Coverage of 1-sample proportion bootstrap methods

Connor Demorest

Background

- Intro Stats courses traditionally teach theory-methods
  - Ex: 1 and 2 sample t-test, Normal approx. to binomial, etc
- Bootstrap methods have some pedagogical value
  - ”Montana State Introductory Statistics with R” – By Stacey!
  - Tim Hesterberg – bootstrap resampling as a teaching tool
- Not a ton of literature on the bootstrap binomial approximation
  - Tons of literature about the t-tests and different ways to do resampling

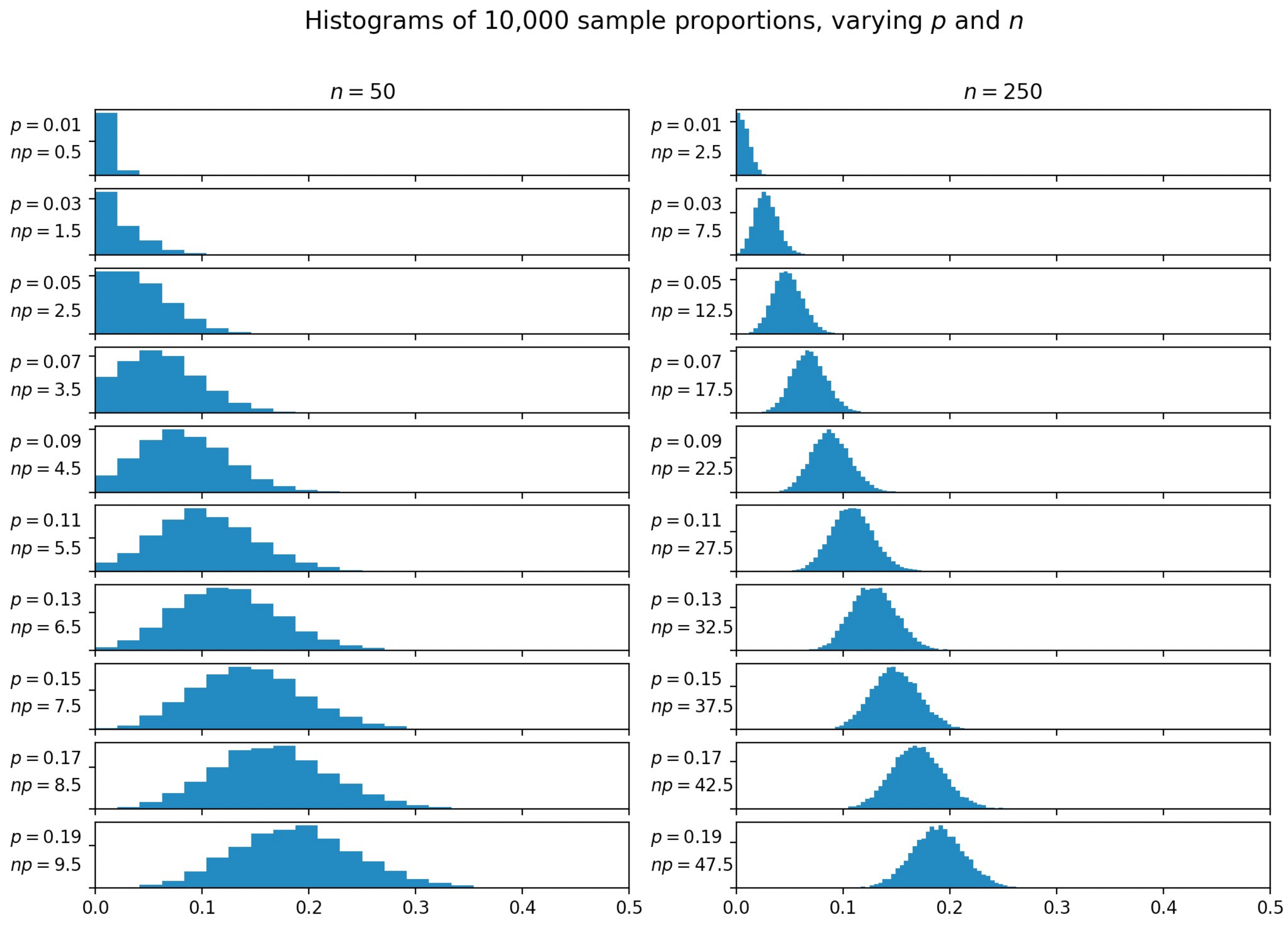
What is bootstrap resampling?

