Bootstrap-t interval

- The Bootstrap-t interval allows us to obtain accurate intervals without having to make normal theory assumptions
- It estimates the distribution of the pivot directly from the data
- Assumes that we have a formula for the standard error of a statistic

Bootstrap-t interval procedure

1. Estimate θ by $\hat{\theta}$ and its standard error by $\widehat{\text{s.e.}}\left(\hat{\theta}\right)$
2. Generate B bootstrap samples
3. Compute $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ corresponding to each bootstrap sample.
4. Compute
$$t_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\widehat{\text{s.e.}}\left(\hat{\theta}_b^*\right)},\ b = 1, \ldots, B,$$
where $\widehat{\text{s.e.}}\left(\hat{\theta}_b^*\right)$ is computed using the same formula as $\widehat{\text{s.e.}}\left(\hat{\theta}\right)$ but using the bootstrap sample.
5. Compute the $\alpha/2$ and $(1-\alpha/2)$-quantiles of $t_1^*, t_2^*, \ldots, t_B^*$. Call the $\hat{t}_{\alpha/2}$ and $\hat{t}_{1-\alpha/2}$.
6. The $(1-\alpha)100\%$ bootstrap-t confidence interval for θ is
$$\left(\hat{\theta} - \hat{t}_{1-\alpha/2}\text{s.e.}\left(\hat{\theta}\right), \hat{\theta} - \hat{t}_{\alpha/2}\text{s.e.}\left(\hat{\theta}\right)\right)$$

Remarks about the bootstrap-t interval

- The quantity $\frac{\hat{\theta}-\theta}{\widehat{\text{s.e.}}(\hat{\theta})}$ is called an approximate pivot. That is, the distribution is approximately the same for each value of θ, which allows us to construct the confidence interval.
- The bootstrap-t quantile applies only to the given sample (unlike standard normal or t-quantiles)
- Theoretically, as n increases, the coverage of the bootstrap-t is closer to the desired nominal level (e.g., 95%) than when the standard normal or t-quantiles are ued
- Bootstrap-t intervals can be asymmetric about 0.
- Bootstrap-t procedure is particularly applicable to location-equivariant statistics (i.e., increasing each data value by a constant c results in increasing the statistic by c)
- Bootstrap-t cannot be used for more general problems (e.g., setting a CI for a correlation coefficient0
- Bootstrap-t cannot be recommended for general nonparametric problems

Bootstrap-t when SE formula is unavailable

- The denominator of t_b^* requires that $\widehat{\text{s.e.}}\left(\hat{\theta}_b^*\right)$, the estimated standard deviation of $\hat{\theta}_b^*$ for the bth bootstrap sample.
- But when $\hat{\theta}$ is a more complicated statistic, it may not have a simple standard error formula.
- To address this, we compute a bootstrap estimate of SE *for each bootstrap sample*. This implies two nested levels of bootstrap sampling!

- Number of bootstrap samples needed:

$$\text{Estimation of the percentiles of the pivot} \longrightarrow B = 1000$$
$$\text{Estimation of SE} \longrightarrow B = 25$$

- This requires $25 \cdot 1,000 = 25,000$ overall bootstrap samples. This might be computationally infeasible.

Transformations and the Bootstrap-t

- The Bootstrap-t interval may perform erratically in small-sample nonparametric settings
- This can be alleviated through proper transformation.
- Result: Let (X_i, Y_i), $i = 1, \ldots, n$ be iid with $E\left(X_i^4\right) < \infty$ and $E\left(Y_i^4\right) < \infty$, then

$$\sqrt{n}\left(\frac{1}{2}\log\frac{1+r}{1-r} - \frac{1}{2}\log\frac{1+\rho}{1-\rho}\right) \xrightarrow{d} N(0,1).$$

- Note: Fisher transformation $r \longmapsto \frac{1}{2}\log\frac{1+r}{1-r} = \tanh^{-1}(r)$

Algorithm 3 Computing a confidence interval for ρ through usual Bootstrap-t when normality is not assumed

