Bootstrap Method

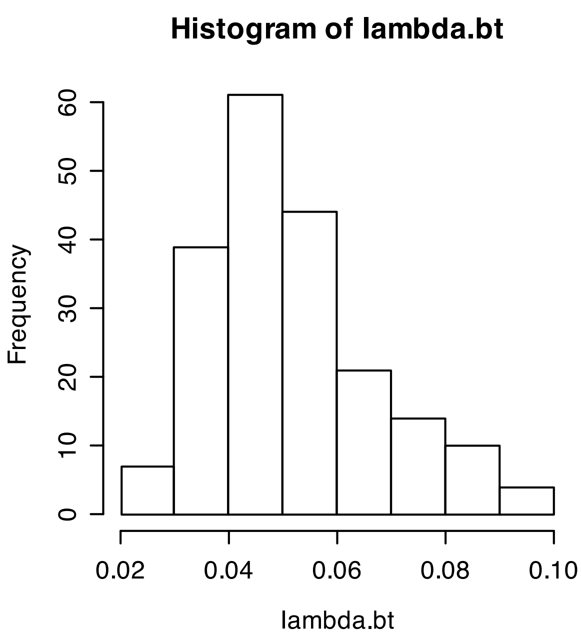
```
> # Purpose:  understand how bootstrap method works
> obs=c(11.96, 5.03, 67.40, 16.07, 31.50, 7.73, 11.10, 22.38)
> n=length(obs)
> mean(obs)
[1] 21.64625
> # estimate of lambda
> lambda = 1/mean(obs); lambda
[1] 0.04619738
> # The exponential distribution with rate=lambda has density f(x) =
lambda e^(- lambda x)
> # draw a random sample of size n from exponential(lambda)
distribution, where lambda is estimated from data
> x=rexp(n, lambda); x
[1]  4.451616 12.097513  6.302449 21.942872 37.191007 76.530816
9.458349 11.386464
> mean(x)
[1] 22.42014
> 1/mean(x)  # Bootstrap estimate of lambda
[1] 0.04460276
> # do it again
> x=rexp(n, lambda); x
[1]  9.4759699  2.5089895  0.1630891  0.7994896 51.2508151 10.9096888
9.2945093  2.8122216
> mean(x)
[1] 10.90185
> 1/mean(x)  # Bootstrap estimate of lambda
[1] 0.09172758
> # here we do a nonparametrical bootstrape by replacing x=rexp(n,
lambda) with x=sample(obs, n, replace=T)
> x=sample(obs, n, replace=T); x
[1] 11.10 22.38 67.40 31.50  5.03  7.73 11.10 22.38
> mean(x)
[1] 22.3275
> 1/mean(x)  # Bootstrap estimate of lambda
[1] 0.04478782
> # do it again
> x=sample(obs, n, replace=T); x
[1]  5.03 11.10 11.96 16.07 11.96 31.50 16.07 11.96
> mean(x)
[1] 14.45625
> 1/mean(x)  # Bootstrap estimate of lambda
[1] 0.06917423
> # repeat the procedure B times, save the B sample means in xbar and
sample sd in ss[]
> B=200; xbar = rep(0, B); ss = rep(0, B)
> for(i in 1:B) { x=sample(obs, n, replace=T); xbar[i] = mean(x)  }
> # the B bootstrap estimate of lambda are
> lambda.bt = 1/xbar
> summary(lambda.bt)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.02127 0.04039 0.04856 0.05142 0.05994 0.09881
> # bootstrap estimate of standard error
> sd(lambda.bt)
[1] 0.01565485
```

