{\hat{\sigma}} \right| ; n - 1 \right) \end{aligned}$$

CDF of t distribution

- What if distribution of y_i unknown?
 - Use CLT to approximate distribution of t-statistic
-
-

Example: testing one mean

- From earlier slide,

$$\sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

- So,

$$P \left(\left| \frac{\frac{1}{n} \sum_{i=1}^n y_i - \mu_0}{s_y} \right| \geq \left| \frac{\bar{y} - \mu_0}{\hat{\sigma}} \right| \right) \rightarrow 2 \underbrace{\Phi \left(- \left| \frac{\bar{y} - \mu_0}{\hat{\sigma}} \right| \right)}_{\text{standard normal CDF}}$$

Data on deworming treatment and school participation from **Miguel and Kremer (2003)**

	Treatment	Control
Sample size ¹	873	352
Infection	0.32	0.54
School attendance	0.808	0.684

Example: testing one mean

- Test $H_0 : E[\text{infected}|\text{control}] = 0.5$ against $H_a : E[\text{infected}|\text{control}] \neq 0.5$
 - Let $I_i = 1$ if infected, else 0
 - What is sample variance of I_i ?

$$\begin{aligned}\hat{\sigma}_I^2 &= \frac{1}{n} \sum_{i=1}^n (I_i - \bar{I})^2 = \left(\frac{1}{n} \sum_{i=1}^n I_i^2 \right) - \bar{I}^2 \\ &= \left(\frac{1}{n} \sum_{i=1}^n I_i \right) - \bar{I}^2 \quad (I_i^2 = I_i) \\ &= \bar{I} - \bar{I}^2 = \bar{I}(1 - \bar{I})\end{aligned}$$

- Test statistic

$$z = \sqrt{n} \frac{\bar{I} - 0.5}{\sqrt{\bar{I}(1 - \bar{I})}} = \sqrt{352} \frac{0.54 - 0.5}{\sqrt{0.54(1 - 0.54)}} = 1.506$$

- P-value = $2(1 - \Phi(z)) = 0.132$
-
-

Example: testing two means

- Often want to test difference between two means
 - H_0 : treatment has no effect $\sim H_0$ treatment mean = control mean
- Let $\mu_T =$ treatment mean, $\mu_C =$ control mean
- $H_0 : \mu_T = \mu_C$
- CLT implies

$$\sqrt{n_T}(\bar{I}_T - \mu_T) \xrightarrow{d} N(0, \sigma_T^2) \text{ and } \sqrt{n_C}(\bar{I}_C - \mu_C) \xrightarrow{d} N(0, \sigma_C^2)$$

i.e. \bar{I}_T is approximately $N(\mu_T, \sigma_T^2/n_T)$

- Sum of two normals is normal
- If treatment and control groups independent, then $\bar{I}_T - \bar{I}_C$ is approximately $N(0, \sigma_T^2/n_T + \sigma_C^2/n_C)$ formally we could show

$$\frac{\bar{I}_T - \bar{I}_C}{\sqrt{\hat{\sigma}_T^2/n_T + \hat{\sigma}_C^2/n_C}} \xrightarrow{d} N(0, 1)$$

Example: testing two means

- Test $H_0 : E[\text{infected}|\text{control}] = E[\text{infected}|\text{treatment}]$ against $H_a : E[\text{infected}|\text{control}] > E[\text{infected}|\text{treatment}]$

– Test statistic:

$$\begin{aligned} z &= \frac{\bar{l}_T - \bar{l}_C}{\sqrt{\hat{\sigma}_T^2/n_T + \hat{\sigma}_C^2/n_C}} \\ &= \frac{0.32 - 0.54}{\sqrt{0.32(1 - 0.32)/873 + 0.54 * (1 - 0.54)/352}} = -7.12 \end{aligned}$$

– P-value = $\Phi(z) = 5 \times 10^{-13}$

- Test $H_0 : E[\text{school}|\text{control}] = E[\text{school}|\text{treatment}]$ against $H_a : E[\text{school}|\text{control}] < E[\text{school}|\text{treatment}]$

