Statistics and Inference

Paul Schrimpf

UBC
Economics 326

January 23, 2018

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

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

Section 1

Properties of estimators

Unbiasedness

- Setup:
  - Sample $\{y_1, ..., y_n\}$
  - Parameter θ
  - Estimator W, some function of sample
- The bias of W is $\text{Bias}(W) = \mathrm{E}[W] - \theta$
- W is unbiased if $\mathrm{E}[W] = \theta$
- 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

Variance of estimators

- The variance of W is $\text{Var}(W)$
- The standard error of W is $\sqrt{\text{Var}(W)}$
- If W_1 and W_2 is 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 $\text{MSE}(W) = \text{E}[(W - \theta)^2]$
  - $\text{MSE}(W) = \text{Bias}(W)^2 + \text{Var}(W)$
- 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:

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

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

Section 2

Finite sample inference

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 $f(x_1, \ldots, x_n) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{N-\sum x_i}$
  - Could use to compute how surprising sample $\bar{x}$ that we observe is if the true $p = p_0$, e.g. compute $\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - p_0\right| \geq |\bar{x} - p_0|\right)$
- 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

Section 3

Asymptotics

Asymptotic inference

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

Convergence in probability

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

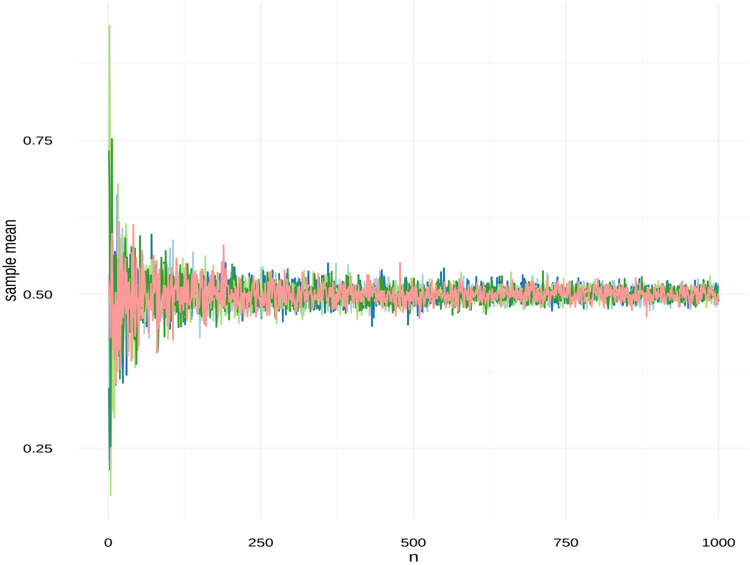
$$\lim_{n\to\infty} \mathrm{P}\left(|W_n - \theta| > \epsilon\right) = 0$$

  denote by $\operatorname{plim} W_n = \theta$ or $W_n \xrightarrow{p} \theta$
- Show using law of large numbers: if $y_1, ..., y_n$ are i.i.d. with mean μ, or if $y_1, ..., y_n$ have finite expectations and $\sum_{i=1}^{\infty} \frac{\mathrm{Var}(y_i)}{i^2}$ is finite, then $\bar{y} \xrightarrow{p} \mu$
- Properties:
  - $\operatorname{plim} g(W_n) = g(\operatorname{plim} W_n)$ 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 $W_n \xrightarrow{p} \theta$

Demonstration of LLN

```r
n <- seq(1,200) ## sample sizes
sims <- 10 ## number of simulations for each sample size
rand <- function(N) { return(runif(N)) }
## simulation
sampleMean <- matrix(0,nrow=length(n),ncol=sims)
for(i in 1:length(n)) {
  N = n[i];
  x <- matrix(rand(N*sims),nrow=N, ncol=sims)
  sampleMean[i,] = colMeans(x)
}
## plot sample means
library(ggplot2)
library(reshape)
df <- data.frame(sampleMean)
df$n <- n
df <- melt(df,id="n")
llnPlot <- ggplot(data=df,aes(x=n, y=value, colour=varia
  geom_line() + scale_colour_brewer(palette="Paired") +
  scale_y_continuous(name="sample mean") +
  guides(colour=FALSE)
llnPlot
```

Demonstration of LLN



Convergence in probability: examples

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

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^n y_i^2}_{\xrightarrow{p} \mathrm{E}[y^2] \text{ by LLN}} - \underbrace{\bar{y}^2}_{\xrightarrow{p} \mathrm{E}[y]^2 \text{ by LLN and CMT}}$$

$$\xrightarrow{p} \mathrm{E}[y^2] - \mathrm{E}[y]^2 = \mathrm{Var}(y)$$

Convergence in probability: examples

- Mean divided by variance:

