9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Lecture 11: Monte Carlo and Bootstrap Methods

I. Objectives

Understand how Monte Carlo methods are used in statistics.

Understand how to apply properly parametric and nonparametric bootstrap methods.

Understand why the bootstrap works.

"John Tukey suggested that the bootstrap be named the shotgun, reasoning that it could blow the head off any statistical problem if you were willing to tolerate the mess."
Paraphrased from Efron (1979).

"Dad, I just freed my mind!"
Elena Brown, 5 years old, explaining to me the secret of how she got control of her skis and successfully negotiated the bottom third of the ski slope on her first run of the new ski season.

II. Monte Carlo Methods

Computer simulations are useful to give insight into complicated problems when detailed analytic studies are not possible. We see this all the time in physics, chemistry, climatology and ecology, just to cite a few examples. Because by definition, stochastic simulations introduce randomness in to the study, they are often referred to as **Monte Carlo** methods after the capital city of the principality of Monaco, which is known for gambling. We have already made use of Monte Carlo methods to check the theoretical properties of some of our statistical procedures such as confidence intervals in **Lecture 8**, and to compute posterior probability densities in our Bayesian analyses in **Lecture 10** and in **Homework Assignment 7**.

In this lecture, we will present a very general, computer-intensive method called the **bootstrap** to compute from a random sample an estimate of a particular function of interest and the uncertainty in that function. The bootstrap was first proposed by Brad Efron at Stanford in 1979 (Efron, 1979). Since then it has become one of the most widely-used statistical methods. As has often been the case for Monte Carlo methods in general, it has made it possible to solve easily a number of estimation problems that would not have been possible or certainly far more difficult by other means. We begin by reviewing two elementary Monte Carlo methods.

A. Inverse Transformation Method. Before beginning with the bootstrap, we re-present one of the most basic Monte Carlo algorithms for simulating draws from a probability distribution. A *cdf* F outputs a number between 0 and 1. Therefore if F is continuous and if we take U as a uniform random variable on $(0,1)$, then we can define

$$X = F^{-1}(U) \tag{11.1}$$

where $F^{-1}(u)$ equals x whenever $F(x) = u.$ Because $F(x)$ is continuous and monotonic, it follows that we can simulate a random variable X from the continuous cdf $F,$ whenever F^{-1} is computable by simulating a random number $U,$ then setting $X = F^{-1}(U).$ These observations imply the following algorithm known as the **Inverse Transformation Method**.

Algorithm 11.1 (Algorithm 3.1, Inverse Transformation Method)
Pick n large.

1. For $i = 1$ to n.
2. Draw U_i from $U(0,1)$ and compute $X_i = F^{-1}(U_i)$
3. The observations $(X_1,...,X_n)$ are a random sample from F.

Example 3.2. Simulating Exponential Random Variables (Revisited). In **Lecture 3**, we defined the exponential probability density as

$$f(x) = \lambda e^{-\lambda x} \tag{11.2}$$

for $x > 0$ and $\lambda > 0$. It follows that the *cdf* $F(x)$ is

$$F(x) = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}. \tag{11.3}$$

To find $F^{-1}(u)$ we note that

$$1 - e^{-\lambda x} = u \tag{11.4}$$

$$x = -\frac{\log(1-u)}{\lambda} \tag{11.5}$$

Hence, if u is uniform on $(0,1)$ then

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda} \tag{11.6}$$

is exponentially distributed with mean λ^{-1}. Since $1-u$ is uniform on $(0,1)$, it follows that $x = -\frac{\log(u)}{\lambda}$ is exponential with mean λ^{-1}. Therefore, we can use **Algorithm 3.1** to simulate draws from an exponential distribution by using F^{-1} in Eq. 11.6.

B. Sampling from an Empirical Cumulative Distribution Function. The same idea we used to simulate draws from a continuous *cdf* may be used to simulate draws from a discontinuous or empirical *cdf*. That is, given $x_1,...,x_n$ define the **empirical cumulative distribution function** as

$$F_n(x) = \frac{\sum_{i=1}^{n} I\{x_i \le x\}}{n} \tag{11.7}$$

where $I\{x_i \le x\}$ is the indicator function of the event $x_i \le x$ for $x_{\min} \le x \le x_{\max}$. We have that $F_n(x)$ is the *cdf* corresponding to the empirical probability mass function that places weight $\frac{1}{n}$ at every observation. We can express the empirical *cdf* alternatively as

