4 Confidence Intervals

Point estimators provide a limited amount of information about the unknown parameter in that they only specify a single value. A more useful approach to estimation is to specify a range of plausible values together with some measure of plausibility. We will develop confidence intervals first for the mean of a normal population and then for general parameters.

4.1 The Z Confidence Interval for a Normal Mean

Again suppose that $X_1, \ldots, X_n$ forms a random sample from a normal distribution $N(\mu, \sigma^2)$ where σ^2 is known. We want to construct a random interval that will contain the mean μ with a specified probability γ. Thus, we form an interval $(l(x_1, \ldots, x_n), u(x_1, \ldots, x_n))$ such that

$$P_\mu[l(X_1, \ldots, X_n) \leq \mu \leq u(X_1, \ldots, X_n)] = \gamma.$$

We will use the sampling distribution of $\bar{X}_n$ from Chapter 4 to derive this interval.

We let $\phi(x)$ and $\Phi(x)$ denote the pdf and cdf, respectively, of the standard normal distribution, and let c be a constant such that

$$P(-c < Z < c) = \int_{-c}^{c} \phi(x)dx = \Phi(c) - \Phi(-c) = \gamma.$$

Next

$$\begin{aligned} -c < Z < c \quad &\Longleftrightarrow \quad -c < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < c \\ &\Longleftrightarrow \quad \bar{X}_n - c\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + c\frac{\sigma}{\sqrt{n}} \end{aligned}$$

Thus,

$$P\left(\bar{X}_n - c\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + c\frac{\sigma}{\sqrt{n}}\right) = \gamma.$$

This says that with probability equal to γ, the unknown value of μ falls between the random variables,

$l(X_1, \ldots, X_n) = \bar{X}_n - c\frac{\sigma}{\sqrt{n}}$ and $u(X_1, \ldots, X_n) = \bar{X}_n + c\frac{\sigma}{\sqrt{n}}$.

The value of c is determined by the fact that $P(-c < Z < c) = \gamma$. This implies that

$$\gamma = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1.$$

Thus, $\Phi(c) = \frac{1+\gamma}{2}$ and $c = \Phi^{-1}((1+\gamma)/2) = z_{(1+\gamma)/2}$,

where z_α denotes the α^{th} quantile of the standard normal distribution.

Remark: We can form the *likelihood interval* $C(x_1, \ldots, x_n)$ for μ by letting

$$C(x_1, \ldots, x_n) = \{\mu : L(\mu | x_1, \ldots, x_n) \geq k(x_1, \ldots, x_n)\}$$

for some $k(x_1, \ldots, x_n)$. The text on page 327 shows that the likelihood interval for μ is the same as the interval that we derived above.

4.2 A General Definition of Confidence Intervals

We now define a confidence interval for a function $\psi(\theta)$ of a parameter θ where the joint distribution of $X_1, \ldots, X_n$ has pdf/pmf $f_\theta(x_1, \ldots, x_n)$ where $\theta \in \Omega$.

We say that an interval $C(X_1, \ldots, X_n) = (l(X_1, \ldots, X_n), u(X_1, \ldots, X_n))$ is a level γ confidence interval for $\psi(\theta)$ if

$$\begin{aligned} P_\theta(\psi(\theta) \in C(X_1, \ldots, X_n)) &= P_\theta(l(X_1, \ldots, X_n) \le \psi(\theta) \le u(X_1, \ldots, X_n)) \\ &\ge \gamma \qquad \text{for every } \theta \in \Omega. \end{aligned}$$

Thus, the random interval $C(X_1, \ldots, X_n)$ has a probability of a least γ (usually 0.95 or larger) of containing the true value of $\psi(\theta)$.

In the last example we saw that
$(\bar{X}_n - z_{(1+\gamma)/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{(1+\gamma)/2}\frac{\sigma}{\sqrt{n}})$ is a level$-\gamma$ confidence interval for μ when sampling from a normal distribution with known variance σ^2.

4.3 The t Confidence Interval for a Normal Mean

We now consider the more realistic setting where $X_1, \ldots, X_n$ is a random sample from a normal (μ, σ^2) distribution where $\theta = (\mu, \sigma^2) \in \Omega = \mathcal{R} \times (0, \infty)$ is the unknown parameter. We will derive a level γ confidence interval for $\psi(\theta) = \mu$. The derivation will be similar to that where σ^2 was known, but we will need to use the t distribution rather than the normal distribution in our derivation.