---

1. For $b = 1, \ldots, B_1$
   (a) Sample $(X_{1b}^*, Y_{1b}^*), \ldots, (X_{nb}^*, Y_{nb}^*)$ with replacement from $(X_1, Y_1), \ldots, (X_n, Y_n)$
   (b) Compute r_b^* = sample correlation based on the bth bootstrap sample
   (c) Take $B_2 = 25$ bootstrap samples WR from $\{(X_{1b}^*, Y_{1b}^*), \ldots, (X_{nb}^*, Y_{nb}^*)\}$ (second-level bootstrap samples)
   (d) Compute r_{bl}^* = Pearson correlation computed from the lth 2nd-level bootstrap sample, and compute
   $$\widehat{\text{s.e.}}(r_b^*) = \sqrt{\frac{\sum_{l=1}^{B_2}\left(r_{bl}^* - \bar{r}_b^*\right)^2}{B_2 - 1}}$$
   (e) Compute
   $$t_b^* = \frac{r_b^* - r}{\widehat{\text{s.e.}}(r_b^*)}$$
2. Compute the $\alpha/2$ and $(1-\alpha/2)$th quantiles of $t_1^*, \ldots, t_{B_1}^*$. Call these $\hat{t}_{\alpha/2}$ and $\hat{t}_{1-\alpha/2}$.
3. The $(1-\alpha)100\%$ CI for ρ is given by
   $$\left(r - \hat{t}_{1-\alpha/2}\widehat{\text{s.e.}}(r), r - \hat{t}_{\alpha/2}\widehat{\text{s.e.}}(r)\right)$$
   where
   $$\widehat{\text{s.e.}}(r) = \sqrt{\frac{\sum_{b=1}^{B_1}\left(r_b^* - \bar{r}^*\right)^2}{B_1 - 1}},$$
   and $\bar{r}^*$ is the sample mean of the r_b^*s.

Example. (Efron& Tibshirani) The law school data. A random sample of size $n = 15$ was taken from a collection of $N = 82$ American law schools participating in a large study of admission practices. Two measurements were made on the entering classes of each school in 1973: LSAT, the average score for the class on a national law test, and GPA, the average undergraduate grade-point average for the class. The table below gives the data Suppose that we want a confidence interval for ρ, the correlation between LSAT and GPA. Normality is not assumed. We shall discuss two ways of using the Bootstrap-t to compute this CI. We first use the approach in Algorithm 3.

School	**LSAT**	**GPA**	**School**	**LSAT**	**GPA**
1	576	3.39	9	651	3.36
2	635	3.30	10	605	3.13
3	558	2.81	11	653	3.12
4	578	3.03	12	575	2.74
5	666	3.44	13	545	2.76
6	580	3.07	14	572	2.88
7	555	3.00	15	594	2.96
8	661	3.43			

——————————————————**R EXAMPLE**——————————————————

Algorithm 4 Computing a confidence interval for ρ through usual Bootstrap-t after transforming

1. For $b = 1, \ldots, B_1$

   (a) Sample $(X_{1b}^*, Y_{1b}^*), \ldots, (X_{nb}^*, Y_{nb}^*)$ with replacement from $(X_1, Y_1), \ldots, (X_n, Y_n)$

   (b) Compute

$$r_b^* = \text{ sample correlation based on the bth bootstrap sample}$$
$$\hat{\phi}_b^* = \frac{1}{2}\log\frac{1+r_b^*}{1-r_b^*} = \tanh^{-1}(r_b^*)$$

   (c) Take $B_2 = 25$ bootstrap samples WR from $\{(X_{1b}^*, Y_{1b}^*), \ldots, (X_{nb}^*, Y_{nb}^*)\}$ (second-level bootstrap samples)

   (d) Compute r_{bl}^* = Pearson correlation computed from the lth 2nd-level bootstrap sample, and compute

$$\widehat{\text{s.e.}}\left(\hat{\phi}_b^*\right) = \sqrt{\frac{\sum_{l=1}^{B_2}\left(\hat{\phi}_{bl}^* - \bar{\hat{\phi}}_b^*\right)^2}{B_2 - 1}}$$
$$\text{where } \hat{\phi}_{bl}^* = \frac{1}{2}\log\left(\frac{1+r_{bl}^*}{1-r_{bl}^*}\right)$$
$$r_{bl}^* = \text{Pearson correlation from } l\text{th bootstrap sample}$$
$$\bar{\hat{\phi}}_b^* = \text{sample mean of } \hat{\phi}_{bl}^*,\ l = 1, \ldots, B_2$$

   (e) Compute

$$t_b^* = \frac{\hat{\phi}_b^* - \hat{\phi}}{\widehat{\text{s.e.}}\left(\hat{\phi}_b^*\right)}$$
$$\text{where } \hat{\phi} = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

2. Compute the $\alpha/2$ and $(1-\alpha/2)$th quantiles of $t_1^*, \ldots, t_{B_1}^*$. Call these $\hat{t}_{\alpha/2}$ and $\hat{t}_{1-\alpha/2}$.

