XII. The Bootstrap

A. Problem: want general procedure for constructing confidence intervals for a parameter with minimal assumptions about distribution

1. Solution:
   a. treat data as though they are the population of interest, so we don’t need true parameter value.
   b. draw new random samples from the population represented by the original sample
      i. do this with replacement, so it won’t be a permutation usually.
   c. Calculate estimator for this sample, and call it \( \hat{\theta}_{B,i} \).
   d. Estimate standard error as \( \hat{\sigma}^2 = \sum_{i=1}^{B} (\hat{\theta}_{B,i} - \theta)^2 / B \)

   ii. Using \( \hat{\theta}_b = \sum_{i=1}^{B} \hat{\theta}_{B,i} / B \) in place of \( \theta \) gives \\
      iii. Use \( B - 1 \) instead of \( B \) in denominator of \( \hat{\sigma}^2 \), or \( \hat{\theta}_b \) in place of \( \theta \).

B. Residual method

1. Suppose a parameter \( \theta \) is estimated by \( \hat{\theta} \).
2. Suppose that distribution of \( \hat{\theta} - \theta \) does not depend on \( \theta \).
3. Let \( v_L \) and \( v_C \) satisfy \( P[v_L \leq \hat{\theta} - \theta \leq v_U] = 1 - \alpha \).
4. Then \( P[\hat{\theta} - v_U \leq \theta \leq \hat{\theta} - v_L] = 1 - \alpha \).
5. Estimate \( v_U \) and \( v_L \) be quantiles of \( \hat{\theta}_{B,i} - \hat{\theta} \).

C. Percentile method.

1. Referred to as the “usual method” by Shao and Tu (1995, p. 132f).
2. Suppose \( \theta - \hat{\theta} \) has distribution symmetric about 0.
3. Then can use \( v_U = \hat{\theta} - u_L \), \( v_L = \hat{\theta} - u_U \).
4. CI is now \((u_L, u_U)\)
5. Problems:
   a. In many cases, we are not prepared to make assumptions about the distribution of the observations.
   b. If we knew the true value of the parameter, we would not have so do statistics.

D. Which order statistic from the bootstrap sample do we use?

1. Most naive approach uses \( \hat{\theta}_{B(i)} \) to represent quantile \( i/N \) of this distribution.
   a. \( \hat{\theta}_{B(1)} \) represents \( 1/N \) quantile
   b. \( \hat{\theta}_{B(N)} \) represents \( 1 \) quantile
2. Conceptually, the estimation problem ought to be symmetric if we swap the order of bootstrap observations
   a. Naive quantiles are not.
   b. Upper one is wrong: Either for residual or for percentile bootstrap, population that \( \hat{\theta}_B \) might be intended to
      represent might not exist on a bounded interval.
3. Make symmetric by making \( \hat{\theta}_{B(i)} \) to represent quantile \( i/(N + 1) \) of this distribution.

E. BCa method:

1. Suppose that we want to give a CI for \( \theta \), with estimator \( \hat{\theta} \).
2. Suppose \( \hat{\theta} \) can be transformed to normality using a transformation \( \phi \) (which is unknown, but we’ll be able to
   estimate it.)
3. Suppose \( \phi(\hat{\theta}) \) has a bias that depends linearly on \( \phi(\theta) \):
   a. More realistically, we’ll let \( u_L \) and \( u_U \) be quantiles of \( \hat{\theta}_{B,i} \), with \( v_L = u_L - \hat{\theta} \), \( v_U = u_U - \hat{\theta} \).
6. CI is \((\hat{\theta} - v_U, \hat{\theta} - v_L) = (2\hat{\theta} - u_U, 2\hat{\theta} - u_L)\)
7. Davison and Hinkley (1997), p. 29 refer to this as the basic bootstrap confidence interval method.

F. Bootstrap- \( t \) confidence interval:

1. Setup:
   a. \( \hat{\theta} \) has distribution centered around \( \theta \),
   b. but with dispersion \( \sigma(\theta) \) depending on \( \theta \)
   c. You have an estimate of \( \sigma \) of \( \sigma(\theta) \)
   d. Ratio is called studentized.
2. Resample \( B \) data sets as above.
3. for each resampled data set, calculate
   a. estimate \( \hat{\theta}_{B,i} \)
   b. Then can use \( v_U = \hat{\theta} - u_L \), \( v_L = \hat{\theta} - u_U \).
    c. CI is now \((u_L, u_U)\)
   d. Ratio is called studentized.
b. standard error $\hat{\sigma}_{B,i}$.

4. Let empirical distribution of $T_{B,i} = (\hat{\theta}_{B,i} - \hat{\theta})/\hat{\sigma}_{B,I}$ stand in for distribution of $T = (\hat{\theta} - \theta)/\sigma$
   a. This distribution doesn’t involve unknown parameters
      i. Such a quantity is called pivotal.
      ii. Typical strategy for inference.
   b. With $\bar{X}$ in for $\mu$.
   c. Then $0.95 \approx \Pr(t_L \leq (\hat{\theta} - \theta)/\hat{\sigma} \leq t_U)$, for $t_L$ and $t_U$ the .025 and .975 points of the distribution of $T$.
   d. Let $t_L$ and $t_U$ be percentage points from bootstrap distribution.

5. Often have $\theta = \mathbb{E}[X_j], \hat{\theta} = \bar{X}, \hat{\sigma} = S/\sqrt{n}$. [Mark B]

G. Can get scale parameter via log scale. [Mark C R]
H. Cautionary example: Coefficient of variation with mean near zero. [Mark D R]