Confidence Intervals

Nathaniel E. Helwig

University of Minnesota

1 Background and Motivation

Suppose that we have an independent and identically distributed (iid) sample of data $x_1, \ldots, x_n$ where $x_i \overset{\text{iid}}{\sim} F$ from some distribution F. Furthermore, suppose that the distribution F depends on some parameter $\theta = t(F)$, and we have formed an estimate of the parameter, denoted by $\hat{\theta} = s(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is the sample of data. As a reminder, the notation $\theta = t(F)$ denotes that the parameter is a function of the distribution, and the notation $\hat{\theta} = s(\mathbf{x})$ denotes that the estimate is a function of the sample. Given that $\hat{\theta}$ is a function of a random sample of data, the estimate $\hat{\theta}$ is a random variable that has some distribution, which we will denote by $F_{\hat{\theta}}$. Note that the distribution of the estimate, i.e., $F_{\hat{\theta}}$, will depend on (i) the form of the estimator, i.e., the function $s(\cdot)$ that is used to compute the estimate, (ii) the form of the data generating distribution F, and (iii) the sample size n.

Given that the estimate $\hat{\theta}$ is a random variable, there is an inherent amount of uncertainty in the estimate. In the previous chapter, we discussed some ways to explore the quality of an estimator: bias, variance, and mean squared error. As a reminder, the bias is concerned with the location of the estimator (i.e., the difference between the estimator's expected value and the true unknown parameter) and the variance is concerned with the spread of the estimator (i.e., the expected squared difference between the estimate and its expected value). If the sampling distribution of the estimator is normally distributed, i.e., if $\hat{\theta} \sim N(\mu_{\hat{\theta}}, \sigma^2_{\hat{\theta}})$ where $\mu_{\hat{\theta}} = E(\hat{\theta})$ is the expected value and $\sigma^2_{\hat{\theta}} = \text{Var}(\hat{\theta})$ is the variance, then we only need the parameters $\mu_{\hat{\theta}}$ and $\sigma^2_{\hat{\theta}}$ to understand how confident we can be in our estimate $\hat{\theta}$. However, if the sampling distribution $F_{\hat{\theta}}$ is some generic distribution, we need to know the distributional form and parameters to assess the confidence we can have in our estimate.

2 What is a Confidence Interval?

Definition. Given a confidence level $\alpha \in (0, 1)$, the probabilistic statement

$$P\left(a(\hat{\theta}) < \theta < b(\hat{\theta})\right) = 1 - \alpha$$

defines a $100(1-\alpha)\%$ confidence interval for the unknown parameter θ. The confidence interval endpoints $a(\cdot)$ and $b(\cdot)$ are functions of the estimate $\hat{\theta}$, which is a random variable. Thus, the $100(1-\alpha)\%$ confidence interval provides a range of values depending on $\hat{\theta}$ such that the probability of θ being within the interval is $1-\alpha$.

***Example* 1.** Suppose that $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ and we want to form a confidence interval for μ. As an estimate of μ, we will use the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Assuming that $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, we know that $\bar{x} \sim N(\mu, \sigma^2/n)$, which implies that $\sqrt{n}(\bar{x}-\mu)/\sigma \sim N(0, 1)$. As a result, we have that

$$P\left(z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha$$

where $z_\alpha = \Phi^{-1}(\alpha)$ with $\Phi^{-1}(\cdot)$ denoting the quantile function for the standard normal distribution. Rearranging the terms inside the above probability statement gives

$$\begin{aligned} 1 - \alpha &= P\left(z_{\alpha/2}\sigma/\sqrt{n} < \bar{x} - \mu < z_{1-\alpha/2}\sigma/\sqrt{n}\right) \\ &= P\left(z_{\alpha/2}\sigma/\sqrt{n} - \bar{x} < -\mu < z_{1-\alpha/2}\sigma/\sqrt{n} - \bar{x}\right) \\ &= P\left(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} > \mu > \bar{x} - z_{1-\alpha/2}\sigma/\sqrt{n}\right) \end{aligned}$$