From Slide 44 of Chapter 4, we know that

$$T = \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1).$$

We let $f_{n-1}(x)$ denote the pdf of the t distribution with $n-1$ df, and let c be a constant such that

$$P(-c < T < c) = \int_{-c}^{c} f_{n-1}(x)dx = \gamma.$$

Next

$$-c < T < c \iff -c < \frac{\bar{X}_n - \mu}{S/\sqrt{n}} < c$$

$$\iff \bar{X}_n - c\frac{S}{\sqrt{n}} < \mu < \bar{X}_n + c\frac{S}{\sqrt{n}}$$

Thus,

$$P\left(\bar{X}_n - c\frac{S}{\sqrt{n}} < \mu < \bar{X}_n + c\frac{S}{\sqrt{n}}\right) = \gamma.$$

This says that the probability equals γ that the unknown value of μ falls between the random variables, $l = \bar{X}_n - c\frac{S}{\sqrt{n}}$ and $u = \bar{X}_n + c\frac{S}{\sqrt{n}}$.

Here $c = t_{(1+\gamma)/2}(n-1)$ where $t_\alpha(\lambda)$ denotes the α^{th} quantile of the $t(\lambda)$ distribution.

In applications, we observe the data $X_1 = x_1, \ldots, X_n = x_n$ and compute $l(x_1, \ldots, x_n) = \bar{x}_n - t_{(1+\gamma)/2}(n-1)\frac{s}{\sqrt{n}}$ and $u(x_1, \ldots, x_n) = \bar{x}_n + t_{(1+\gamma)/2}(n-1)\frac{s}{\sqrt{n}}$.

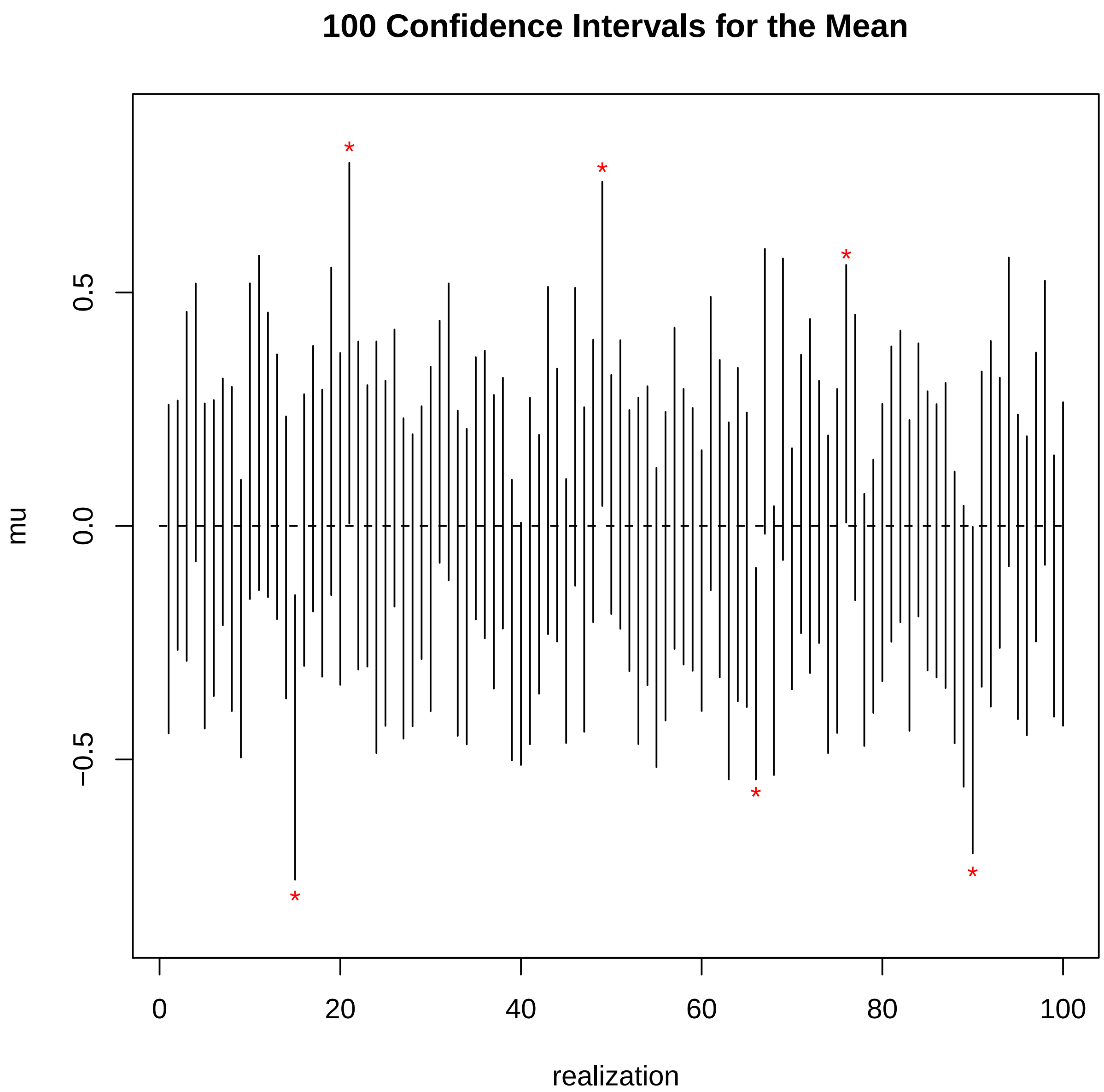
We say that the interval (l, u) is a confidence interval for μ with confidence coefficient γ.

