Lecture 4, Part I: Using the Bootstrap to Estimate the Uncertainty of a Statistic

Text references: Chapter 6 in Shalizi

Big Picture: The bootstrap is a technique for approximating the uncertainty (e.g. bias, variance and confidence sets) of estimators using simulations. It can be applied in a wide range of situations.

What is the source of statistical uncertainty?

Well, suppose we re-run the experiment (survey, census,...) and get different data. Then, everything we calculate from data (estimates, test statistics, policies,....) will change from trial to trial as well. This variability is the source of **statistical uncertainty**. Quantifying this uncertainty is a way to be honest about what we actually know.

General setting (point estimation)

In your previous classes, you have used at least two main approaches to estimating standard errors:

1. Direct calculation where we replace unknowns with their best estimates.

2. Large-sample approximations.

In many other cases, however, we can't even come up with an expression for $\mathbb{V}(T)$:

The Bootstrap Idea: *If* it were possible to obtain/simulate additional data sets (*each of size* n), we could use these to approximate the variance and, more generally, the sampling distribution of T:

The sample variance of $T_1^*, T_2^*, \ldots, T_B^*$ is an estimator for $\mathbb{V}(T)$, i.e., the bootstrap estimate $\hat{\mathsf{se}}_{\mathrm{boot}}$ of the standard error is defined as:

Similarly, to approximate the **sampling distribution** of T or to construct confidence intervals, we can histogram the bootstrap replications or compute sample quantiles. However....

Of course, we cannot actually draw from F, because we don't know F. Instead, we make the draws from an **estimate** of F. That is, to approximate the sampling distribution, we simulate data and treat the simulated data just like real data.

There are two main versions of the bootstrap:

I. Parametric Bootstrap

Here F is assumed to belong to some parametric family, e.g. F is $Normal(\mu, \sigma^2)$. So, we write F_θ to indicate that F is fully specified by the value of θ. In the parametric bootstrap: if F_θ depends on a parameter θ and $\widehat{\theta}$ is an estimate of θ, then we simply sample from $F_{\widehat{\theta}}$ instead of F . This is just as accurate, but much simpler than, the delta method. Here is more detail.

Suppose that $X_1, ..., X_n \sim f(x; \theta)$. Let $\widehat{\theta}$ be the mle. Let $\tau = g(\theta)$. Then $\widehat{\tau} = g(\widehat{\theta})$. To get the standard error of $\widehat{\tau}$ we need to compute the Fisher information and then apply the delta method. The bootstrap allows us to avoid both steps. We just do the following:

Example 1. Suppose that $X_1, X_2, \ldots, X_n$ are modeled as iid Gamma(α, β). The observed data are

```
13.79, 8.52, 7.10, 10.82, 5.35, 11.06, 11.91, 11.36,
8.58, 7.20, 7.68, 8.60, 9.36, 6.95, 12.20, 11.42,
12.54, 12.06, 9.73, 4.36
```

Report an estimate of α, along with an (approximate) standard error for that estimator.

