MATHEMATICS E-156, SPRING 2014
MATHEMATICAL FOUNDATIONS OF STATISTICAL SOFTWARE
Module #11 (Confidence intervals)

Last modified: April 23, 2014

Reading from Chihara and Hesterberg

- Chapter 7, starting with section 7.1.3. There is no harm in rereading sections 7.1.1 and 7.1.2.

R scripts

- 11P-ConfidenceUniform.R
  Topic 1 – a location parameter for a uniform distribution
  Topic 2 – a scale parameter for a uniform distribution

- 11A-DifferenceMeans.R
  Topic 1 – confidence interval for the difference of means
  Topic 2 – using Student t to compute a P value
- ConfidenceOneSided.R
  Topic 1 – computing a one-sided confidence interval from sample mean and sd
  Topic 2 – simulation, using the mean mu as a scale parameter
  Topic 3 – computing a one-sided confidence interval from real data
- 11C-Confidence Proportion.R
  Topic 1 – using a normal approximation to the binomial distribution

- 11D-Bootstrap t.R
  Topic 1 – bootstrapping the Student t statistic
  Topic 2 - checking the bootstrap t confidence interval by simulation

Mathematical notes

1. Pivotal statistic – location parameter

   Suppose that $X_1, X_2, \cdots X_n$ are random variables from a distribution with parameter θ. A *pivotal statistic* is a function $h(X_1, X_2, \cdots X_n, \theta)$ whose distribution does not depend on θ or on any parameters with unknown values. The parameter θ is called a *location parameter* if the distribution of $X - \theta$ does not depend on θ.

   (a) Show that for a normal distribution with known σ, μ is a location parameter and that given one sample X from this distribution, there are random variables L and U, with values that depend on X but not on θ, such that $P(\mu < L) = 0.025$ and $P(\mu > U) = 0.025$

   (b) Show that if $X \sim \text{Unif}(\theta - 1, \theta + 1)$, then θ is a location parameter, and explain how to determine L and R for a 92% confidence interval.

2. Pivotal statistic – scale parameter

   A parameter θ is called a *scale parameter* if the distribution of X/θ does not depend on θ.

   (a) Show that if X has the distribution $N(2\sigma, \sigma^2)$, then σ is a scale parameter, and find a formula for L if you want a 97.5% one-sided confidence interval for σ.

   (b) Show that if X has the distribution $\mathrm{Unif}(0, \theta)$, then θ is a scale parameter. Find formulas for L and U that can be used if you know only $X = \max(X_1, X_2)$ and want a 90% confidence interval.

3. Confidence interval for a proportion

   Suppose that you have one sample from a binomial distribution, $X \sim \text{Binom}(n, p)$. By using a normal approximation to the binomial distribution, find formulas for random variables L and U that will function as the ends of a 95% confidence interval; i.e. $P(L > p) = 0.025, P(U < p) = 0.025$. Let $q = 1.96$ denote the .975 quantile of $N(0, 1)$, and set $X/n = \hat{p}$. (Messy algebra – p is neither a location parameter nor a scale parameter.)

Section problems

1. Exercise 11 on page 202 - answer on page 403.

2. Exercise 23 on page 205 - partial answers on page 403. After solving the problem, make a dataframe to simulate the test results. Do a permutation test to see whether the drug is effective, and construct bootstrap t confidence intervals to compare with your answers to (a), (b), and (d).

3. Exercise 34 on page 208. The quantity $2\lambda X$ is a "pivotal statistic," since it has a known distribution (chi square) that does not depend on lambda. Do a simulation where you draw a random X from Gamma$(2, 3)$4000 times, compute the lower bound L and upper bound U of the confidence interval in each case, and count how many times the true value of λ falls outside your confidence interval on either side.

Homework assignment This assignment should be submitted as a single R script. Include enough comments so that it is clear what you are doing and where each problem begins. You can upload it to the dropbox on the Class 11 page of the Web site.

It is OK to paste and edit lines from the scripts on the course Web site. It is not OK to paste lines from your classmates' solutions!

1. Exercise 12 on page 202. Also find a 95% t confidence interval, as in exercise 11 (the first section problem).

2. Exercise 20 on page 205. After solving the problem, make a dataframe to simulate the survey results. Do a permutation test to see whether there is gender difference in voter preference, and construct bootstrap t confidence intervals to compare with your answers to (a), (b), and (c).

   This problem is very similar to the second section problem!

3. Exercise 36 on page 208. This problem will encourage you to read section 7.5 carefully!

4. (a) Exercise 38 on page 208. You will need the result of exercise 37, which in turn relies on familiar theorem B.16. In this case $(n-1)S^2/\sigma^2$ is a pivotal statistic because its chi-square distribution does not depend either on σ or on the unknown parameter μ.

   (b) Do a simulation where the weights of the eight cereal boxes are drawn from $N(560, 10^2)$, and check that, when you calculate the confidence interval as in part (a), the true variance of 100 falls within the confidence interval in roughly 95% of the cases.

5. Suppose that you draw two samples from $N(\mu, \sigma)$ with $\mu = 5, \sigma = 2$. In this case the quantity $T = \sqrt{2}(\overline{X} - \mu)/S$ is a pivotal statistic, since its distribution (Student t with one degree of freedom, a.k.a. Cauchy) does not depend on the unknown parameter σ. You can therefore calculate from the sample mean $\overline{X}$ two quantities L and U such that the events $\mu < L$, $L \leq \mu \leq U$, and $U < \mu$ all have probability 1/3. Do a simulation with 3000 trials to show that this idea works. You can get the required quantiles either from `qcauchy()` or from `qt()`.