WSABI Summer Research Lab **Summer 2024**

Lecture 9: The Bootstrap

Instructor: Ryan Brill *Scribe: Jonathan Pipping*

9.1 Asymptotic Confidence Intervals

To this point of the course, we have used the Central Limit Theorem's asymptotic normality to construct confidence intervals for the mean of a random variable. We'll begin by reviewing a few examples of asymptotic confidence intervals.

Example 9.1 (Wald Confidence Interval for the Binomial Parameter)**.** The $(1-\alpha)\cdot 100\%$ Wald Confidence Interval for the binomial parameter p is given by

$$\widehat{p} \pm z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \tag{9.1}$$

where n is the number of observations and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. *Recall that this was proven by the Central Limit Theorem, which states that the binomial sum is asymptotically normal.*

Example 9.2 (Agresti-Coull Confidence Interval for the Binomial Parameter)**.** The $(1-\alpha)\cdot 100\%$ Agresti-Coull Confidence Interval for the binomial parameter p is given by

$$\widehat{p} \pm z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{p}'(1-\widehat{p}')}{n+4}}, \text{ where } \widehat{p}' = \frac{S_n+2}{n+4} \tag{9.2}$$

where n is the number of observations and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. *Recall that this was proven by the Central Limit Theorem, which states that the binomial sum is asymptotically normal.*

Example 9.3 (Confidence Interval for Regression Coefficients)**.** The $(1-\alpha)\cdot 100\%$ confidence interval for the linear regression coefficient β_j is given by

$$\widehat{\beta}_j \pm t_{n-k,1-\alpha/2} \cdot \mathrm{SE}(\widehat{\beta}_j) \tag{9.3}$$

where n is the number of observations, k is the number of parameters in the model and $t_{n-k,1-\alpha/2}$ is the $1-\alpha/2$ quantile of the t-distribution with $n-k$ degrees of freedom. *Note that this carries the implicit assumption that the errors ϵ_i are normally distributed.*

All three of these examples assume normality, but what if we don't want to make this assumption? Or what if we want to construct a confidence interval that is too complicated to compute analytically? It's for this reason that we introduce the bootstrap.

9.2 The Bootstrap

Definition 9.4 (The Bootstrap)**.** *The bootstrap is a non-parametric method for obtaining standard errors or constructing confidence intervals without assuming a distribution. It accomplishes this by* ***resampling data*** *from the original sample.*

How can we use the bootstrap to construct confidence intervals? We'll begin by implementing the bootstrap for the examples from the previous section.

9.2.1 Bootstrap Estimated Binomial Proportion

Let T represent the full training dataset with observations $\{X_1, \ldots, X_n\}$. To bootstrap with B replications, for each $b = 1, \ldots, B$:

1. Generate $T^{(b)}$: Re-sample m observations X_i from T with replacement.
2. Compute $\widehat{p}^{(b)}$: Estimate $\widehat{p}$ from $T^{(b)}$ as $\widehat{p}^{(b)} = \frac{S_m^{(b)}}{m}$ where $S_m^{(b)} = \sum_{i=1}^{m} X_i^{(b)}$.
3. Sort the $\widehat{p}^{(b)}$s: Order the $\widehat{p}^{(b)}$s from smallest to largest so that $\widehat{p}^{(1)} \leq \widehat{p}^{(2)} \leq \ldots \leq \widehat{p}^{(B)}$.

Then we can easily calculate the following quantities of interest:

- $SE(\widehat{p}) = SD(\widehat{p}^{(1)}, \ldots, \widehat{p}^{(B)}) = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{p}^{(b)} - \bar{\widehat{p}}\right)^2}$
- 95% CI for p: $[\widehat{p}^{(2.5^{th} \text{ quantile})}, \widehat{p}^{(97.5^{th} \text{ quantile})}]$

9.2.2 Bootstrap Estimated Regression Coefficients

Let T represent the full training dataset with observations $\{(X_1, y_1), \ldots, (X_n, y_n)\}$. To bootstrap with B replications, for each $b = 1, \ldots, B$:

1. Generate $T^{(b)}$: Re-sample m observations (X_i, y_i) from T with replacement.
2. Compute $\widehat{\beta}^{(b)}$: Estimate $\widehat{\beta}$ from $T^{(b)}$ using the OLS solution $\widehat{\beta}^{(b)} = ({X^{(b)}}^T X^{(b)})^{-1} {X^{(b)}}^T y^{(b)}$ where $X^{(b)}$ is the design matrix and $y^{(b)}$ is the response vector for the b^{th} bootstrap sample.
3. Sort the $\widehat{\beta}_j^{(b)}$s: Order the $\widehat{\beta}_j^{(b)}$s from smallest to largest so that $\widehat{\beta}_j^{(1)} \leq \widehat{\beta}_j^{(2)} \leq \ldots \leq \widehat{\beta}_j^{(B)}$.