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \frac{k}{n} & x_{(k)} \le x < x_{(k+1)} \\ 1 & x_{(n)} < x \end{cases} \tag{11.8}$$

where the $x_{(i)}$ are the ordered observations or **order statistics** from $i = 1$ to n and $k = \sum_{\mathrm{i=1}}^{\mathrm{n}} I\{x_i \le x\}$. Remember we use the order statistics to construct the Q-Q plots. We define $F_n^{-1}(u)$ as

$$F_n^{-1}(u) = x_{(k)} \tag{11.9}$$

if $\frac{k}{n} \le u < \frac{k+1}{n}$ for $k = 1, \ldots, n-1.$

An example of an empirical *cdf* is shown in **Figure 11A** which shows the retinal ISI data from Example 8.1. There are 971 observations here so that the jump of $\frac{1}{n}$ at each order statistic corresponds to a jump of approximately 10^{-3}. For this reason, this empirical *cdf* looks very smooth.

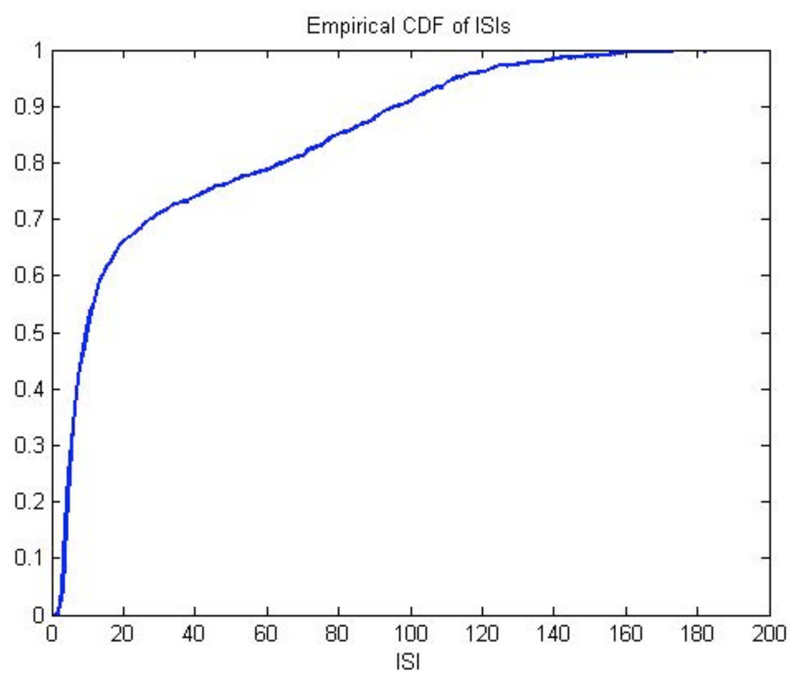

Figure 11A. Empirical Cumulative Distribution Function for Retinal ISI Data (Example 8.1).

Remark 11.1. $F_n(x)$ is an estimate of the unknown cumulative distribution function $F(x)$. This can be justified by the Law of Large Numbers (**Lecture 7**) and more strongly by another important result in probability theory called the **Glivenko-Cantelli Lemma.** This lemma says that as n goes to infinity $F_n(x)$ converges uniformly to the true $F(x)$ in distribution.

Note on Notation. In our lectures we have used θ to denote the parameter in our probability models. In this lecture we represent this parameter by ξ in order to let θ, as we show below define the generic function whose properties are characterized in a bootstrap analysis.

Remark 11.2. In **Lectures 8** to **10** we studied method-of-moments, likelihood and Bayesian methods for estimating the parameters of probability models. In each case, we wrote down a specific probability model to describe the data and used one of these three approaches to estimate the model parameters. These models, such as the binomial, Poisson, Gaussian, gamma, beta and inverse Gaussian are **parametric models** because given the formula for the probability distribution, each is completely defined once we specify the values of the model parameters. In this regard our method-of-moments, likelihood and Bayesian methods provided **parametric estimates** of the probability distributions. That is, given a probability distribution $f(x \mid \xi)$ the associated *cdf* is $F(x \mid \xi) = \int_{-\infty}^{x} f(u \mid \xi) du$. Hence, given $\hat{\xi}$ an estimate of the parameter ξ we have that $F(x \mid \hat{\xi}) = \int_{-\infty}^{x} f(u \mid \hat{\xi}) du$ is the parametric estimate of the *cdf* is $F(x \mid \hat{\xi})$. Therefore, given $F(x \mid \hat{\xi})$, we can use **Algorithm 3.1** to simulate estimated samples from $F(x \mid \xi)$. The type of optimality properties of $F(x \mid \hat{\xi})$ and the samples from it will follow from the optimality properties of the type of estimation procedure used to compute $\hat{\xi}$.