Often we will write the γ confidence interval for μ as

$$\bar{x} \pm t_{(1+\gamma)/2}(n-1)\frac{s}{\sqrt{n}}$$

We say that μ lies in this observed interval with confidence (not probability) γ. The frequentist interpretation of confidence intervals follows:

> If we construct a large number of level γ confidence intervals for μ, about $100(\gamma)\%$ of them will contain μ. We illustrate this in the figure on the next slide.


100 Confidence Intervals for the Mean
mu
0.5
0.0
−0.5
0
20
40
60
80
100
realization

Remark: The construction of the confidence interval for μ depended on the random variable T which was a function of the data and the parameter. The important aspect of T was that it had a distribution that did not depend on the unknown parameter. We call such a random variable a pivot.

Example 45 again Suppose that $X_1, \ldots, X_n$ form a random sample from a $N(\mu, \sigma^2)$ distribution. We want to construct a level γ confidence interval for σ^2.

Earlier we learned that

$$W = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi^2(n-1).$$

We can use W as a pivot.

Let $g_{n-1}(x)$ be the pdf of the χ^2 distribution with $n-1$ df, and let c_1 and c_2 be constants such that

$$P(c_1 < W < c_2) = \int_{c_1}^{c_2} g_{n-1}(x)dx = \gamma.$$

Then

$$\begin{aligned} c_1 < W < c_2 \quad &\Longleftrightarrow \quad c_1 < \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2} < c_2 \\ &\Longleftrightarrow \quad \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{c_2} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{c_1}. \end{aligned}$$

Thus,

$$\left(\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{c_2}, \frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{c_1} \right)$$

is a level γ confidence interval for σ^2.

4.4 Pivots

We summarize our approach to forming confidence intervals through the use of pivots. Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution with pmf or pdf $f_\theta(x)$.

A pivot, $W(X_1, \ldots, X_n, \theta)$, is a random variable whose distribution does not depend on θ. (Often $W(x_1, \ldots, x_n, \theta)$ is a monotone function of θ.)

Define a and b to be constants such that

$$P(a < W(X_1, \ldots, X_n, \theta) < b) = \gamma.$$

When we observe $X_1 = x_1, \ldots, X_n = x_n$, we define the level γ confidence interval for θ as

$$C(x_1, \ldots, x_n) = \{\theta : a < W(x_1, \ldots, x_n, \theta) < b\}.$$

4.5 Approximate Confidence Intervals for Based on the MLE

Whenever the asymptotic distribution of an estimator $\hat{\theta}$ is normal, i.e.,

$$\hat{\theta} \overset{\text{approx}}{\sim} N(\theta_0, \hat{V}_n),$$

we can form an approximate pivot:

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{V}_n}} \overset{\text{approx}}{\sim} N(0, 1)$$

We use this pivot to form an approximate $100(\gamma)\%$ confidence interval for θ:

$$\hat{\theta} \pm Z_{(1+\gamma)/2} ASE(\hat{\theta})$$

where $Z_{(1+\gamma)/2}$ is a value such that $P[Z \leq Z_{(1+\gamma)/2}] = (1+\gamma)/2$ and Z is a standard normal rv. This interval is called the Wald confidence interval for θ.