Then we can easily calculate the following quantities of interest:

- $SE(\widehat{\beta}_j) = SD(\widehat{\beta}_j^{(1)}, \ldots, \widehat{\beta}_j^{(B)}) = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{\beta}_j^{(b)} - \bar{\widehat{\beta}}_j\right)^2}$
- 95% CI for β_j: $[\widehat{\beta}_j^{(2.5^{th} \text{ quantile})}, \widehat{\beta}_j^{(97.5^{th} \text{ quantile})}]$

As we can see, once we re-sample the data and re-estimate the parameters of interest, we can easily calculate the standard error and confidence interval for the parameters with elementary statistics. We now turn to visualizing the bootstrap resampling scheme.

9.2.3 Visualizing the Bootstrap

The re-sampling scheme described above is illustrated in Figure 9.1: first we re-sample the data B times with replacement, estimate the parameters of interest, then sort the estimates and calculate the standard error and confidence interval for the parameter of interest.

Why does this resampling scheme work? Let's compare to how we would ideally obtain standard errors and confidence intervals (which is impossible in practice). First, we would need to sample the population B times with replacement, get B different parameter estimates, then calculate the standard error and confidence interval as before. We visualize this in Figure 9.2.

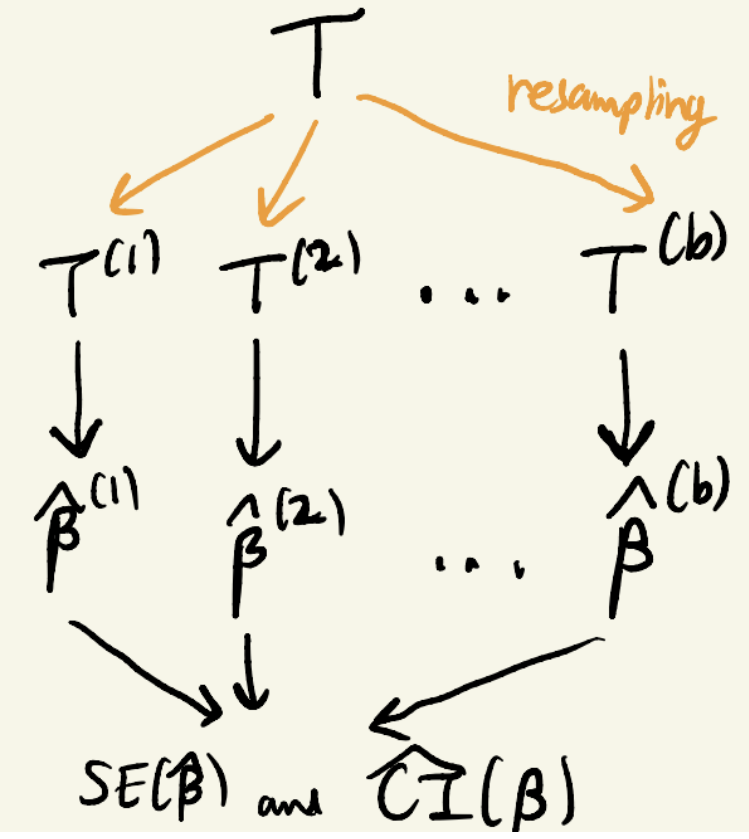

Figure 9.1: Visualization of the Bootstrap Resampling Scheme