3. Compute the $(1-\alpha)100\%$ CI for ϕ is given by

$$\left(\hat{\phi} - \hat{t}_{1-\alpha/2}\widehat{\text{s.e.}}\left(\hat{\phi}\right), r - \hat{t}_{\alpha/2}\widehat{\text{s.e.}}\left(\hat{\phi}\right)\right)$$

   where

$$\widehat{\text{s.e.}}\left(\hat{\phi}\right) = \sqrt{\frac{\sum_{b=1}^{B_1}\left(\hat{\phi}_b^* - \bar{\hat{\phi}}^*\right)^2}{B_1 - 1}},$$
$$\text{where } \bar{\hat{\phi}}^* = \text{sample mean of } \hat{\phi}_b^*$$

4. The $(1-\alpha)100\%$ bootstrap-t CI for ρ is given by

$$\left(\frac{e^{2L}-1}{e^{2L}+1}, \frac{e^{2U}-1}{e^{2U}+1}\right)$$

   where $L = \hat{\phi} - \hat{t}_{1-\alpha/2}\widehat{\text{s.e.}}\left(\hat{\phi}\right)$ and $U = r - \hat{t}_{\alpha/2}\widehat{\text{s.e.}}\left(\hat{\phi}\right)$.

————————————————————————R EXAMPLE————————————————————————

Problem with the bootstrap-t

- The bootstrap-t approach is not transformation-respecting. That is, it makes a difference which scale is used to contract the interval. The transformation $\phi = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ works well for the bivariate normal. But in most problems we don't know the appropriate transformation: a major stumbling block!
- Solution: Use the bootstrap to estimate the appropriate transformation from the data itself, then use this transformation for the construction of a bootstrap-t interval

Example. Let $g(r) = \hat{\phi} = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$ and $g(\rho) = \phi = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$, then

$$\left(\hat{\phi} - \phi\right) \dot{\sim} N\left(0, \frac{1}{n-3}\right)$$

- Bootstrap-t intervals work better for variance-stabilized parameters!
- The transformation g is a variance-stabilizing transformation.

Using the bootstrap to estimate the transformation

- Suppose X is a random variable with mean θ and standard deviation $s(\theta)$ that varies as a function of θ
- Then a Taylor series argument shows that choosing a transformation g such that

$$g'(x) = \frac{1}{s(x)}$$

  has the property that the variance of $g(X)$ is approximately constant.
- Therefore

$$g(x) = \int^{x} \frac{1}{s(u)} du \tag{1.1}$$
$$s(u) = \text{s.e.}\left(\hat{\theta} | \theta = u\right)$$

- So our task is to estimate $s(u)$using the bootstrap then compute a bootstrap-t interval for $\phi = g(\theta)$, then finally transform this back by applying g^{-1} on the quantiles.

Algorithm 5 Computation of the variance-stabilized bootstrap-t interval

1. Generate B_1 bootstrap samples and for each sample $\boldsymbol{x}^{*b}$ compute the bootstrap replication $\hat{\theta}^*_b$. Take B_2 (second-level) bootstrap samples from $\boldsymbol{x}^{*b}$ and estimate the standard error $\widehat{\text{s.e.}}\left(\hat{\theta}^*_b\right)$
2. Fir a curve to the points $\left[\hat{\theta}^*_b, \widehat{\text{s.e.}}\left(\hat{\theta}^*_b\right)\right]$ to produce a smooth estimate of the function $s(u) = \text{s.e.}\left(\hat{\theta} | \theta = u\right)$
3. Estimate the variance stabilizing transformation $g\left(\hat{\theta}\right)$ from (1.1) using some sort of numerical integration.
4. Using B_3 new bootstrap samples, compute a bootstrap-t interval for $\phi = g(\theta)$. Since the standard error of $g\left(\hat{\theta}\right)$ is roughly constant as a function of θ, we don't need to estimate the denominator in the quantity $\left(g\left(\hat{\theta}^*\right) - g\left(\hat{\theta}\right)\right) / \widehat{\text{s.e.}}^*$ and can set it equal to one.
5. Map the endpoints of the interval back to the θ scale via the transformation g^{-1}.