WS #10 - Bootstrap t

Monday, October 6, 2025

Math 154 - Jo Hardin

Your Name: ____

Names of people you worked with: ____

Name the people sitting one table over from you. Tell your partner one fantastic thing from your weekend.

Task:

Put in the back of your head the distribution of: $\frac{\overline{X}-\mu}{s/\sqrt{n}}$ (which, incidentally, we know is distributed according to t_{n-1} if $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$.)

Additionally, let

$$\begin{aligned}
\hat{\theta}^*(b) &= \text{estimate of } \theta \text{ from the } b^{th} \text{ resample} \\
\hat{SE}^* &= \left[\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^*(b) - \hat{\theta}^*)^2\right]^{1/2}
\end{aligned}$$

1. If you sample B times from a population, how many copies of $\overline{X}$ will there be? How many copies of $s/\sqrt{n}$ will there be?
2. If you re-sample B times from a single dataset, how many copies of $\hat{\theta}^*(b)$ will there be? How many copies of $\hat{SE}^*$?
3. Gosset realized that s varies from sample to sample. In bootstrapping, we want to mimic the process of sampling from a population. What is the problem with using the bootstrap values given above to produce a bootstrapped test statistic?
4. To address the problem, suggest a way of estimating the SE of $\hat{\theta}$ separately for each b.

Solution:

1. When sampling from a population, there will be B copies each of $\overline{X}$ and $s/\sqrt{n}$.
2. When re-sampling from a dataset, there will be B copies of $\hat{\theta}^*(b)$ and 1 copy of $\hat{SE}^*$.
3. Somehow we need to create a test statistic where both the numerator and the denominator are random variables.
4. To find $\hat{SE}^*(b)$, we must bootstrap twice. The algorithm is as follows:

   a. Generate B_1 bootstrap samples (resamples from the original data), and for each sample $\underline{X}^{*b}$ compute the bootstrap estimate $\hat{\theta}^*(b)$.

   b. Take B_2 bootstrap samples (resamples from the bootstrapped data) from $\underline{X}^{*b}$, and estimate the standard error, $\hat{SE}^*(b)$.

   c. The resulting distribution will be based on B_1 values for $T^*(b) = \frac{\hat{\theta}^*(b)-\hat{\theta}}{\hat{SE}^*(b)}$.