which implies that a $100(1-\alpha)\%$ confidence interval for μ defines $a(\bar{x}) = \bar{x} - z_{1-\alpha/2}\sigma/\sqrt{n}$ and $b(\bar{x}) = \bar{x} - z_{\alpha/2}\sigma/\sqrt{n}$. Note that since $-z_{\alpha/2} = z_{1-\alpha/2}$ we can write the two endpoints of the confidence interval as

$$\bar{x} \pm z_{1-\alpha/2}\text{SE}(\bar{x})$$

where $\text{SE}(\bar{x}) = \sigma/\sqrt{n}$ is the standard error of the sample mean. In practice, it is typical to form a 90% confidence interval (i.e., $\alpha = 0.1$), which corresponds to $z_{0.95} \approx 1.65$, a 95% confidence interval (i.e., $\alpha = 0.05$), which corresponds to $z_{0.975} \approx 1.96$, or a 99% confidence interval (i.e., $\alpha = 0.01$), which corresponds to $z_{0.995} = 2.58$.

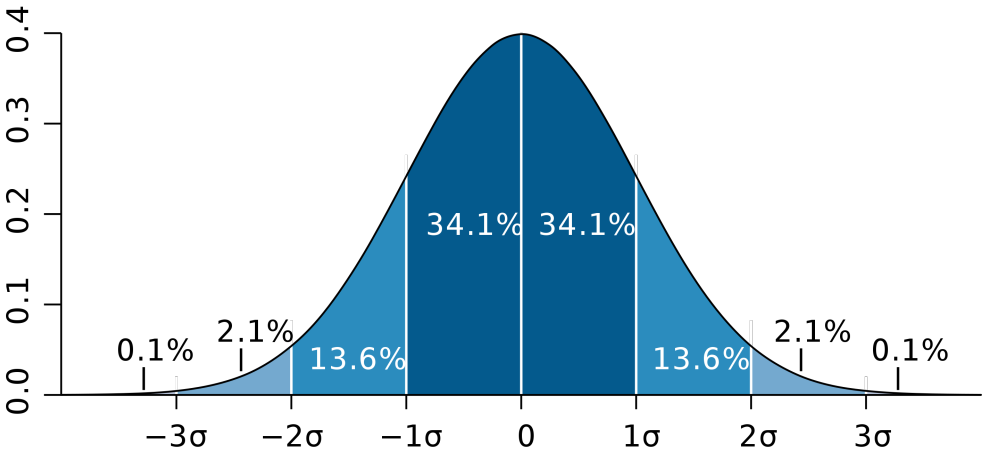

Figure 1: Area under the normal distribution from https://en.wikipedia.org/wiki/Normal_distribution

***Example* 2.** Suppose that $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ and we want to form a confidence interval for σ^2. As an estimate of σ^2, we will use the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Assuming that $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, we know that $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$. As a result, we have that

$$P\left(q_{n-1;\alpha/2} < (n-1)\frac{s^2}{\sigma^2} < q_{n-1;1-\alpha/2}\right) = 1 - \alpha$$

where $q_{n-1;\alpha} = Q_{n-1}(\alpha)$ with $Q_{n-1}(\cdot)$ denoting the quantile function for the χ^2_{n-1} distribution. Rearranging the terms inside the above probability statement gives

$$\begin{aligned}
1 - \alpha &= P\left(\frac{q_{n-1;\alpha/2}}{n-1} < \frac{s^2}{\sigma^2} < \frac{q_{n-1;1-\alpha/2}}{n-1}\right) \\
&= P\left(\frac{q_{n-1;\alpha/2}}{s^2(n-1)} < \frac{1}{\sigma^2} < \frac{q_{n-1;1-\alpha/2}}{s^2(n-1)}\right) \\
&= P\left(\frac{s^2(n-1)}{q_{n-1;\alpha/2}} > \sigma^2 > \frac{s^2(n-1)}{q_{n-1;1-\alpha/2}}\right)
\end{aligned}$$

which implies that a $100(1-\alpha)\%$ confidence interval for σ^2 defines $a(s^2) = (n-1)s^2/q_{n-1;1-\alpha/2}$ and $b(s^2) = (n-1)s^2/q_{n-1;\alpha/2}$. In this case, the quantile values $q_{n-1;1-\alpha/2}$ and $q_{n-1;\alpha/2}$ depend on the sample size n, so there are not standard quantiles values (such as 1.65, 1.96, and 2.58). But it is still typical to use a 90%, 95%, or 99% confidence interval.