II. Nonparametric Bootstrap (often just called the "Bootstrap")

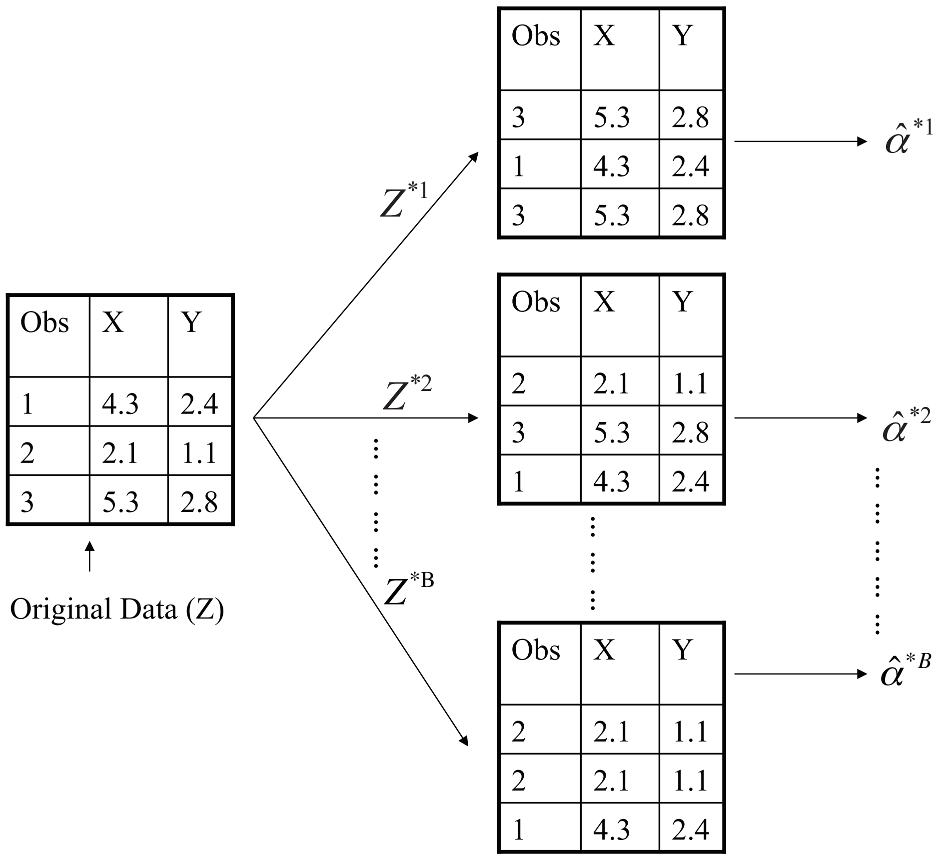

Figure 1: A graphical illustration of the bootstrap approach on a small sample containing $n = 3$ observations. Each bootstrap data set (Z^{*b}, $b = 1, \ldots, B$) contains n observations, sampled with replacement from the original data set (Z). Each bootstrap data set is used to obtain an estimate of α, denoted by $\hat{\alpha}^{*b}$, $b = 1, \ldots, B$. [From "An Introduction to Statistical Learning" by James et al.]

This is the classic version most people refer to when they say "bootstrap". Suppose we are unwilling to assume F is of some parametric family. How do we then approximate simulating from F? Well, in bootstrap, we simulate from the *empirical distribution function* $\hat{F}$ that gives probability $1/n$ to each data point in the original sample; that is

Q: How good is the approximation of the variance of T, the statistic of interest? What are the sources of error?

Example 2. Suppose that $X_1, X_2, \ldots, X_n$ are modeled as iid, but with the distribution unspecified/unknown. The observed data are

```
8.13, 8.37, 17.85, 23.12, 29.10, 2.86, 3.89, 0.52, 3.52,
15.65, 8.15, 27.59, 3.86, 10.08, 8.19, 0.59, 22.84, 8.04,
15.84, 12.34
```

Report an estimate of $\mathbb{V}(X_i)$, along with an (approximate) standard error for that estimator.

Example 3. (Robustness of median versus mean estimates).

We have 26 measurements of the heat of sublimation of platinum. There are 5 outliers.