- Different types: we teach the "percentile interval" method
  - Must have been a pedagogical decision because they have the worst theoretical power of all the common bootstrap methods
    - Does power of a test really matter in intro stats? Discuss
- "Percentile interval" bootstrapping
  - Take sample proportion (p) and sample size (n) and generate many new simulations of the sample to get an estimated sampling distribution
  - Take $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ percentiles of the estimated sampling distribution to create a $100(1-\alpha)\%$ confidence interval

Review of theory-based methods

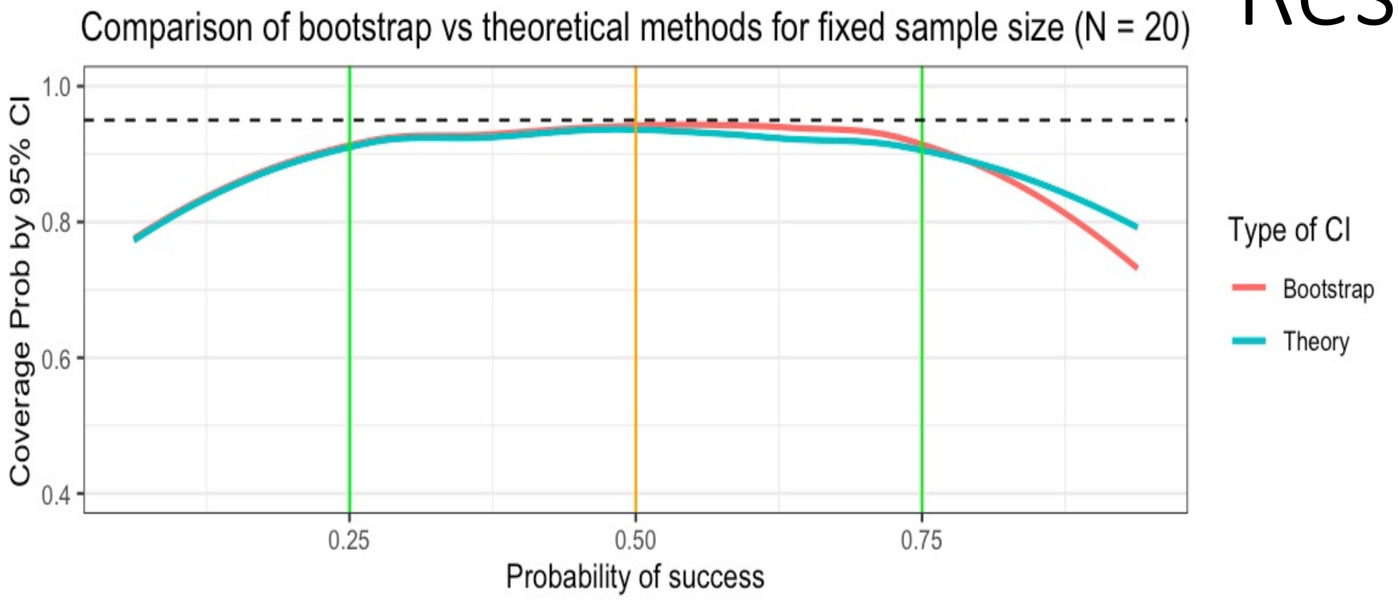
- With "large" n, we can approximate the binomial distribution as a normal distribution and use the normal distribution quantiles
- To create a 95% CI using theoretical methods:
  - $\hat{p} \pm 1.96 * sqrt(\frac{\hat{p}*(1-\hat{p})}{n})$ where $\hat{p}$ is the sample proportion
- When should this approximation be reasonable?
  - Rule of thumb: if n*p > 5 (or 10) and n*(1-p) > 5 (or 10)
  - How reasonable is this rule for the theory and bootstrap methods?