Remark 11.3. In contrast to $F(x \mid \hat{\xi})$ in **Remark 11.2**, we term the empirical cumulative distribution function $F_n(x)$ a **nonparametric estimate** of $F(x)$. This is because we can compute it without making any assumption about a specific parametric form of $F(x)$. Just as we can use **Algorithm 3.1** applied to $F(x \mid \hat{\xi})$, to simulate estimated samples from $F(x \mid \xi)$, we can use $F_n(x)$ to simulate samples from $F(x)$. The prescription is given in the next algorithm.

Algorithm 11.2 (Sampling from an Empirical Cumulative Distribution Function).
Pick n^*.

1. For $i = 1, ..., n^*$
2. Draw U_i from $U(0,1)$.
3. Compute $X_i^* = F_n^{-1}(U_i)$ in Eq. 11.9.

The sample $(X_1^*, ..., X_n^*)$ is a random sample from $F_n(x)$.

The parametric and nonparametric estimates of the *cdf* are the key tools we will need to construct the bootstrap procedures.

III. Bootstrap Methods (Efron and Tibshirani, 1993; DeGroot and Schervish, 2002; Rice 2007).
A. Motivation.
To motivate the bootstrap we return to two problems we considered previously.

Example 8.1 (continued). Retinal Neuron ISI Distribution: Method-of-Moments. In this problem, we have been studying probability models to describe the ISI distribution of a retinal neuron. One of the models we proposed for these data was the gamma distribution with unknown parameters α and β. We found that the method-of-moments estimates were

$$\hat{\beta}_{MM} = \bar{X}(\hat{\sigma}^2)^{-1} \tag{11.10}$$

$$\hat{\alpha}_{MM} = \bar{X}^2(\hat{\sigma}^2)^{-1}. \tag{11.11}$$

One of the points we made in **Lecture 8** was that while we could find approximations to the distributions of $\bar{X}$ using the central limit theorem and that of $\hat{\sigma}^2$ under the special assumption of Gaussian data, it was in general difficult to find the distribution of the sample moments and functions of these moments. However, this is exactly what we need in order to determine the uncertainty in the method-of-moments estimates of α and β in Eqs. 11.10 and 11.11.

Example 8.1. (continued). Retinal Neuron ISI Distribution: Five-Number Summary. One of the first things we did to analyze these data was to compute the five-number summary in **Lecture 8** (See Figure 8B). These are the minimum 2 msec, 25^{th} percentile 5 msec, median (50^{th} percentile) 10 msec, the 75^{th} percentile 43 msec and the maximum of 190 msec. Each of these is a statistic and each estimates respectively, the true minimum, 25^{th} percentile, median (50^{th} percentile), 75^{th} percentile, maximum. Hence, because each is a statistic, each has uncertainty. Each of these is a nonparametric estimate of the corresponding true function of the distribution. Suppose we wish to compute nonparametric estimates of the uncertainty in these statistics. None of the method-of-moments, likelihood or Bayesian theory we have developed would allow us to compute the uncertainty in these estimates. Indeed, the analytics necessary to study these estimates falls under the heading of the quite advanced areas of mathematical statistics called order statistics (Rice, 2007) and robust estimation theory (DeGroot and Schervish, 2002).

These two problems can be generally formulated as follows. Suppose we have a random sample $x = (x_1, ..., x_n)$ from an unknown probability distribution F and suppose we wish to estimate the quantity of interest $\theta = T(F(x))$. The quantity θ can be either a parameter of a parametric probability distribution, or a nonparametric quantity, such as one of the components of the five-number summary. Suppose we estimate θ as $\hat{\theta} = T(\hat{F}(x))$. How can we compute the uncertainty in $\hat{\theta}$?