3 Interpreting Confidence Intervals

Confidence intervals are often misinterpreted. The most common misinterpretation of a confidence interval is that there is a $100(1-\alpha)\%$ chance that $a(\hat{\theta}) < \theta < b(\hat{\theta})$ for a given estimate $\hat{\theta}$. Note that this interpretation is incorrect because for any given estimate $\hat{\theta}$ and corresponding confidence interval $[a(\hat{\theta}), b(\hat{\theta})]$, the inequality statement $a(\hat{\theta}) < \theta < b(\hat{\theta})$ is either true or false. The correct interpretation of a confidence interval is as follows. Suppose that we repeat our experiment a large number of independent times, i.e., we collect R independent samples of data each of size n. Let $\hat{\theta}_r$ denote the estimate of θ for the r-th sample of data, and let $[a(\hat{\theta}_r), b(\hat{\theta}_r)]$ denote confidence interval formed from $\hat{\theta}_r$ (for $r = 1, \ldots, R$). As the number of replications $R \to \infty$, we have that

$$\frac{1}{R}\sum_{r=1}^{R} I\left(a(\hat{\theta}_r) < \theta < b(\hat{\theta}_r)\right) = 1 - \alpha$$

where $I(\cdot)$ is an indicator function, i.e., $I(\cdot) = 1$ if the inequality statement is true, and $I(\cdot) = 0$ otherwise. Note that for the r-th replication, the true parameter θ is either in the interval or not, i.e., $I(\cdot)$ is either equal to 1 or 0.

***Example* 3.** For the example of forming a confidence interval for μ with $x_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, here is a simple demonstration of forming a 95% confidence interval using $R = 10000$ replications with $n = 25$ observations. Note that $\mu = 0$ and $\sigma^2 = 1$ in the below example.

```
> R <- 10000
> n <- 25
> set.seed(1)
> xbar <- replicate(R, mean(rnorm(n)))
> ci.lo <- xbar - qnorm(.975) / sqrt(n)     # 95% CI lower bound
> ci.up <- xbar - qnorm(.025) / sqrt(n)     # 95% CI upper bound
> ci.in <- (ci.lo <= 0) & (0 <= ci.up)
> summary(ci.in)
   Mode   FALSE    TRUE
logical     499    9501
> mean(ci.in)
[1] 0.9501
```

4 Lower and Upper Confidence Bounds

A confidence interval involves finding both the lower bound $a(\hat{\theta})$ and the upper bound $b(\hat{\theta})$ that provides a range (or interval) of values such that the probability of θ being within the interval is $1-\alpha$. However, in some cases, it may be preferable to define a lower or upper bound, i.e., just one of the endpoints, instead of an interval. Note that lower and upper confidence bounds are appropriate if there is a specific direction of interest that is important for making a decision.

Definition. Given a confidence level $\alpha \in (0,1)$, the probabilistic statement

$$P\left(a(\hat{\theta}) < \theta\right) = 1-\alpha$$

defines a $100(1-\alpha)\%$ lower confidence bound for the unknown parameter θ, and the probabilistic statement

$$P\left(\theta < b(\hat{\theta})\right) = 1-\alpha$$

defines a $100(1-\alpha)\%$ upper confidence bound for the unknown parameter θ.

Confidence bounds are frequently used in studies that are attempting to establish the effectiveness of a treatment, e.g., clinical trials. This is because establishing that a treatment is effective (in a hypothesized direction of interest) only requires using one of the two endpoints of a confidence interval. As a result, it makes more sense to use a lower or upper confidence bound (instead of an interval), because the bound places all of the uncertainty on the side of the interval that is of interest to the substantive problem/question.