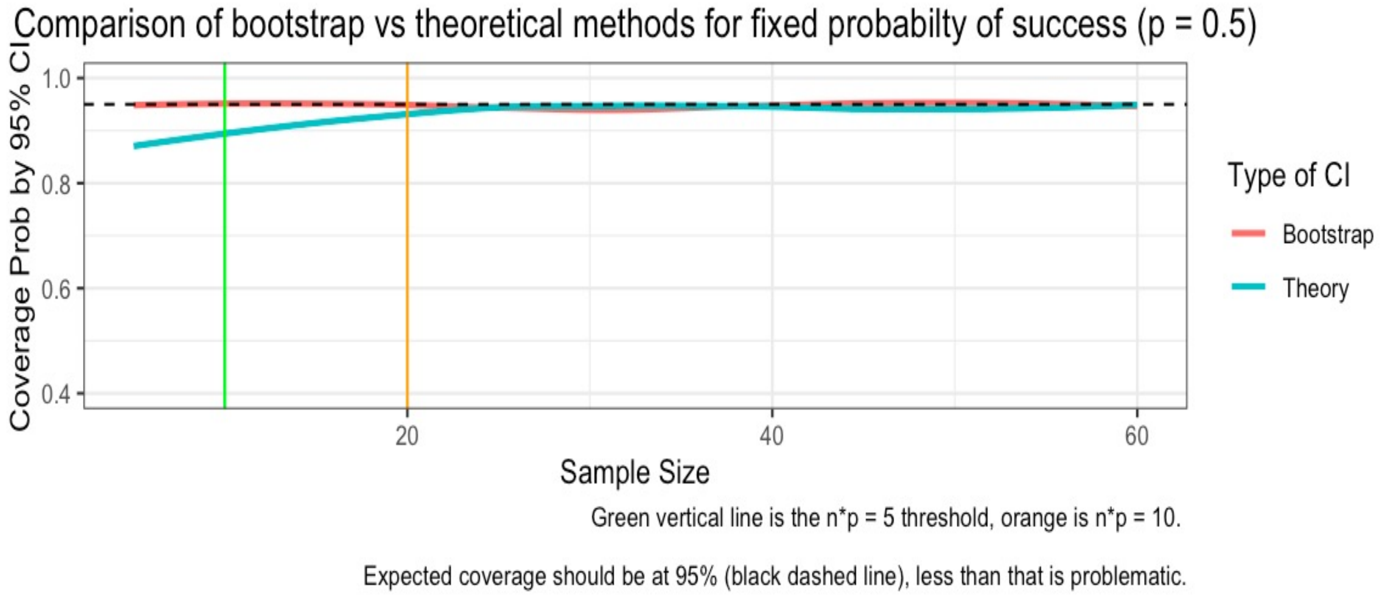
Sampling distributions of values of n*p



- When n*p = 5, there is some “bunching” at 0
- N*p > 10 seems to be too strong of an assumption based on this plot
- I suggest n*p > 7.5 could be a good middle ground