B. The Nonparametric Bootstrap
Bootstrap methods depend on the concept of a bootstrap sample. Let $\hat{F}$ be an empirical *cdf* as defined in Eq. 11.8. It puts mass $1/n$ at every observation x_i for $i = 1, ..., n$. A **bootstrap sample**

is a random sample of size n drawn from $\hat{F}$ which we denote as $x^* = (x_1^*, ..., x_n^*)$. We use the * to indicate that x^* is not the actual data set x. Alternatively stated, the bootstrap sample $x_1^*, ..., x_n^*$ is a random sample of size n drawn with replacement from the set of n objects $x_1, ..., x_n$. The bootstrap sample $x_1^*, ..., x_n^*$ consists of members of the original sample some appearing, zero times, some appearing once, some appearing twice, etc. Given a bootstrap sample, we can compute a **bootstrap replication** of $\hat{\theta} = T(\hat{F}(x))$ defined as

$$\hat{\theta}^* = T(\hat{F}(x^*)) \tag{11.12}$$

The quantity in Eq. 11.12 comes about by applying the function T to the bootstrap sample. For example, for the method-of-moments in Eqs. 11.10 and 11.11

$$\hat{\beta}_{MM}^* = \bar{X}^*(\hat{\sigma}^{2*})^{-1} \tag{11.13}$$

$$\hat{\alpha}_{MM}^* = \bar{X}^{2*}(\hat{\sigma}^{2*}).^{-1} \tag{11.14}$$

In the case of the five-number summary, it would be the five-number summary computed at the bootstrap sample $x_1^*, ..., x_n^*$.

It follows that we can compute the uncertainty in $\hat{\theta} = T(\hat{F}(x))$ by simply drawing a large number of bootstrap samples, say B, from $\hat{F}$ and we compute $\hat{\theta}^* = T(\hat{F}(x^*))$ for each sample. A histogram of $\hat{\theta}^*$ shows the uncertainty in $\hat{\theta}$. We could also summarize the uncertainty in $\hat{\theta}$ by computing the standard error of the bootstrap replicates defined as

$$se_B(\hat{\theta}^*) = \left[(B-1)^{-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}^*)^2 \right]^{\frac{1}{2}} \tag{11.15}$$

where

$$\bar{\theta}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*. \tag{11.16}$$

Notice here we divide by $B-1$ instead of B because for many problems only a small number of bootstrap replicates is required to accurately establish the uncertainty. The quantity $\bar{\theta}^*$ in Eq. 11.16 is the bootstrap estimate. Dividing by $B-1$ provides an unbiased estimate of the standard error. We could also compute an empirical 95% confidence interval based on the histogram of the bootstrap replicates. We term the procedure of drawing random samples with replacement from the empirical *cdf* $\hat{F}$ to compute the uncertainty in $\hat{\theta}$ a **nonparametric bootstrap procedure**. The procedure is nonparametric because we use the nonparametric estimate of F defined in Eq. 11.8. We can state this in the following algorithm.

Algorithm 11.3 (Nonparametric Bootstrap)
Pick B.

1. Draw $x^{*1},...,x^{*B}$ independent bootstrap samples from $\hat{F}$, where $x^{*b} = (x_1^{*b},...,x_n^{*b})$ is a vector of n observations which contains the b^{th} bootstrap sample for $b = 1,...,B.$

2. Evaluate the bootstrap replication corresponding to each bootstrap sample $\hat{\theta}^*(b) = T(\hat{F}(x^{*b}))$ for $b = 1,...,B.$

3. Compute the uncertainty in $\hat{\theta}$ as either the:
   i. sample standard error as $se_B(\hat{\theta}^*)$ defined in 11.15
   ii. histogram of the $\hat{\theta}^*(b) = T(\hat{F}(x^{*b}))$
   iii. 95% confidence interval from the histogram.

Remark 11.4. Notice that in the case of the method-of-moments estimates of the parameters of the gamma distribution, even though the functions of interest are parameters, we are using a nonparametric bootstrap procedure to estimate their uncertainty.

C. The Parametric Bootstrap

Suppose as in **Remark 11.2** that given our random sample $x_1,...,x_n$ we estimate F not by the empirical cdf $F_n(x)$ but by a parametric estimate of the *cdf* $F(x \mid \hat{\xi})$ using a method-of-moments or likelihood approach. Therefore, it follows that we can carry out our bootstrap procedure using $F(x \mid \hat{\xi})$ instead of the empirical $F_n(x)$ as the estimate of F. We term this procedure of drawing random samples with replacement from the parametric *cdf* $\hat{F} = F(x \mid \hat{\xi})$ to compute the uncertainty in $\hat{\theta}$ a **parametric bootstrap procedure**. We can state this procedure formally in the following algorithm