<u>Example 46 again</u> Suppose that $X \sim$ Binomial (n, θ). We can represent X as $X_1 + \ldots + X_n$ where $X_i \sim$ Bernoulli (θ). Then the Fisher's information in a single X_i is

$$I(\theta) = \frac{1}{\theta(1-\theta)},$$

Then

$$V_n(\theta) = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$$

and

$$ASE(\hat{\theta}) = \sqrt{\frac{1}{nI(\hat{\theta})}} = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

A level γ confidence interval for θ is given by

$$\hat{\theta} \pm Z_{(1+\gamma)/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \quad \text{where} \quad \hat{\theta} = \frac{x}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Alternatively, we could form a different pivot that is also based on the asymptotic properties of the MLE. We know that

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \longrightarrow \infty$$

Thus,

$$P\left[-Z_{(1+\gamma)/2} < \frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{\theta(1-\theta)}} < Z_{(1+\gamma)/2}\right] = \gamma$$

We obtain the level $1-\alpha$ confidence interval:

$$\left\{\theta : n(\hat{\theta}-\theta)^2 < \theta(1-\theta)Z^2_{(1+\gamma)/2}\right\}$$

or
$$\frac{\hat{\theta}+\frac{Z^2_{(1+\gamma)/2}}{2n} \pm Z_{(1+\gamma)/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}+\frac{Z^2_{(1+\gamma)/2}}{4n^2}}}{1+Z^2_{(1+\gamma)/2}/n}$$

This interval is called the score confidence interval for θ.

Discussion of the Confidence Intervals for the Binomial Distribution

The two forms of confidence intervals for the binomial proportion will produce similar results for large sample sizes ($n > 100$) as long as the observed number of successes x and observed number of failures $n - x$ are not close to zero (at least 15). For smaller sample sizes and in the situation where x or $n - x$ is small, the score interval has much better coverage probability.

Example: In a survey of 277 randomly selected adult shoppers, 69 stated that if an advertised item is unavailable they request a rain check, construct 95% CIs for θ.

The 95% Wald interval is

$$0.249 \pm 1.96\sqrt{0.249(1-0.249)/277} = 0.249 \pm 0.051$$

The 95% Wald interval is $(.198, .300)$.

We can also compute the 95% score confidence interval:

$$\frac{0.249 + \dfrac{1.96^2}{2(277)} \pm 1.96\sqrt{\dfrac{0.249 \cdot 0.751}{277} + \dfrac{1.96^2}{4(277)^2}}}{1 + (1.96^2)/277}$$

We obtain the 95% score confidence interval for θ: $(0.202, 0.303)$. This appears similar to the 95% Wald interval, $(.198, .300)$.

Bill of Rights Data: Recall that $x = 14$ and $n = 50$ in this example.

The 95% Wald interval is $(0.156, 0.404)$ whereas the 95% score interval is $(0.175, 0.417)$. There is a greater difference between the two intervals due to the small sample size.

Example 46 again Suppose that $X_1, \ldots, X_n$ is a random sample from an exponential (λ) distribution. Obtain exact and approximate level γ confidence interval for λ.

Exact confidence interval: Consider the random variable $W = 2\lambda \sum_{i=1}^n X_i$. Its moment generating function is

$$M_W(s) = M_{2\lambda \sum_{i=1}^n X_i}(s) = \prod_{i=1}^n M_{X_i}(2\lambda s) = \left(\frac{\lambda}{\lambda - 2\lambda s}\right)^n = \left(\frac{1}{1-2s}\right)^{2n/2}.$$

Thus, W has a chi-squared distribution with $2n$ degrees of freedom. Let a and b be values such that $G_{2n}(b) - G_{2n}(a) = \gamma$ where $G_{2n}(x)$ is the cdf of the chi-squared distribution with $2n$ degrees of freedom. Then

$$\begin{aligned} \gamma &= P[a \le W \le b] = P[a \le 2\lambda \sum_{i=1}^n X_i \le b] \\ &= P\left[\frac{a}{2\sum_{i=1}^n X_i} \le \lambda \le \frac{b}{2\sum_{i=1}^n X_i}\right] \end{aligned}$$