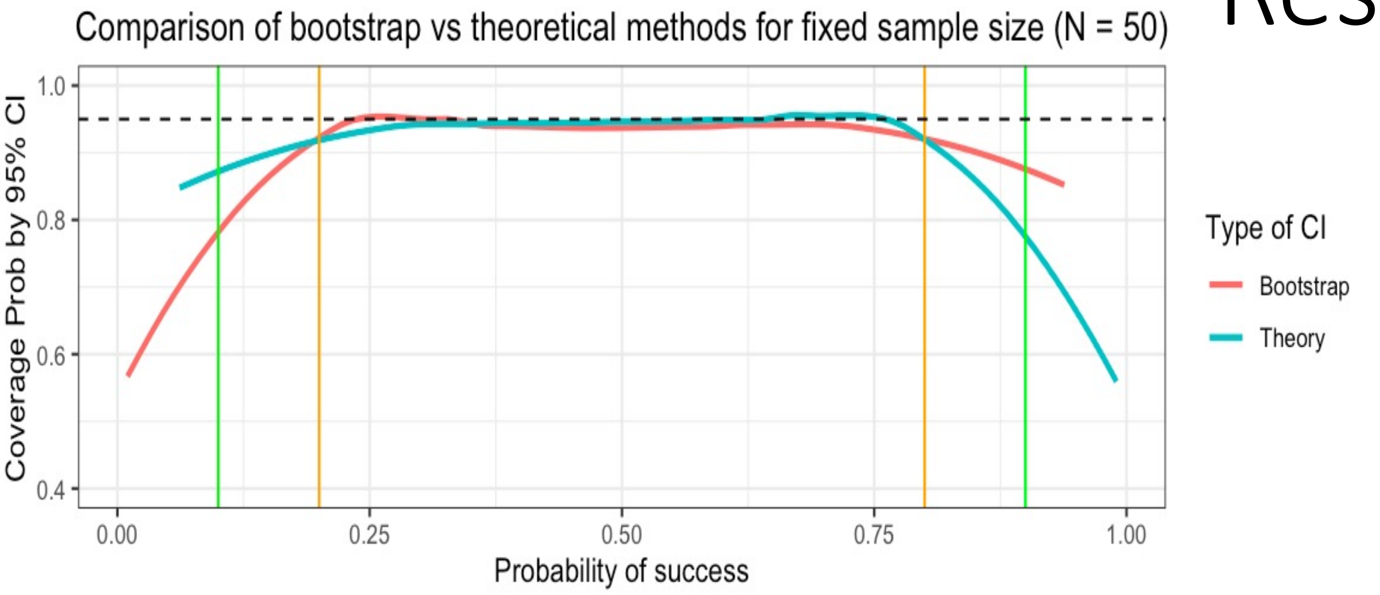
Results: n = 20, p = 0.5

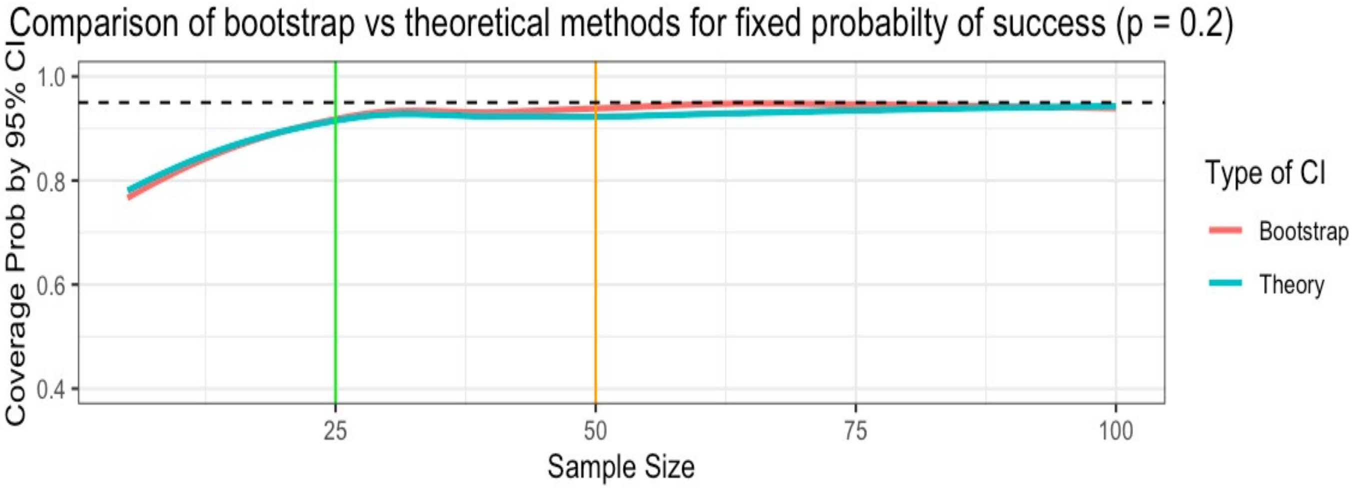




- For a fixed n = 20, bootstrap methods and theory methods are very similar
- For a fixed p = 0.5, bootstrap is slightly better for very small sample sizes
- Below green (n*p = 5), there is weak coverage around 80-90%
- At n*p = 10 (orange), almost perfect coverage for both methods

Results: n = 50, p = 0.2





Green vertical line is the n*p = 5 threshold, orange is n*p = 10.

Expected coverage should be at 95% (black dashed line), less than that is problematic.

- Bootstrap and theory seem very similar here
- Between green and orange there is strong coverage at least 90% for both methods
- Might support the n*p > 10 is too strong theory
- See R shiny app for more (if time) ☺
- I’m not sure what next steps would make sense. T-tests have been done to death for this sort of thing