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

Convergence in probability: examples

- Sample correlation:

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

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

Convergence in distribution

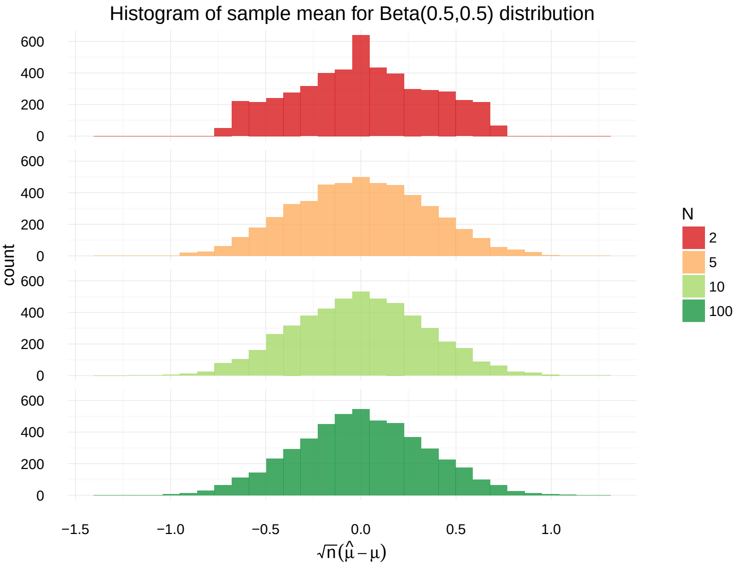
- Let F_n be the CDF of W_n and W be a random variable with CDF F
- W_n converges in distribution to W, written $W_n \xrightarrow{d} W$, if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x where F is continuous
- Central limit theorem: Let $\{y_1, ..., y_n\}$ be i.i.d. with mean μ and variance σ^2 then $Z_n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

Demonstration of CLT

```
N <- c(2,5,10,100)
simulations <- 5000
means <- matrix(0,nrow=simulations,ncol=length(N))
for(i in 1:length(N)) {
  n <- N[i]
  dat <- matrix(rbeta(n*simulations,
                      shape1=0.5,shape2=0.5),
                nrow=simulations, ncol=n)
  means[,i] <- (apply(dat, 1, mean) - 0.5)*sqrt(n)
}

# Plotting
df <- data.frame(means)
df$n <- N
df <- melt(df)
cltPlot <- ggplot(data=df, aes(x=value, fill=variable)) +
  geom_histogram(alpha=0.8, position="identity") +
  scale_x_continuous(name=expression(sqrt(n)(hat(mu) - mu))
  scale_fill_brewer(type="div",palette="RdYlGn",
                    name="N",label=N) +
  labs(title="Histogram of sample mean for Beta(0.5,0.5) di
  facet_grid(variable ~.,label=function(x) {""}) + theme_mi
```

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

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

CLT examples

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

Section 4

Hypothesis testing

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 $\sqrt{n}\frac{\bar{y}-\mu}{\hat{\sigma}^2} \sim t(n-1)$

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

$$p = \mathrm{P}\left(\left| \frac{\frac{1}{n}\sum_{i=1}^n y_i - \mu_0}{s_y} \right| \geq \left| \frac{\bar{y}-\mu_0}{\hat{\sigma}} \right| \right) = 2 \underbrace{F_t\left(-\left| \frac{\bar{y}-\mu_0}{\hat{\sigma}} \right| ; n-1 \right)}_{\text{CDF of t distribution}}$$

- 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,

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

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

[1] Only the sample size for observations of infections. The real data has more observations of school participation, but for illustration we will pretend this is the sample size.

Example: testing one mean

- Test $H_0 : \mathrm{E}[\text{infected}|\text{control}] = 0.5$ againt $H_a : \mathrm{E}[\text{infected}|\text{control}] \neq 0.5$
  - 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[infected|control] = E[infected|treatment] against H_a : E[infected|control] > E[infected|treatment]
  - Test statistic:

$$z = \frac{\bar{I}_T - \bar{I}_C}{\sqrt{\hat{\sigma}_T^2/n_T + \hat{\sigma}_C^2/n_C}}$$

$$= \frac{0.32 - 0.54}{\sqrt{0.32(1-0.32)/873 + 0.54 * (1-0.54)/352}} = -7.12$$

  - P-value $= \Phi(z) = 5 \times 10^{-13}$
- Test H_0 : E[school|control] = E[school|treatment] against H_a : E[school|control] < E[school|treatment]
  - Test statistic:

$$z = \frac{\bar{S}_T - \bar{S}_C}{\sqrt{\hat{\sigma}_T^2/n_T + \hat{\sigma}_C^2/n_C}}$$

$$= \frac{0.808 - 0.684}{\sqrt{0.808(1-0.808)/873 + 0.684 * (1-0.684)/352}} = 4.41$$