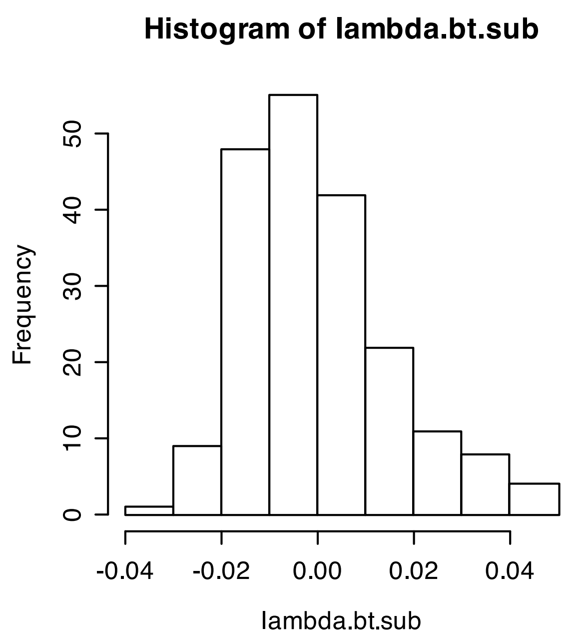
```
> # distribution of the bootstrap estimates
> stem(lambda.bt)

  The decimal point is 2 digit(s) to the left of the |

  2 | 1
  2 | 55789
  3 | 00112222333344444444
  3 | 55556666777888889999
  4 | 00000011111112222222233344444
  4 | 55556666666777788888888889999999
  5 | 00011112222233333333444444444
  5 | 6666667888889
  6 | 000000122234
  6 | 566677789
  7 | 000001122234
  7 | 57
  8 | 01124
  8 | 56667
  9 | 034
  9 | 89

> hist(lambda.bt)
> lambda.bt.sub = lambda.bt-mean(lambda.bt)
> hist(lambda.bt.sub)
```






```
> # 5 and 95 percentiles of lambda.bt-mean(lambda.bt)
> quantile(lambda.bt.sub, c(.05, .95) )
         5%         95%
-0.01940520  0.03217117
> # 90% bootstrap CI
> c(lambda-quantile(lambda.bt.sub,.95), lambda-
quantile(lambda.bt.sub,.05) )
       95%         5%
0.01402620 0.06560257
```

7-2.5 Bootstrap Estimate of the Standard Error (CD Only)

There are situations in which the standard error of the point estimator is unknown. Usually, these are cases where the form of $\hat{\Theta}$ is complicated, and the standard expectation and variance operators are difficult to apply. A computer-intensive technique called the **bootstrap** that was developed in recent years can be used for this problem.

Suppose that we are sampling from a population that can be modeled by the probability distribution $f(x; \theta)$. The random sample results in data values $x_1, x_2, \ldots, x_n$ and we obtain $\hat{\theta}$ as the point estimate of θ. We would now use a computer to obtain *bootstrap samples* from the distribution $f(x; \hat{\theta})$, and for each of these samples we calculate the bootstrap estimate $\hat{\theta}^*$ of θ. This results in

Bootstrap Sample	Observations	Bootstrap Estimate
1	$x_1^*, x_2^*, \ldots, x_n^*$	$\hat{\theta}_1^*$
2	$x_1^*, x_2^*, \ldots, x_n^*$	$\hat{\theta}_2^*$
$\vdots$	$\vdots$	$\vdots$
B	$x_1^*, x_2^*, \ldots, x_n^*$	$\hat{\theta}_B^*$

Usually $B = 100$ or 200 of these bootstrap samples are taken. Let $\bar{\theta}^* = (1/B)\sum_{i=1}^{B}\hat{\theta}_i^*$ be the sample mean of the bootstrap estimates. The bootstrap estimate of the standard error of $\hat{\Theta}$ is just the sample standard deviation of the $\hat{\theta}_i^*$, or

$$s_{\hat{\Theta}} = \sqrt{\frac{\sum_{i=1}^{B}(\hat{\theta}_i^* - \bar{\theta}^*)^2}{B - 1}} \tag{S7-1}$$

In the bootstrap literature, $B - 1$ in Equation S7-1 is often replaced by B. However, for the large values usually employed for B, there is little difference in the estimate produced for $s_{\hat{\Theta}}$.