Algorithm 11.4 (Parametric Bootstrap)
Pick $B.$

1. Draw $x^{*1},...,x^{*B}$ independent bootstrap samples from $\hat{F} = F(x \mid \hat{\xi}),$ where $x^{*b} = (x_1^{*b},...,x_n^{*b})$ is a vector of n observations which contains the b^{th} bootstrap sample for $b = 1,...,B.$

2. Evaluate the bootstrap replication corresponding to each bootstrap sample $\hat{\theta}^*(b) = T(\hat{F}(x^{*b}))$ for $b = 1,...,B.$

3. Compute the uncertainty in $\hat{\theta}$ as either the:
   i. sample standard error as $se_B(\hat{\theta}^*)$ defined in 11.15
   ii. histogram of the $\hat{\theta}^*(b) = T(\hat{F}(x^{*b}))$
   iii. 95% confidence interval from the histogram.

Remark 11.5. An important difference between the nonparametric and parametric bootstrap procedures is that in the nonparametric procedure, only values of the original sample appear in the bootstrap samples. In the parametric bootstrap, the range of values in the bootstrap sample is the entire support of $F.$ For example, in the case of the retinal ISI data, there are 971 values. Therefore a nonparametric bootstrap procedure applied to these data would have only 971 different values in any bootstrap sample. In the parametric bootstrap of this problem using the gamma distribution the values in the bootstrap sample could be any value between zero and positive infinity.

Remark 11.6. We can use the parametric bootstrap to analyze the uncertainty in the nonparametric estimators such as the five-number summary.

D. Applications of the Bootstrap

1. Example 8.1 (continued) Nonparametric and Parametric Bootstrap Analyses of the Method-of-Moments Estimates for the Gamma Distribution Parameters.

We used **Algorithms 11.3** and **11.4** to perform nonparametric and parametric bootstrap analyses in order to estimate the uncertainty in the method-method of-moments estimates for the gamma probability density parameters for the retinal ISI data. The histograms of the analyses are shown in **Figures 11B** and **11C** for $\hat{\alpha}_{MM}$ and in Figures 11D and 11E for $\hat{\beta}_{MM}$. The findings from the analysis are summarized in **Tables 11.1** for $\hat{\alpha}_{MM}$ and **11.2** for $\hat{\beta}_{MM}$.

	$\hat{\alpha}_{MM}$	$\hat{\alpha}_B$	$bias_{MM} = \hat{\alpha}_{MM} - \hat{\alpha}_B$	$se(\hat{\alpha}_B)$	**95% CI**
Nonparametric	**0.6284**	**0.6299**	**-0.0015**	**0.0281**	**(0.5749, 0.6849)**
Parametric	**0.6284**	**0.6324**	**-0.0040**	**0.0452**	**(0.5437, 0.7211)**

Table 11.1 Comparison of Nonparametric and Parametric Bootstrap Analyses of $\hat{\alpha}_{MM}$.

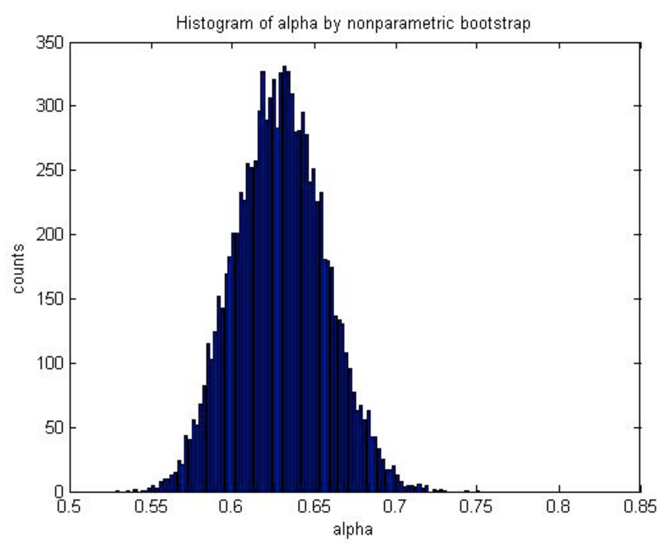

Figure 11B. Histogram of nonparametric bootstrap samples of $\hat{\alpha}_{MM}$.