– Test statistic:

$$\begin{aligned} z &= \frac{\bar{S}_T - \bar{S}_C}{\sqrt{\hat{\sigma}_T^2/n_T + \hat{\sigma}_C^2/n_C}} \\ &= \frac{0.808 - 0.684}{\sqrt{0.808(1 - 0.808)/873 + 0.684 * (1 - 0.684)/352}} = 4.41 \end{aligned}$$

– P-value = $1 - \Phi(z) = 5 \times 10^{-6}$

The RAND health insurance experiment randomly assigned 3985 Americans to one of 14 different insurance plans from 1974-1982. All insurance plans included catastrophic coverage – families total health-care spending beyond a proportion of their income was completely covered. The least generous plan in the experiment only had catastrophic coverage. The most generous plan offered comprehensive care for free. There were also plans that had a deductible of \$150 per person or \$450 per family, after which everything was covered. Finally, there were plans with coinsurance that required families to pay between 25% and 50% of the cost of health care out of pocket, up to the catastrophic limit. You can read more about the RAND health insurance experiment in chapter 1 of [Angrist and Pischke \(2014\)](#) or in [Aron-Dine, Einav, and Finkelstein \(2013\)](#).

Example: RAND health insurance experiment

	Means		Differences between plan groups		
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
A. Health-care use					
Face-to-face visits	2.78 [5.50]	.19 (.25)	.48 (.24)	1.66 (.25)	.90 (.20)
Outpatient expenses	248 [488]	42 (21)	60 (21)	169 (20)	101 (17)
Hospital admissions	.099 [.379]	.016 (.011)	.002 (.011)	.029 (.010)	.017 (.009)
Inpatient expenses	388 [2,308]	72 (69)	93 (73)	116 (60)	97 (53)
Total expenses	636 [2,535]	114 (79)	152 (85)	285 (72)	198 (63)

Example: RAND health insurance experiment

	Means		Differences between plan groups		
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
B. Health outcomes					
General health index	68.5 [15.9]	–.87 (.96)	.61 (.90)	–.78 (.87)	–.36 (.77)
Cholesterol (mg/dl)	203 [42]	.69 (2.57)	–2.31 (2.47)	–1.83 (2.39)	–1.32 (2.08)
Systolic blood pressure (mm Hg)	122 [19]	1.17 (1.06)	–1.39 (.99)	–.52 (.93)	–.36 (.85)
Mental health index	75.5 [14.8]	.45 (.91)	1.07 (.87)	.43 (.83)	.64 (.75)
Number enrolled	759	881	1,022	1,295	3,198

Example: TRUMP health insurance experiment?



"It says our health insurance is being replaced by a series of tweets calling us losers."

p-value pitfalls

- Hypothesis tests and p-values are a tool for quantifying uncertainty, but not the only tool
- Significance thresholds often over-emphasized
- Difference in significance is not necessarily a significant difference
 - A statistically significantly different from 0 and B not statistically significantly different from 0 does not imply that $A - B$ is statistically significant
- When testing many hypotheses, likely to find significant results by chance alone
 - Researcher degrees of freedom and garden of forking paths can introduce many tests inadvertently

These ideas are frequently discussed by the statistician Andrew Gelman on his blog.
http://andrewgelman.com/2005/06/14/the_difference/
<http://andrewgelman.com/2011/09/09/the-difference-between-significant-and-not-significant/>
<http://andrewgelman.com/2016/01/04/plausibility-vs-probability-prior-distributions-garden-fork/>
<http://andrewgelman.com/2013/11/06/marginally-significant/>

¹New Yorker daily cartoon, January 5th 2017, <http://www.newyorker.com/cartoons/daily-cartoon/thursday-january-5th-health-care?intcid=mod-latest>

4.2 Asymptotic interpretation of p-values

When we calculate p-values using the asymptotic distribution of an estimator, then in a given sample, we know that our p-value is not exactly correct. However, we also know that for a sufficiently large sample, the error in the p-value will be arbitrarily small. Exactly how large of a sample we need for the error to be negligible depends on the true distribution of our sample and the number of parameters we estimated to calculate the p-value.