Thus, an exact level γ confidence interval for λ is

$$\left[\frac{a}{2\sum_{i=1}^{n} X_i}, \frac{b}{2\sum_{i=1}^{n} X_i}\right]$$

Approximate confidence interval I: From slide 64,we have the result that $\hat{\lambda} \sim AN(\lambda_0, \lambda_0^2/n)$. For the Wald interval, we will use the pivot $Z = \frac{\hat{\lambda}-\lambda_0}{\hat{\lambda}/\sqrt{n}}$. This has an approximate $N(0,1)$ distribution and we have

$$\begin{aligned}\gamma &= P\left[-Z_{(1+\gamma)/2} \leq \frac{\hat{\lambda}-\lambda_0}{\hat{\lambda}/\sqrt{n}} \leq Z_{(1+\gamma)/2}\right] \\ &= P\left[\hat{\lambda} - \frac{Z_{(1+\gamma)/2}\hat{\lambda}}{\sqrt{n}} \leq \lambda_0 \leq \hat{\lambda} + \frac{Z_{(1+\gamma)/2}\hat{\lambda}}{\sqrt{n}}\right].\end{aligned}$$

Thus,

$$\left[\hat{\lambda}(1 - \frac{Z_{(1+\gamma)/2}}{\sqrt{n}}), \hat{\lambda}(1 + \frac{Z_{(1+\gamma)/2})}{\sqrt{n}}\right]$$

is an approximate level γ confidence interval for λ.

Approximate confidence interval II: From slide 64,we have the result that $\hat{\lambda} \sim AN(\lambda_0, \lambda_0^2/n)$. For the score interval, we will use the pivot $Z = \frac{\hat{\lambda}-\lambda_0}{\lambda_0/\sqrt{n}}$. This has an approximate $N(0,1)$ distribution and we have

$$\begin{aligned} \gamma &= P\left[-Z_{(1+\gamma)/2} \leq \frac{\hat{\lambda}-\lambda_0}{\lambda_0/\sqrt{n}} \leq Z_{(1+\gamma)/2}\right] \\ &= P\left[\frac{-Z_{(1+\gamma)/2}}{\sqrt{n}} \leq \frac{\hat{\lambda}}{\lambda_0} - 1 \leq \frac{Z_{(1+\gamma)/2}}{\sqrt{n}}\right] \\ &= P\left[\frac{\hat{\lambda}}{1+\frac{Z_{(1+\gamma)/2}}{\sqrt{n}}} \leq \lambda_0 \leq \frac{\hat{\lambda}}{1-\frac{Z_{(1+\gamma)/2}}{\sqrt{n}}}\right]. \end{aligned}$$

Thus,

$$\left[\frac{\hat{\lambda}}{1+\frac{Z_{(1+\gamma)/2}}{\sqrt{n}}}, \frac{\hat{\lambda}}{1-\frac{Z_{(1+\gamma)/2)}}{\sqrt{n}}}\right]$$

is an approximate level γ confidence interval for λ.

<u>Example 39 again</u> Find a 95% confidence intervals for the rate parameter in the nerve pulse data. For this set of data, $n = 799$ and $\sum_{i=1}^{799} = 174.64$.

- Exact 95% confidence interval with equal tail probabilities: Let G denote the cdf of the $\chi^2(1598)$ distribution. Then the exact 95% ci is

$$\left[\frac{G^{-1}(0.025)}{2\sum_{i=1}^{799} X_i}, \frac{G^{-1}(0.975)}{2\sum_{i=1}^{799} X_i}\right] = \left[\frac{1489.1}{2(174.64)}, \frac{1710.7}{2(174.64)}\right] = (4.262, 4.898)$$

- 95% Wald interval based on the mle:

$$\hat{\lambda} \pm Z_{0.975}\hat{\lambda}/\sqrt{n} = 4.575 \pm 1.96 \times 4.575/\sqrt{799} = 4.575 \pm 0.317$$

Thus, the approximate 95% confidence interval is $(4.258, 4.892)$

- 95% Score interval based on the mle:

$$\left[\frac{\hat{\lambda}}{1 + \frac{Z_{0.975}}{\sqrt{n}}}, \frac{\hat{\lambda}}{1 - \frac{Z_{0.975}}{\sqrt{n}}}\right] = \left[\frac{4.575}{1 + \frac{1.96}{\sqrt{799}}}, \frac{4.575}{1 - \frac{1.96}{\sqrt{799}}}\right] = (4.278, 4.916)$$

4.6 Bootstrap Confidence Intervals

The bootstrap can be used to obtain approximate confidence intervals for a function of a population parameter, $\psi(\theta) = T(F_\theta)$ where T is a function of the distribution F_θ. For example, T could refer to a moment of F_θ or a quantile of F_θ. We can write the estimate based on the original data as $\hat{\psi} = T(F_n)$ where F_n is the empirical cdf,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i).$$

This is a step function with jumps of size $1/n$ at each single value of x_i. We note that F_n is a cdf that puts weight $1/n$ at each observed value of x_i.