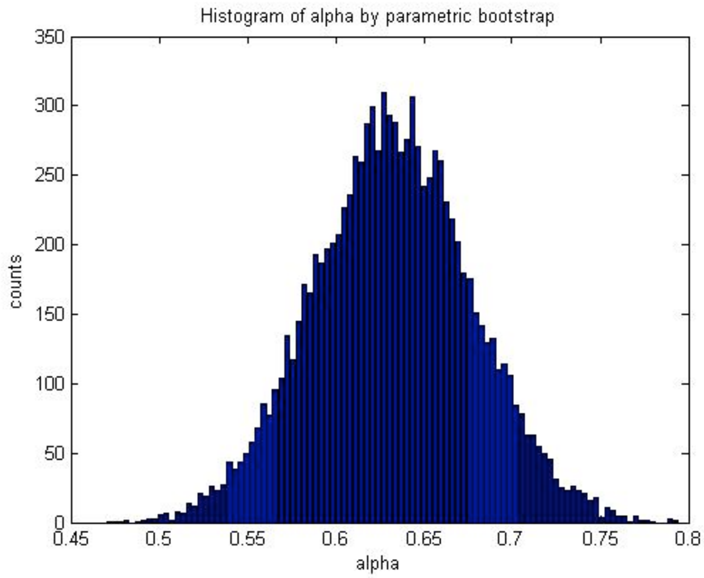

Figure 11C. Histogram of parametric bootstrap samples of $\hat{\alpha}_{MM}$.

The parametric and nonparametric bootstrap estimates of α agree and both procedures show little to no bias (**Table 11.1 column 3**). The parametric bootstrap procedure has a larger standard error and a wider 95% confidence interval (**Table 11.1 columns 4** and **5**) suggesting that there is greater uncertainty in the estimate of α than suggested by the nonparametric analysis.

	$\hat{\beta}_{MM}$	$\hat{\beta}_B$	$bias_{MM} = \hat{\beta}_{MM} - \hat{\beta}_B$	$se(\hat{\beta}_B)$	**95% CI**
Nonparametric	**0.0204**	**0.0205**	**0.0001**	**0.0007**	**(0.0191, 0.0218)**
Parametric	**0.0204**	**0.0206**	**0.0002**	**0.0017**	**(0.0173, 0.0239)**

Table 11.2 Comparison of Nonparametric and Parametric Bootstrap Analyses of $\hat{\beta}_{MM}$.

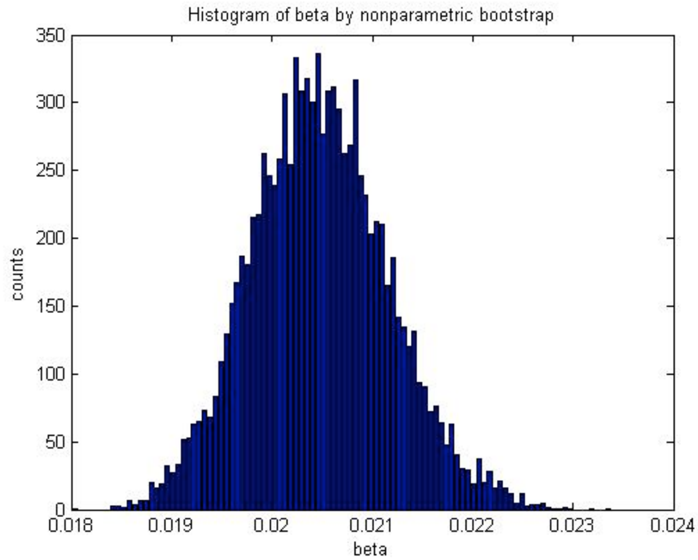

Figure 11D. Histogram of nonparametric bootstrap samples of $\hat{\beta}_{MM}$.

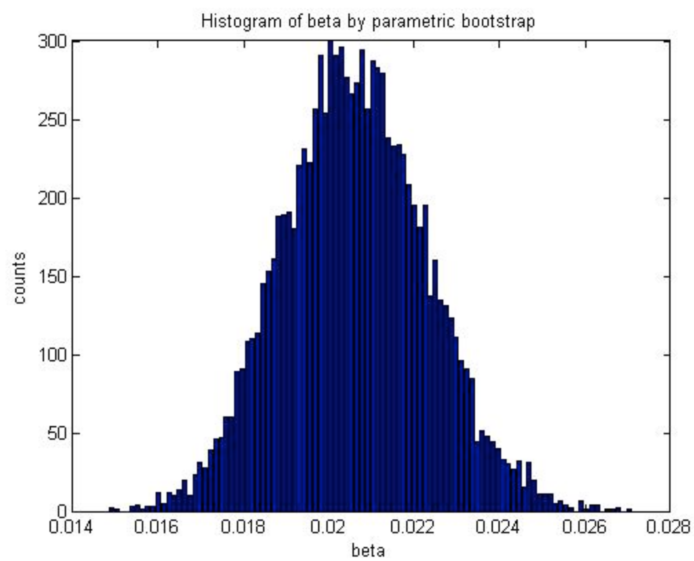

Figure 11E. Histogram of parametric bootstrap samples of $\hat{\beta}_{MM}$.