EXAMPLE S7-1 The time to failure of an electronic module used in an automobile engine controller is tested at an elevated temperature in order to accelerate the failure mechanism. The time to failure is exponentially distributed with unknown parameter λ. Eight units are selected at random and tested, with the resulting failure times (in hours): $x_1 = 11.96$, $x_2 = 5.03$, $x_3 = 67.40$, $x_4 = 16.07$, $x_5 = 31.50$, $x_6 = 7.73$, $x_7 = 11.10$, and $x_8 = 22.38$. Now the mean of an exponential distribution is $\mu = 1/\lambda$, so $E(X) = 1/\lambda$, and the expected value of the sample average is $E(\bar{X}) = 1/\lambda$. Therefore, a reasonable way to estimate λ is with $\hat{\lambda} = 1/\bar{X}$. For our sample, $\bar{x} = 21.65$, so our estimate of λ is $\hat{\lambda} = 1/21.65 = 0.0462$. To find the bootstrap standard error we would now obtain $B = 200$ (say) samples of $n = 8$ observations each from an exponential distribution with parameter $\lambda = 0.0462$. The following table shows some of these results:

Bootstrap Sample	Observations	Bootstrap Estimate
1	8.01, 28.85, 14.14, 59.12, 3.11, 32.19, 5.26, 14.17	$\hat{\lambda}_1^* = 0.0485$
2	33.27, 2.10, 40.17, 32.43, 6.94, 30.66, 18.99, 5.61	$\hat{\lambda}_2^* = 0.0470$
$\vdots$	$\vdots$	$\vdots$
200	40.26, 39.26, 19.59, 43.53, 9.55, 7.07, 6.03, 8.94	$\hat{\lambda}_{200}^* = 0.0459$

The sample average of the $\hat{\lambda}_i^*$ (the bootstrap estimates) is 0.0513, and the standard deviation of these bootstrap estimates is 0.020. Therefore, the bootstrap standard error of $\hat{\lambda}$ is 0.020. In this case, estimating the parameter λ in an exponential distribution, the variance of the estimator we used, $\hat{\lambda}$, is known. When n is large, $V(\hat{\lambda}) = \lambda^2/n$. Therefore the estimated standard error of $\hat{\lambda}$ is $\sqrt{\hat{\lambda}^2/n} = \sqrt{(0.0462)^2/8} = 0.016$. Notice that this result agrees reasonably closely with the bootstrap standard error.

Sometimes we want to use the bootstrap in situations in which the form of the probability distribution is unknown. In these cases, we take the n observations in the sample as the *population* and select B random samples each of size n, with replacement, from this population. Then Equation S7-1 can be applied as described above. The book by Efron and Tibshirani (1993) is an excellent introduction to the bootstrap.

7-3.3 Bayesian Estimation of Parameters (CD Only)

This book uses methods of statistical inference based on the information in the sample data. In effect, these methods interpret probabilities as relative frequencies. Sometimes we call probabilities that are interpreted in this manner **objective probabilities.** There is another approach to statistical inference, called the **Bayesian approach,** that combines sample information with other information that may be available prior to collecting the sample. In this section we briefly illustrate how this approach may be used in parameter estimation.

Suppose that the random variable X has a probability distribution that is a function of one parameter θ. We will write this probability distribution as $f(x \mid \theta)$. This notation implies that the exact form of the distribution of X is conditional on the value assigned to θ. The classical approach to estimation would consist of taking a random sample of size n from this distribution and then substituting the sample values x_i into the estimator for θ. This estimator could have been developed using the maximum likelihood approach, for example.

Suppose that we have some additional information about θ and that we can summarize that information in the form of a probability distribution for θ, say, $f(\theta)$. This probability distribution is often called the **prior distribution** for θ, and suppose that the mean of the prior is μ_0 and the variance is σ_0^2. This is a very novel concept insofar as the rest of this book is concerned because we are now viewing the parameter θ as a random variable. The probabilities associated with the prior distribution are often called **subjective probabilities,** in that they usually reflect the analyst's degree of belief regarding the true value of θ. The Bayesian approach to estimation uses the prior distribution for θ, $f(\theta)$, and the joint probability distribution of the sample, say $f(x_1, x_2, \ldots, x_n \mid \theta)$, to find a **posterior distribution** for θ, say, $f(\theta \mid x_1, x_2, \ldots, x_n)$. This posterior distribution contains information both from the sample and the prior distribution for θ. In a sense, it expresses our degree of belief regarding the true value of θ after observing the sample data. It is easy conceptually to find the posterior distribution. The joint probability distribution of the sample $X_1, X_2, \ldots, X_n$ and the parameter θ (remember that θ is a random variable) is