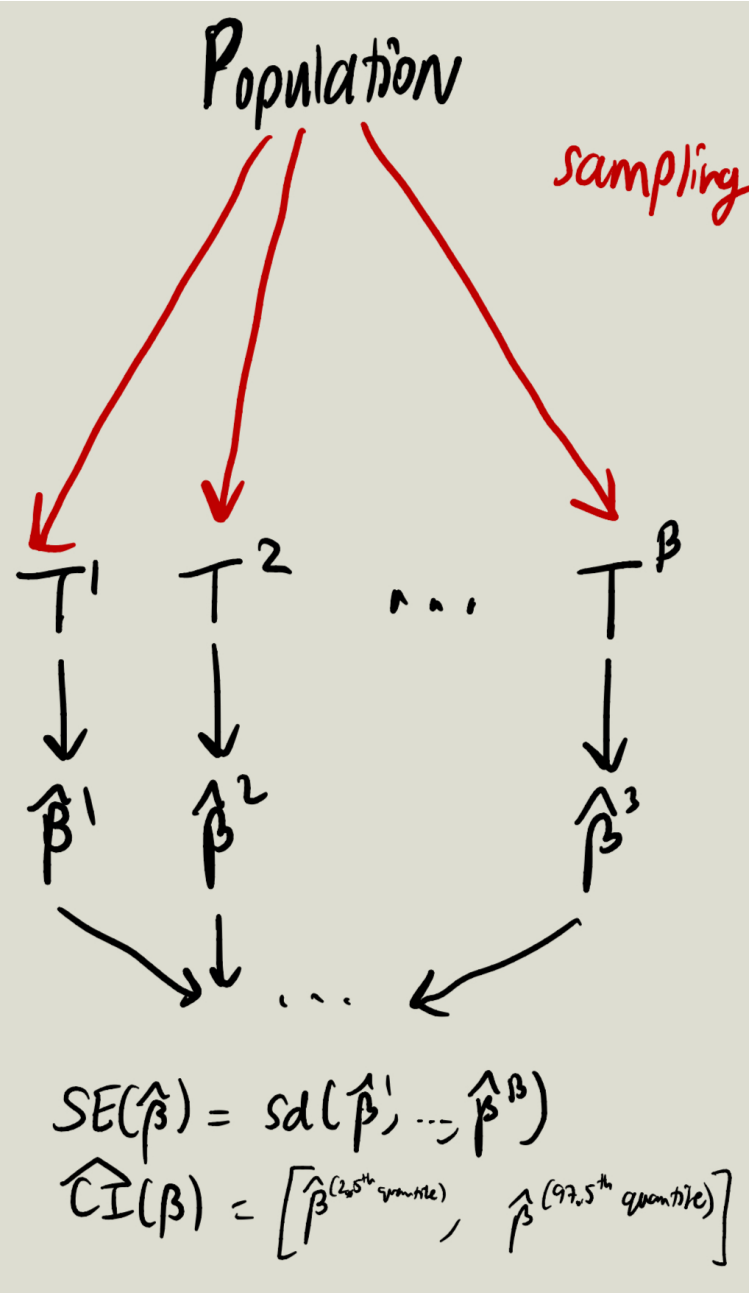

Figure 9.2: Visualization of the Population Resampling Scheme

Bootstrapping works because the training dataset T is itself a sample from the population; thus resampling from T minimics sampling from the population. These two approaches are pictured together in Figure 9.3.

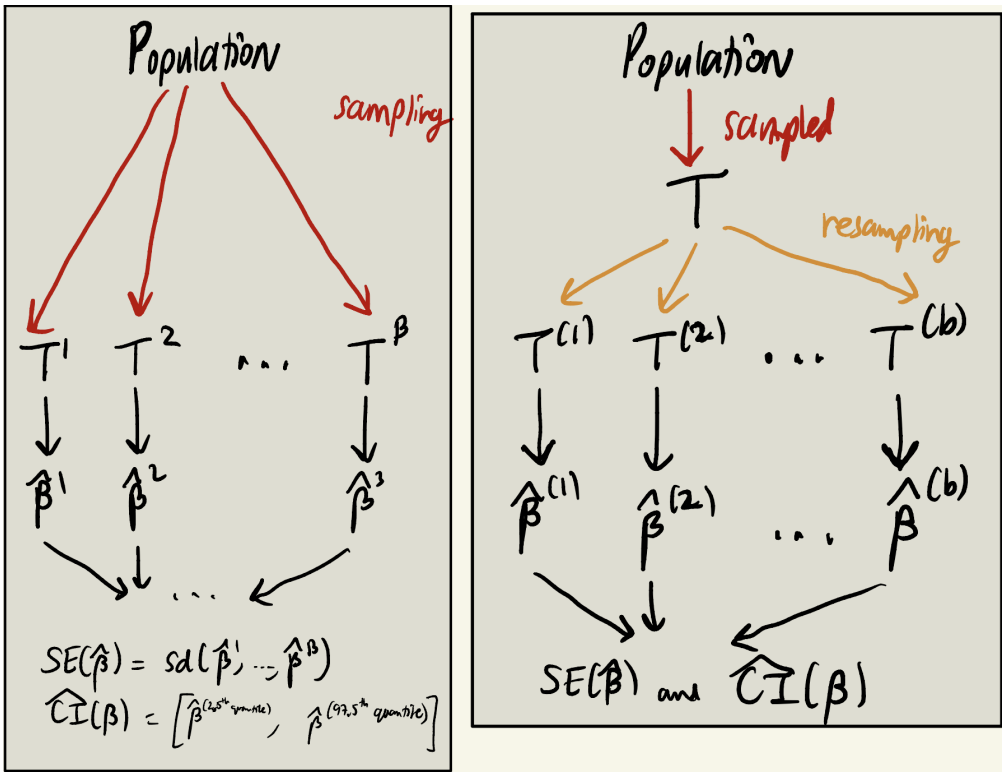

Figure 9.3: Comparison of the Population Sampling Scheme and the Bootstrap Resampling Scheme

9.2.4 Drawbacks

Like any other method, the bootstrap has its drawbacks. For one, it's not always applicable. For example, bootstrapping works well for the standard error of means, but not for extrema like the max and min. Additionally, the bootstrap is computationally intensive, and can sometimes underestimate uncertainty (meaning the SE is too small and the CI is too narrow). However, we can calibrate the bootstrap to achieve desired coverage probabilities. For more on this, see [BYW].

References

[AC] Agresti, A., & Coull, B.A., *Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions*, The American Statistician, 1998.

[BCD] Brown, T.C., Cai, T.T., & Dasgupta, A., *Interval Estimation for a Binomial Proportion*, Statistical Science, 2001.

[BYW] Brill, R. S., Yurko, R., & Wyner, A. J., *Analytics, Have Some Humility: A Statistical View of Fourth-Down Decision Making*, The American Statistician, 2025.