***Example* 4.** Suppose that a drug company has created a new medication that is designed to treat depression. To study the efficacy of the treatment, it would be preferable to use a confidence bound (instead of interval), given that establishing the treatment's effectiveness only requires showing that the level of depressive symptoms has been reduced (on average) by the treatment. For example, if $Z = Y - X$ denotes the difference in the depression severity (Y is post-treatment and X is pre-treatment), and if $Z \sim N(\mu, \sigma^2)$, then it would make sense to use an upper confidence bound. This is because observing $b(\hat{\theta}) < 0$ would suggest an effective treatment, i.e., that depression severity has been reduced, so it is preferable to only focus on the upper confidence bound in this case.

5 Properties of Confidence Intervals

Definition. Given a procedure for forming confidence interval, the coverage rate refers to the proportion of times that the parameter θ is included within the interval using some specified number of replications $R \gg 1$.

For example, the observed coverage rate in the sample mean example was 0.9501 using $R = 10000$ replications. If the coverage rate is above the nominal level of $1 - \alpha$, then the confidence interval procedure is said to be conservative. In contrast, if the coverage rate is below the nominal level, then the confidence interval procedure is said to be liberal. Finally, the coverage rate is said to be "accurate" if it is approximately equal to the nominal rate.

Definition. The width of the confidence interval refers to the distance between the upper and lower bounds, i.e., width $= b(\hat{\theta}) - a(\hat{\theta})$.

Note that if two different confidence interval procedures have accurate coverage rates, then the narrower confidence interval (i.e., the one with the smaller width) should be preferred. This is because the width of the confidence interval relates to its efficiency, such that wider intervals are less efficient.

Definition. The shape of the confidence interval refers to the ratio of the distance between the upper bound and the estimate relative to the distance between the estimate and the lower bound, i.e., shape $= [b(\hat{\theta}) - \hat{\theta}]/[\hat{\theta} - a(\hat{\theta})]$.

If shape > 1, this indicates that the confidence interval is wider on the right side, and if shape < 1, this indicates that the confidence interval is wider on the left side. Note that if shape $= 1$, then the confidence interval is symmetric around the estimate $\hat{\theta}$, which is typically the case for confidence intervals of mean parameters.

IMPORTANT: The purpose of any method for forming a confidence interval is to produce endpoints $a(\hat{\theta})$ and $b(\hat{\theta})$ that provide accurate coverage rates. Thus, the coverage rate is the most important aspect of any procedure for forming a confidence interval. If the coverage rate is not accurate, then the width and shape do not really matter. So it is only typical to discuss the width and shape of the confidence interval *after* ensuring that the confidence interval procedure results in accurate coverage rates.

6 Forming Confidence Intervals

There are three general frameworks that can be used to form confidence intervals:

- *Parametric*: If the distribution of the estimate $\hat{\theta}$ can be exactly derived, then an exact confidence interval can be formed.
- *Asymptotic*: If the distribution of the estimate $\hat{\theta}$ can be asymptotically derived, then an (asymptotically) approximate confidence interval can be formed.
- *Nonparametric*: If the distribution of the estimate $\hat{\theta}$ is unknown, then resampling methods can be used to estimate $F_{\hat{\theta}}$, and the estimated distribution $\hat{F}_{\hat{\theta}}$ can be used to form approximate confidence intervals.

The first procedure (i.e., parametric) was used in the previous examples. More specifically, in the previous examples, we knew the exact probability distributions of $\bar{x}$ and s^2 under the specified data generation assumptions. Using these known distributions, we were able to construct and manipulate probability statements that enabled us to derive the lower and upper bounds of the confidence interval.

The second procedure (i.e., asymptotic) can be used to construct confidence intervals for large samples of data when we know the asymptotic (i.e., limiting) distribution of $\hat{\theta}$. If we are interested in forming a confidence interval for the population mean $\mu = E(X)$, then the central limit theorem (CLT) can be used to form an asymptotic confidence interval. As a reminder, the CLT tells us that the sample mean is asymptotically normally distributed, i.e., $\bar{x} \dot{\sim} N(\mu, \frac{\sigma^2}{n})$ for large n, if the data generating distribution F is non-normal.