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n \mid \theta) f(\theta)$$

and the marginal distribution of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \sum_{\theta} f(x_1, x_2, \ldots, x_n, \theta), & \theta \text{ discrete} \\ \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n, \theta)\, d\theta, & \theta \text{ continuous} \end{cases}$$

8-2.6 Bootstrap Confidence Intervals (CD Only)

In Section 7-2.5 we showed how a technique called the **bootstrap** could be used to estimate the standard error $\sigma_{\hat{\theta}}$, where $\hat{\theta}$ is an estimate of a parameter θ. We can also use the bootstrap to find a confidence interval on the parameter θ. To illustrate, consider the case where θ is the mean μ of a normal distribution with σ known. Now the estimator of θ is $\overline{X}$. Also notice that $z_{\alpha/2}\sigma/\sqrt{n}$ is the $100(1-\alpha/2)$ percentile of the distribution of $\overline{X}-\mu$, and $-z_{\alpha/2}\sigma/\sqrt{n}$ is the $100(\alpha/2)$ percentile of this distribution. Therefore, we can write the probability statement associated with the $100(1-\alpha)\%$ confidence interval as

$$P(100(\alpha/2) \text{ percentile} \le \overline{X}-\mu \le 100(1-\alpha/2) \text{ percentile}) = 1-\alpha$$

or

$$P(\overline{X}-100(1-\alpha/2) \text{ percentile} \le \mu \le \overline{X}-100(\alpha/2) \text{ percentile}) = 1-\alpha$$

This last probability statement implies that the lower and upper $100(1-\alpha)\%$ confidence limits for μ are

$$L = \overline{X} - 100(1-\alpha/2) \text{ percentile of } \overline{X}-\mu = \overline{X} - z_{\alpha/2}\sigma/\sqrt{n}$$
$$U = \overline{X} - 100(\alpha/2) \text{ percentile of } \overline{X}-\mu = \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}$$

We may generalize this to an arbitrary parameter θ. The $100(1-\alpha)\%$ confidence limits for θ are

$$L = \hat{\theta} - 100(1-\alpha/2) \text{ percentile of } \hat{\theta}-\theta$$
$$U = \hat{\theta} - 100(\alpha/2) \text{ percentile of } \hat{\theta}-\theta$$

Unfortunately, the percentiles of $\hat{\theta}-\theta$ may not be as easy to find as in the case of the normal distribution mean. However, they could be estimated from **bootstrap samples.** Suppose we find B bootstrap samples and calculate $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ and $\overline{\theta}^*$ and then calculate $\hat{\theta}_1^*-\overline{\theta}^*$, $\hat{\theta}_2^*-\overline{\theta}^*, \ldots, \hat{\theta}_B^*-\overline{\theta}^*$. The required percentiles can be obtained directly from the differences. For example, if $B = 200$ and a 95% confidence interval on θ is desired, the fifth smallest and fifth largest of the differences $\hat{\theta}_i^*-\overline{\theta}^*$ are the estimates of the necessary percentiles.

We will illustrate this procedure using the situation first described in Example 7-3, involving the parameter λ of an exponential distribution. Following that example, a random sample of $n = 8$ engine controller modules were tested to failure, and the estimate of λ obtained was $\hat{\lambda} = 0.0462$, where $\hat{\lambda} = 1/\overline{X}$ is a maximum likelihood estimator. We used 200 bootstrap samples to obtain an estimate of the standard error for $\hat{\lambda}$.