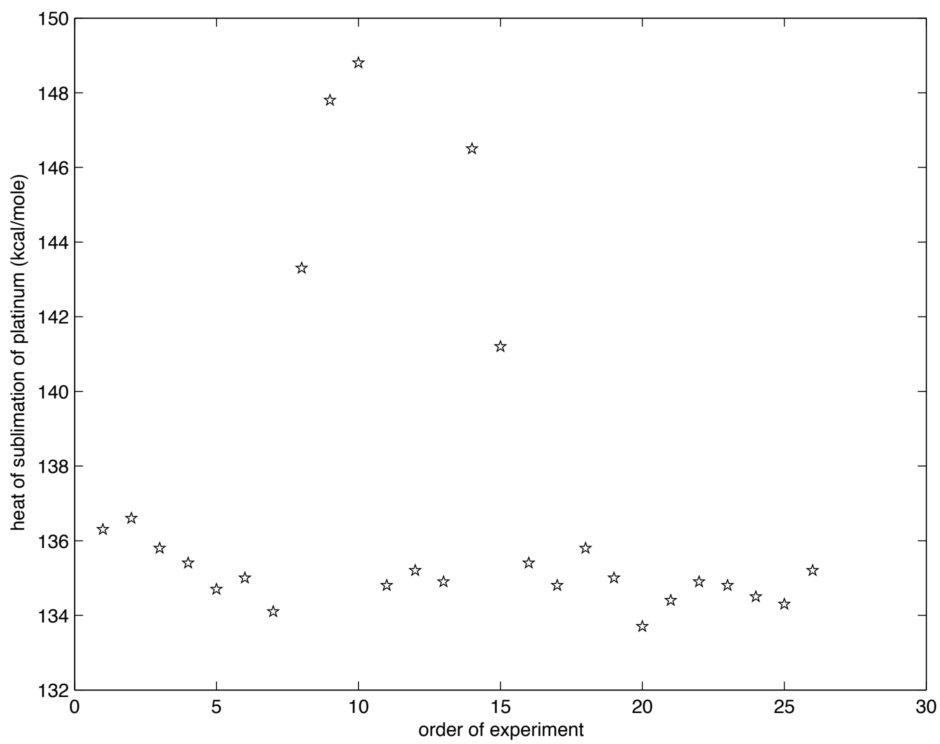

Figure 2: Measurements of the heat of sublimation of platinum

Compute the sample mean and the sample median. Use bootstrap to estimate the sampling distributions of these two summary statistics. What differences do you see in the estimated distributions?

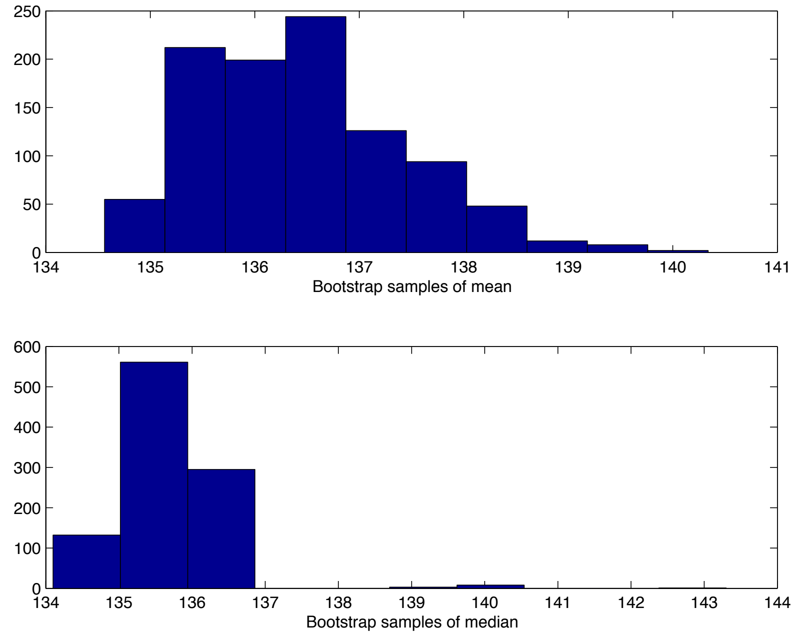

Figure 3: Histograms of bootstrap replications of the sample mean (top) and the sample median (bottom)

Estimating Bias

We can also use the bootstrap to estimate the bias of our estimator. That $\hat{\mathsf{se}}_{\mathrm{boot}}$, defined above, is a reasonable approximation of standard error is more or less very intuitive, but the bias argument is not as obvious. The idea is to make the following two approximations:

Once we believe the first approximation (1), the second approximation (2) clearly follows. But why should (1) be reasonable? It will remain a valid approximation as long as the distributions of $\hat{\theta} - \theta$ and $\hat{\theta}^* - \hat{\theta}$ are close. This is weaker than saying that the distributions of $\hat{\theta}$ and $\hat{\theta}^*$ should be close, or that $\mathbb{E}(\hat{\theta})$ and θ should be close.