Asymptotic interpretation of p-values

- CLT: under H_0 ,

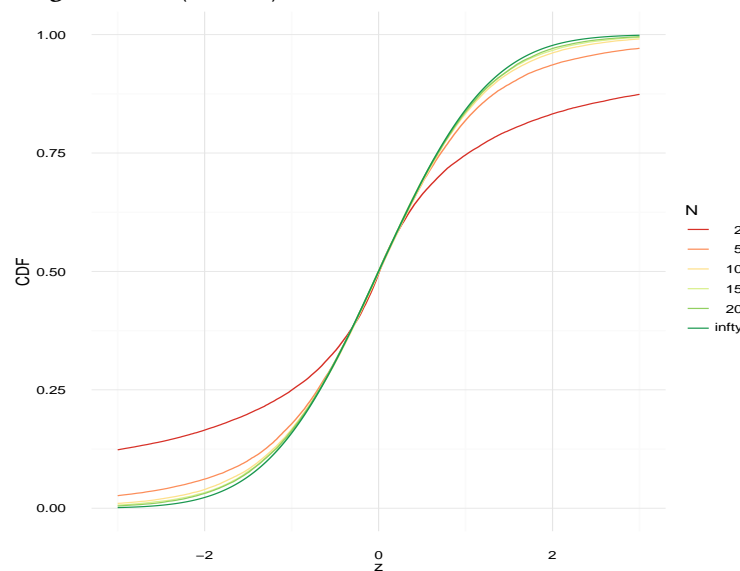
$$z_n = \sqrt{n} \frac{\bar{y} - \mu}{\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

or equivalently,

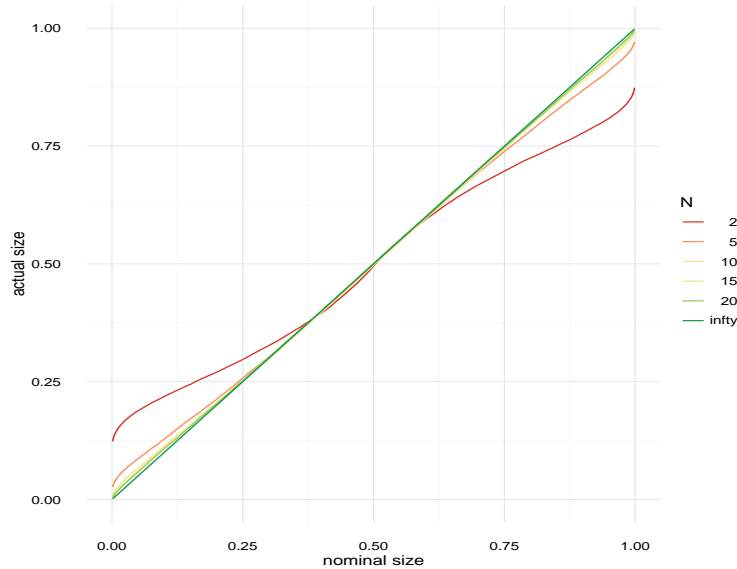
$$\lim_{n \rightarrow \infty} P(z_n \leq x) = \Phi(x)$$

- P-values are only correct asymptotically (for large samples)
 - Small finite sample p-values will be somewhat off
 - With exact p-values, under H_0 , p-values from distributed $U[0, 1]$
 - Asymptotic p-values, under H_0 , p-values only $U[0, 1]$ in large samples
-
-

Convergence of $P(z_n \leq x)$



Convergence of p-values



[Code for graphs](#)

References

References

- Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives* 27 (1):197–222. URL <http://www.aeaweb.org/articles.php?doi=10.1257/jep.27.1.197>.
- Diez, David M, Christopher D Barr, and Mine Cetinkaya-Rundel. 2012. *OpenIntro Statistics*. OpenIntro. URL <http://www.openintro.org/stat/textbook.php>.
- Linton, Oliver B. 2017. *Probability, Statistics and Econometrics*. Academic Press. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001868500&title=Probability%2C%20statistics%20and%20econometrics.
- Menzel, Konrad. 2009. "14.30 Introduction to Statistical Methods in Economics." Massachusetts Institute of Technology: MIT OpenCourseWare. URL <http://ocw.mit.edu/courses/economics/14-30-introduction-to-statistical-methods-in-economics-spring-2009/lecture-notes/>.

Miguel, E. and M. Kremer. 2003. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72 (1):159–217. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2004.00481.x/abstract>.

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.