Similarly, the parametric and nonparametric bootstrap estimates of β agree (**Table 11.2**) and both procedures show little to no bias (**Table 11.2, column 3**). The parametric bootstrap here as well suggests that there is greater uncertainty in the estimate of β than suggested by the nonparametric analysis (**Table 11.2, column 4** and **5**).

2. Example 8.1 (continued) Nonparametric Bootstrap of the 75^th^ Percentile.

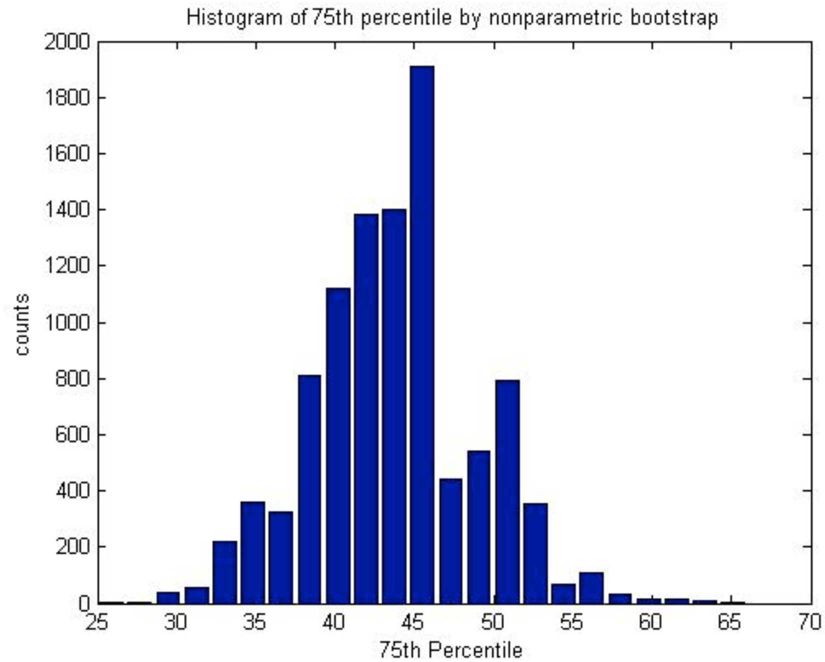

Figure 11F . Histogram of Nonparametric Bootstrap samples of 75th Percentile.

	$75th\ Estimate$	$75th_B$	$bias_{75th} = 75th - 75th_B$	$se(75th_B)$	**95% CI**
Nonparametric	**43**	**43.73**	**0.73**	**5.12**	**(33.69 53.77)**

Table 11.3 Nonparametric Bootstrap Analyses of 75th Percentile.

We used **Algorithms 11.3** to perform nonparametric bootstrap analysis of the uncertainty in the empirical (order statistic) estimate of the 75th percentile. The histogram of the analyses is shown in **Figure 11F** and the findings from the analysis are summarized in **Table 11.3**. The order statistic and bootstrap estimates agree indicating that there is very little bias. The standard error and 95% confidence intervals suggest that the true 75th percentile could be as small as 34 msec or as large as 54 msec. This analysis shows one of the important features of the bootstrap namely, except for the assumption of independence of the observations, we need to make no assumptions about the specific probability model that generated the data.

E. A Heuristic Look at the Bootstrap Theory

While the bootstrap is a very simple to apply, computer-intensive estimation procedure, the theory underlying it has been a subject of much investigation by some of the best minds in modern statistics. See the references in Efron and Tibshirani (1993) and DeGroot and Schervish (2002). The essential idea behind why the procedure gives a "correct" assessment of uncertainty has two components. First, for the nonparametric bootstrap, by the Glivenko-Cantelli Lemma, $F_n(x)$ converges in distribution to the unknown cumulative distribution function F. If an observed sample has n observations, then there are $\binom{2n-1}{n} = \frac{(2n-1)!}{(n-1)!n!}$ distinct bootstrap samples. As $B \to \infty$, the bootstrap samples converge to the population of distinct bootstrap samples. Therefore, if $F_n(x^*)$ converge $F_n(x)$ and $F_n(x)$ converges to F, then it follows that $\hat{\theta}^* = T(\hat{F}(x^*))$ converges to $\theta = T(F(g))$.