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

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	.19	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(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	−.87	.61	−.78	−.36
	[15.9]	(.96)	(.90)	(.87)	(.77)
Cholesterol (mg/dl)	203	.69	−2.31	−1.83	−1.32
	[42]	(2.57)	(2.47)	(2.39)	(2.08)
Systolic blood pressure (mm Hg)	122	1.17	−1.39	−.52	−.36
	[19]	(1.06)	(.99)	(.93)	(.85)
Mental health index	75.5	.45	1.07	.43	.64
	[14.8]	(.91)	(.87)	(.83)	(.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."

[1] *New Yorker* daily cartoon, January 5th 2017, http://www.newyorker.com/cartoons/daily-cartoon/

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

Asymptotic interpretation of p-values

- CLT: under H_0,

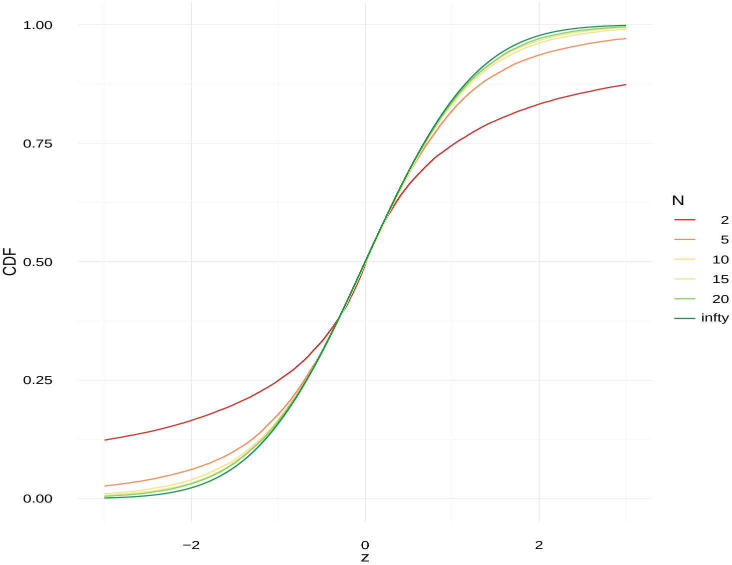
$$z_n = \sqrt{n}\frac{\bar{y} - \mu}{\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

or equivalently,

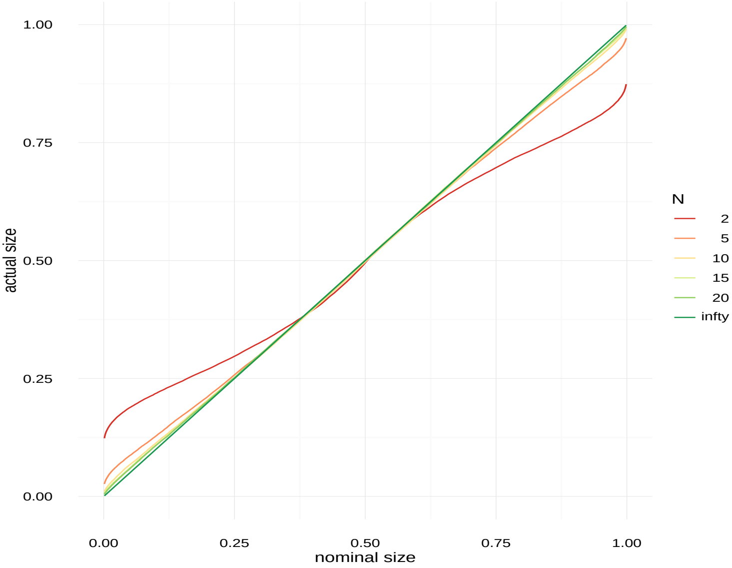
$$\lim_{n \to \infty} \mathrm{P}(z_n \leq x) = \Phi(x)$$

- 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 $\mathbb{P}(z_n \leq x)$



Convergence of p-values



Code for graphs

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

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.

References

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-econom 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.