The third procedure (i.e., nonparametric) is more computationally intensive than the other two procedures, but is much more general. Note that the first two procedures can only be used in the small number of circumstances where we know the exact or asymptotic distribution of $\hat{\theta}$, which is typically only the case for mean parameters. For example, if we want to form a confidence interval for the median, we don't have any nice theoretical argument that can be used. To form confidence intervals for generic parameters, the nonparametric bootstrap can be used (Efron, 1979; Efron and Tibshirani, 1993).

7 Nonparametric Bootstrap

7.1 Basic Procedure

Suppose that we have an iid sample of n observations from some unknown distribution F, i.e., $x_i \overset{\text{iid}}{\sim} F$, and assume that the distribution F depends on some parameter $\theta = t(F)$. Furthermore, given a sample of data $\mathbf{x} = (x_1, \ldots, x_n)^\top$, suppose that we can compute an estimate $\hat{\theta} = s(\mathbf{x})$, and assume that the distribution of $\hat{\theta}$ is unknown. In such cases, the nonparametric bootstrap uses the empirical cumulative distribution function $\hat{F}_n$ in place of the unknown data generating distribution F to approximate the distribution of $\hat{\theta}$.

The nonparametric bootstrap begins by forming the bootstrap distribution, which consists of $R \gg 1$ replications of the estimate (or statistic) $\hat{\theta}$. For $r = 1, \ldots R$, the bootstrap:

1. Defines $\mathbf{x}_r = (x_{1r}, \ldots, x_{nr})^\top$ where x_{ir} is sampled *with replacement* from $\{x_1, \ldots, x_n\}$
2. Calculates $\hat{\theta}_r = s(\mathbf{x}_r)$

Note that $\hat{\theta}_r$ is referred to as the r-th replicate of the statistic (or estimate), and the collection of R replicates $\{\hat{\theta}_r\}_{r=1}^R$ is referred to as the bootstrap distribution. In practice, it is typical to use $R \geq 10000$ resamples, but it may be necessary to set R even larger when the data generating distribution F is relatively skewed (Hesterberg, 2015). The bootstrap distribution $\{\hat{\theta}_r\}_{r=1}^R$ can be used to assess the uncertainty the estimate or to form different types of confidence intervals (Helwig, 2017a,b).

7.2 Standard Error and Bias

As an estimate of the variance of $\hat{\theta}$, the bootstrap uses $\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{R-1}\sum_{r=1}^R (\hat{\theta}_r - \bar{\theta})^2$, where $\bar{\theta} = \frac{1}{R}\sum_{r=1}^R \hat{\theta}_r$ is the mean of the bootstrap distribution. The corresponding estimate of the standard error has the form $\widehat{\text{SE}}(\hat{\theta}) = \left(\widehat{\text{Var}}(\hat{\theta})\right)^{1/2}$, which is simply the sample standard deviation of the bootstrap distribution. To estimate the bias of the estimator $\hat{\theta} = s(\mathbf{x})$, the bootstrap uses $\widehat{\text{Bias}}(\hat{\theta}) = \bar{\theta} - t(\hat{F}_n)$, where $\bar{\theta}$ is an estimate of $E(\hat{\theta})$ and $t(\hat{F}_n)$ is an estimate of the parameter θ. As a reminder, $\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$ where the expectation is calculated with respect to F and $\theta = t(F)$. With real data, we never know F or the true parameter θ, so the bootstrap estimate of bias uses the ECDF $\hat{F}_n$ in place of the unknown true CDF F.