A similar argument holds for the parametric bootstrap except for the fact that $F(x \mid \hat{\xi})$ converges to $F(x \mid \xi)$ follows from the large sample properties of $\hat{\xi}$. That is, $\hat{\xi}$ converges to ξ as the sample size becomes large due to the optimality properties of the procedure used to estimate ξ.Therefore, since $F(x \mid \xi)$ is a usually well-behaved function of ξ it generally follows that $F(x \mid \hat{\xi})$ converges to $F(x \mid \xi)$. The balance of the argument for the nonparametric bootstrap is then applied and a similar conclusion about the convergence of the bootstrap procedure follows in this case as well.

Remark 11.7. The above argument shows that the bootstrap is a frequentist method. Like the standard confidence intervals, the bootstrap is justified based on its long-run properties over repeated sampling.

Remark 11.8. A key assumption in the bootstrap analysis is that the original sample is independent and identically distributed, i.e., every observation is drawn from the same distribution function F. When this is the case, the bootstrap samples can be drawn as stated above in **Algorithms 11.3** and **11.4.** In many common problems in neuroscience, such as analyses of neural spike trains, regression problems, and EEG and local field potential analyses, these assumptions clearly do not hold. For example, in the standard regression problem as we will study shortly, every observation is Gaussian with the same variance yet with a difference mean given by the regression function. Neural spike trains, EEG and local field potential data are never collections of independent observations. Therefore, to apply the bootstrap to any of these commonly encountered problems, special care must be taken to implement the procedure correctly. We will investigate some of these issues when we study linear model, point processes, time-series and spectral methods.

Remark 11.9. The bootstrap is a computer-intensive estimation procedure. The quantity $\bar{\theta}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$ is the bootstrap estimate of θ. It follows that $b_{\hat{\theta}} = (\bar{\theta}^* - \hat{\theta})$ provides an estimate of the bias in the estimation procedure that produced $\hat{\theta}$. Indeed, the bootstrap can be used in this way to produce a bias correction for a procedure. This was the original intended purpose of the jackknife, a resampling method that was the predecessor of the bootstrap.

Remark 11.10. A very appealing feature of the bootstrap is that very often, few bootstrap replicates are needed to accurately assess the uncertainty in the statistic of interest. For example, Efron and Tibshirani (1993) report that only as few as 25 to 200 bootstrap samples may be required to accurately compute the bootstrap estimate of the standard error $se_B(\hat{\theta}^*)$ in Eq. 11.15 for a given parameter of interest. Larger numbers of samples may be necessary to compute *p*-values and a histogram estimate of a probability density.

Remark 11.11. One challenge to applying the bootstrap, though not an insurmountable one, is that it does require reestimating $\hat{\theta}^*(b) = T(\hat{F}(x^{*b}))$ B times. When $\hat{\theta} = T(F(x))$ exists in closed form, such as in the case of the method-of-moments estimates, this is easy to do. However, the maximum likelihood estimates of α and β for the gamma distribution do not exist in closed form. Here, a Newton's procedure would have to be applied to each bootstrap sample to obtain the bootstrap replicates.

IV. Summary

The bootstrap is a widely used, computer-intensive method for estimating uncertainty in statistics of interest. It has made it possible to attack a wide range of problems that would have been otherwise intractable. We will use it extensively in our analyses in later lectures.

Acknowledgments

I am grateful to Uri Eden for making the figures, to Julie Scott for technical assistance and to Jim Mutch for helpful proofreading and comments.

Text References

DeGroot MH, Schervish MJ. *Probability and Statistics*, 3rd edition. Boston, MA: Addison Wesley, 2002.

Efron B, Tibshirani R. *An Introduction to the Bootstrap*. London: Chapman and Hall, 1993.

Rice JA. *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.

Ross SM. *Introduction to Probability Models*, 5th edition. Boston: Academic Press, Inc., 1993.

Literature Reference

Efron, B. Bootstrap methods: another look at the jackknife. *Ann. Statist*., 7: 1-26, 1979.