To construct a confidence interval for $\psi(\theta) = T(F_\theta)$, we need to know the distribution of $\hat{\psi} = T(F_n)$. The bootstrap provides a computational method for approximating the distribution of $\hat{\psi}$.

Since F_θ is not known, we can use the nonparametric bootstrap to approximate the distribution of $\hat{\psi} = T(F_n)$:

- Generate B bootstrap samples of size n from the distribution with cdf F_n. That is, take a sample of size n with replacement from $\{x_1, \ldots, x_n\}$.
- For each bootstrap sample $\{x_1^*, \ldots, x_n^*\}$, compute the estimate of $\psi(\theta)$, say $\hat{\psi}_j^*,\ j = 1, \ldots, B$.
- Compute the *bootstrap percentile confidence interval* for $\psi(\theta)$ by computing

$$(\hat{\psi}_{(1-\gamma)/2}, \hat{\psi}_{(1+\gamma)/2}),$$

where $\hat{\psi}_p$ refers to the p^{th} quantile of the bootstrap estimates, $\{\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*\}$.

Remark: If the form of F_θ is known, the parametric bootstrap replaces the first step by taking a random sample of size n from a distribution with cdf $F_{\hat{\theta}}$ where $\hat{\theta}$ is the mle of θ.

t **Confidence Intervals** An alternative approach to forming a bootstrap confidence interval is to form a t interval centered at the estimate for the observed data and then using the bootstrap estimate of the standard error of the estimator. This results in the t confidence interval for $\psi(\theta)$:

$$\hat{\psi} \pm t_{(1+\gamma)/2} \sqrt{\widehat{\mathrm{Var}}_{\hat{F}}(\hat{\psi})},$$

where $\widehat{\mathrm{Var}}_{\hat{F}}(\hat{\psi})$ is the sample variance of the B bootstrap estimates of ψ.

Remark: The above methods for forming confidence intervals are very basic applications of the bootstrap. There are other approaches using the bootstrap for confidence intervals that result in better performance, especially when the bootstrap distribution of the estimator is skewed. See the text by Efron and Tibshirani for more details.

<u>Example 39 again</u> Find a 95% confidence interval for the rate parameter in the nerve pulse data.

- Wald interval based on the mle: From Slide 64, $I(\lambda) = 1/\lambda^2$. Thus, the approximate 95% confidence interval is $(4.258, 4.892)$:

  $$\hat{\lambda} \pm Z(0.025)\hat{\lambda}/\sqrt{n} = 4.575 \pm 1.96 \times 4.575/\sqrt{799} = 4.575 \pm 0.317$$

- Bootstrap confidence intervals:

```
> temp=rep(0,10000)
> for (i in 1:10000)temp[i]=1/mean(sample(nerve,replace=TRUE))
> quantile(temp,c(0.025,0.975))
    2.5%    97.5%
4.280029 4.896433
> 1/mean(nerve)-1.96*sd(temp)
[1] 4.26618
> 1/mean(nerve)+1.96*sd(temp)
[1] 4.884072
```


The resulting confidence intervals are $(4.280, 4.896)$ and $(4.266, 4.884)$.

- Parametric bootstrap confidence intervals:

  When F_θ has a known form, we can use the parametric bootstrap to approximate the distribution of $\hat{\psi} = T(F_n)$:

```
> temp=rep(0,10000)
> for (i in 1:10000)temp[i]=1/mean(rexp(799,rate=1/mean(nerve)))
> quantile(temp,c(0.025,0.975))
    2.5%    97.5%
4.275859 4.904623
> 1/mean(nerve)-1.96*sd(temp)
[1] 4.258208
> 1/mean(nerve)+1.96*sd(temp)
[1] 4.892044
```

The resulting confidence intervals are $(4.276, 4.905)$ and $(4.258, 4.892)$.

We summarize the various 95% confidence intervals:

Method	Interval
Exact	$(4.262, 4.898)$
Wald	$(4.258, 4.892)$
Score	$(4.278, 4.916)$
Nonparametric bootstrap quantile	$(4.280, 4.896)$
Nonparametric t	$(4.266, 4.884)$
Parametric bootstrap quantile	$(4.276, 4.905)$
Parametric t	$(4.258, 4.892)$