7.3 Confidence Intervals

7.3.1 Naive Methods

The bootstrap distribution can be used to form a variety of different types of confidence intervals (see Helwig, 2017b, for a discussion). The simplest approach is to define a confidence interval as one normally would, but using the bootstrap estimate of the standard error $\widehat{\text{SE}}(\hat{\theta})$ in place of the true standard error, i.e., $\hat{\theta} \pm z_{1-\alpha/2}\widehat{\text{SE}}(\hat{\theta})$. Note that this simple approach is only appropriate if (i) the sampling distribution of $\hat{\theta}$ is approximately normal, and (ii) the theoretical form of the standard error is unknown. A better approach is to use the corresponding quantiles of the bootstrap distribution, which is referred to as the "percentile method". More specifically, the percentile method would define the $100(1-\alpha)\%$ confidence interval of θ by defining the endpoints as $a(\hat{\theta}) = \hat{Q}_R(\alpha/2)$ and $b(\hat{\theta}) = \hat{Q}_R(1-\alpha/2)$, where $\hat{Q}_R(\cdot)$ is the sample quantile function, which is computed from the bootstrap distribution. Although this approach is better than the first method, it is still less accurate than some of the more sophisticated methods for calculating bootstrap confidence intervals.

7.3.2 t Table Method

One of the better methods for calculating bootstrap confidence intervals is referred to as the "t table" method. This approach calculates a t-like statistic for each bootstrap replicate

$$t_r = \frac{\hat{\theta}_r - \hat{\theta}}{\widehat{\text{SE}}(\hat{\theta}_r)}$$

where $\widehat{\text{SE}}(\hat{\theta}_r)$ is an estimate of the standard error of $\hat{\theta}_r$. In most cases, the standard error of $\hat{\theta}_r$ is unknown, so this confidence interval method typically involves using a bootstrap within the bootstrap to estimate $\widehat{\text{SE}}(\hat{\theta}_r)$. In other words, for each bootstrap sample $\mathbf{x}_r = (x_{1r}, \ldots, x_{nr})^\top$, we would typically need to use bootstrap resampling to estimate $\widehat{\text{SE}}(\hat{\theta}_r)$. In this inner bootstrap, the r-th bootstrap sample $\mathbf{x}_r$ is treated as the observed data $\mathbf{x}$, and the r-th replicate $\hat{\theta}_r$ is treated as the observed estimate $\hat{\theta}$. The bootstrap t statistic distribution $\{t_r\}_{r=1}^R$ is then used to form the $100(1-\alpha)\%$ confidence interval, such that $a(\hat{\theta}) = \hat{\theta} - \hat{q}_R(1-\alpha/2)\widehat{\text{SE}}(\hat{\theta})$ and $b(\hat{\theta}) = \hat{\theta} - \hat{q}_R(\alpha/2)\widehat{\text{SE}}(\hat{\theta})$, where $\hat{q}_R(\cdot)$ are the sample quantiles of the bootstrap t statistic distribution $\{t_r\}_{r=1}^R$.

7.3.3 BCa Method

Another good method for forming bootstrap confidence intervals is the "bias corrected and accelerated" (BCa) method. This method is similar to (but better than) the percentile method, given that it uses a corrected version of the bootstrap quantiles $\hat{Q}_R(\cdot)$. More specifically, the BCa method defines the endpoints of the $100(1-\alpha)\%$ confidence interval as $a(\hat{\theta}) = \hat{Q}_R(\alpha_1)$ and $b(\hat{\theta}) = \hat{Q}_R(\alpha_2)$, where α_1 and α_2 are the corrected probabilities that are input into the sample quantile function (in place of $\alpha/2$ and $1-\alpha/2$). The corrected probabilities have the form

$$\alpha_1 = \Phi\left(z_0 + \frac{z_0 + z_{\alpha/2}}{1 - \gamma(z_0 + z_{\alpha/2})}\right)$$
$$\alpha_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha/2}}{1 - \gamma(z_0 + z_{1-\alpha/2})}\right)$$

where $\Phi(\cdot)$ is the CDF for the standard normal distribution, $z_\alpha = \Phi^{-1}(\alpha)$ is the α-th quantile of the standard normal distribution, z_0 is the bias correction factor, and γ is the acceleration parameter. If $z_0 = \gamma = 0$, then $\alpha_1 = \alpha/2$ and $\alpha_2 = 1 - \alpha/2$, which reveals that the BCa method is equivalent to the percentile method when no corrections are needed.