Figure S8-1(a) is a histogram of the 200 bootstrap estimates $\hat{\lambda}_i^*$, $i = 1, 2, \ldots, 200$. Notice that the histogram is not symmetrical and is skewed to the right, indicating that the sampling distribution of $\hat{\lambda}$ also has this same shape. We subtracted the sample average of these bootstrap estimates $\overline{\lambda}^* = 0.5013$ from each $\hat{\lambda}_i^*$. The histogram of the differences $\hat{\lambda}_i^*-\overline{\lambda}^*$, $i = 1, 2, \ldots, 200$, is shown in Figure S8-1(b). Suppose we wish to find a 90% confidence interval for λ. Now the fifth percentile of the bootstrap samples $\hat{\lambda}_i^*-\overline{\lambda}^*$ is -0.0228 and the ninety-fifth percentile is 0.03135. Therefore the lower and upper 90% bootstrap confidence limits are

$$L = \hat{\lambda} - 95 \text{ percentile of } \hat{\lambda}_i^* - \overline{\lambda}^* = 0.0462 - 0.03135 = 0.0149$$
$$U = \hat{\lambda} - 5 \text{ percentile of } \hat{\lambda}_i^* - \overline{\lambda}^* = 0.0462 - (-0.0228) = 0.0690$$

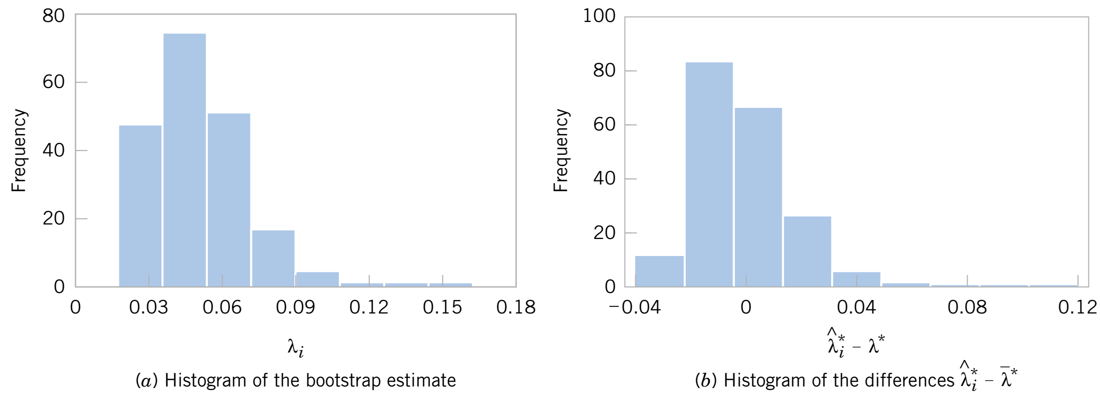


Figure S8-1 Histograms of the bootstrap estimates of λ and the differences $\hat{\lambda}_i^* - \hat{\lambda}^*$ used in finding the bootstrap confidence interval.

Therefore, our 90% bootstrap confidence interval for λ is $0.0149 \le \lambda \le 0.0690$. There is an exact confidence interval for the parameter λ in an exponential distribution. For the engine controller failure data following Example 7-3, the exact 90% confidence interval* for λ is $0.0230 \le \lambda \le 0.0759$. Notice that the two confidence intervals are very similar. The length of the exact confidence interval is $0.0759 - 0.0230 = 0.0529$, while the length of the bootstrap confidence interval is $0.0690 - 0.0149 = 0.0541$, which is only slightly longer. The percentile method for bootstrap confidence intervals works well when the estimator is unbiased and the standard error of $\hat{\theta}$ is approximately constant (as a function of θ). An improvement, known as the *bias-corrected and accelerated* method, adjusts the percentiles in more general cases. It could be applied in this example (because $\hat{\lambda}$ is a biased estimator), but at the cost of additional complexity.

8-3.2 Development of the *t*-Distribution (CD Only)

We will give a formal development of the *t*-distribution using the techniques presented in Section 5-8. It will be helpful to review that material before reading this section.

First consider the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

This quantity can be written as

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma\sqrt{n}}}{\sqrt{S^2/\sigma^2}} \tag{S8-1}$$

*The confidence interval is $\chi^2_{\alpha/2,2n}/(2\sum x_i) \le \lambda \le \chi^2_{1-\alpha/2,2n}/(2\sum x_i)$ where $\chi^2_{\alpha/2,2n}$ and $\chi^2_{1-\alpha/2,2n}$ are the lower and upper $\alpha/2$ percentage points of the chi-square distribution (which was introduced briefly in Chapter 4 and discussed further in Section 8-4), and the x_i are the n sample observations.