More generally, you may consider (1) to be a reasonable approximation as long as $\hat{\theta} - \theta$ is (roughly) *pivotal*, meaning that its distribution does not depend on the unknown parameter θ.

Bootstrap Confidence Intervals

An extremely useful application of the bootstrap is the construction of *confidence intervals*. Recall that a $(1-\alpha)$ confidence interval for θ, computed for $X_1, \ldots X_n$, is defined as:

We stress that the lower and upper limits L and U are **random** (i.e., L and U depend on the data $X_1, \dots X_n$), and it is this randomness that is being considered in the probability statement above — the underlying parameter θ itself is fixed.

The *basic bootstrap confidence interval* for θ computes the bootstrap statistics $\hat{\theta}_1^*, \dots \hat{\theta}_B^*$ as above, and then approximates the distribution of $\hat{\theta} - \theta$ by $\hat{\theta}^* - \hat{\theta}$

That is, we compute the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\hat{\theta}_1^*, \dots \hat{\theta}_B^*$, call them $q_{\alpha/2}$ and $q_{1-\alpha/2}$, and then argue that

In other words, an approximate $(1-\alpha)$ bootstrap confidence interval for θ is given by:

Code for Example 1: Parametric bootstrap, estimating α

```
# How many bootstrap replications to perform

B = 10000

# The entire bootstrap process will be repeated reps times,
# in order to assess the amount of variability

reps = 5

# The observed data

x = c(13.79, 8.52, 7.10, 10.82, 5.35, 11.06, 11.91, 11.36, 8.58, 7.20,
      7.68, 8.60, 9.36, 6.95, 12.20, 11.42, 12.54, 12.06, 9.73, 4.36)

# The estimates of alpha and beta

n = length(x)
alphahat = sum(x)^2/(n*sum(x^2)-sum(x)^2)
betahat = (n*sum(x^2)-sum(x)^2)/(n*sum(x))

print(alphahat)

# The outer loop is over the replications of the entire process
for(j in 1:reps)
{

# Initialize alphahatstar, which will hold the bootstrap
# estimates of alpha

   alphahatstar = rep(0,B)
```

```
# This is the bootstrap loop

   for(i in 1:B)
   {

# Choose from the distribution and calculate alphahatstar
      xstar = rgamma(n,alphahat,betahat)
      alphahatstar[i] = sum(xstar)^2/(n*sum(xstar^2)-sum(xstar)^2)
   }
   print(sd(alphahatstar))
}
```

The Output:

```
> print(alphahat)
[1] 14.43762

> print(sd(alphahatstar))
[1] 6.261482
[1] 6.443681
[1] 6.285241
[1] 6.272167
[1] 6.121731
```

Code for Example 2: Nonparametric bootstrap, estimating $\mathbb{V}(X_i)$

```
# How many bootstrap replications to perform

B = 10000

# The entire bootstrap process will be repeated reps times,
# in order to assess the amount of variability

reps = 5

# The observed data

x = c(8.13, 8.37, 17.85, 23.12, 29.10, 2.86, 3.89, 0.52, 3.52, 15.65, 8.15,
      27.59, 3.86, 10.08, 8.19, 0.59, 22.84, 8.04, 15.84, 12.34)
```

```
n = length(x)

# The sample variance

print(var(x))

# The outer loop is over the replications of the entire process

for(j in 1:reps)
{

# Initialize varhatstar, which will hold the bootstrap
# estimates of the variance

   varhatstar = rep(0,B)

# This is the bootstrap loop

   for(i in 1:B)
   {

# Resample the data WITH REPLACEMENT, and calculate the variance
      xstar = sample(x, size=n, replace=T)
      varhatstar[i] = var(xstar)
   }
   print(sd(varhatstar))
}
```

The Output:

```
> print(var(x))
[1] 76.88741

print(sd(varhatstar))
[1] 19.24616
[1] 19.32063
[1] 19.24954
[1] 19.41074
[1] 19.47081
```