In practice, the bias correction and acceleration factors can be estimated as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{1}{R}\sum_{r=1}^{R} I(\hat{\theta}_r < \hat{\theta})\right) \qquad \text{and} \qquad \hat{\gamma} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right]^{3/2}}$$

where $\hat{\theta}_{(i)}$ is the estimate of θ that would be obtained if x_i was excluded from the sample (which is referred to as the "jackknife estimate") and $\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}$ is the average of the jackknife estimates. The bias correction estimate $\hat{z}_0$ quantifies the median bias of the bootstrap distribution, which relates to the difference between median$(\hat{\theta}_r)$ and $\hat{\theta}$. The acceleration estimate $\hat{\gamma}$ quantifies the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter θ. Note that the bias correction estimate $\hat{z}_0$ is simple to compute, whereas the acceleration estimate $\hat{\gamma}$ is more computationally costly. However, assuming that $n \ll R$, estimating $\hat{\gamma}$ is less costly than estimating $\widehat{\text{SE}}(\hat{\theta}_r)$, which often makes the BCa method less computationally costly than the t table method.

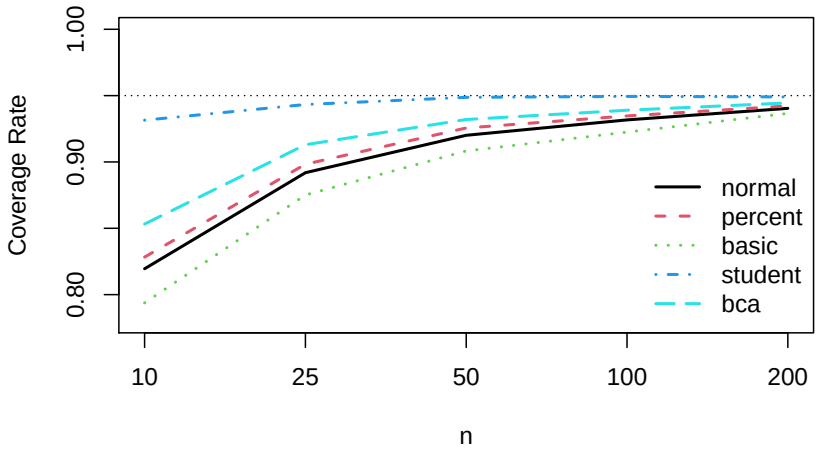

Figure 2: Coverage rates for different bootstrap confidence interval methods.

7.4 Example

A simple simulation study was conducted to explore the quality of the different confidence interval methods. The simulation involved generating $n \in \{10, 20, 50, 100, 200\}$ observations from a χ^2_1 distribution, and then using the nonparametric bootstrap to form a confidence interval for the mean μ. As a reminder, the mean of a chi square distribution is the degrees of freedom parameter k, which in this case is $k = 1$. For each sample size, I generated 10,000 independent samples of data of size n, and then I used the nonparametric bootstrap methods to form 95% confidence intervals. The nonparametric bootstrap was implemented using the `np.boot` function in the **nptest** R package (Helwig, 2020). The coverage rate was defined as the proportion of the 10,000 replications where the given confidence interval method contained the true parameter $\mu = 1$. Comparing the results, the studentized (i.e., t table) method performed best, but this is not surprising. The parameter of interest is the mean, so we have a nice formula for calculating $\widehat{\text{SE}}(\hat{\theta}_r)$, which means that we didn't need to rely on an inner bootstrap to estimate the standard error of $\hat{\theta}_r$. For other parameters, where we don't have a nice formula for calculating $\widehat{\text{SE}}(\hat{\theta}_r)$, we wouldn't expect the studentized method to perform as well. Also, note that the BCa method performed better than the three basic methods, which is not surprising.

References

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics* *7*(1), 1–26.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Boca Raton: Chapman & Hall.

Helwig, N. E. (2017a). Bootstrap Resampling. http://users.stat.umn.edu/~helwig/notes/boot-Notes.pdf.

Helwig, N. E. (2017b). Bootstrap Confidence Intervals. http://users.stat.umn.edu/~helwig/notes/bootci-Notes.pdf.

Helwig, N. E. (2020). *nptest: Nonparametric Tests*. R package version 1.0-2.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician* *69*(4), 371–386.