Bootstrapping

Charlie Gibbons
University of California, Berkeley

ARE 210

Fall 2015

Notes

Outline

1. Concept
   - Parametric and non-parametric bootstrapping
   - Analogy principle
   - Potential for error
2. Mean and variance of the estimator
   - Bias
   - Variance
3. Inference
   - General comments
   - Confidence intervals
   - Hypothesis testing

Notes

Distribution of the estimator

We are very interested in the distribution of our estimator.

We typically appeal to asymptotic results to derive it (*i.e.*, LLN, the delta method, the CLT).

How well do these approximations work in finite samples?
How well do they work for a particular estimator?

We don't know because we only get one realization of our estimator.

Notes

Distributions of the samples

Bootstrapping is a way to provide many realizations of our estimator so that we can study its distribution.

Bootstrapping works by taking many samples *from our true sample* and analyzing the distribution of the estimator on those samples of the sample.

Then, we apply these results to the population.

Our true sample is a sample from F; our bootstrap samples are sampled from $\hat{F}$.

Notes

Non-parametric bootstrapping

There are two ways to obtain $\hat{F}$. First:

Non-parametric bootstrap

Suppose that $X_i \sim F$. We calculate

$$\hat{F}(x) = \frac{\#\{x_i \leq x\}}{N}.$$

See that sampling from $\hat{F}$ is equivalent to sampling from $\{X_i\}$ with replacement.

Sampling with replacement guarantees independence of the observations in the sample.

Notes

Parametric bootstrap

The second way is:

Parametric bootstrap

Suppose that X_i comes from a parametric distribution parameterized by ψ, $X_i \sim F_\psi$, and we estimate $\hat{\psi}$ using our sample. Here,

$$\hat{F} = F_{\hat{\psi}}.$$

The parametric bootstrap is most useful for sampling from the null distribution for hypothesis testing.

Notes

Analogy principle

Bootstrapping works based on the *analogy principle*:

a bootstrap sample is to the true sample as
the true sample is to the population;

or

a bootstrap sample is like a sample from the population.

We use repeated resampling to learn about the properties of the estimator for observations that are i.i.d. $\hat{F}$, then assume that the estimator performs similarly for observations that are i.i.d. F.

The last connection is valid because $\hat{F} \xrightarrow{P} F$.

Notes

Potential for error

Consider a sample of size N with $X_i \sim F$.

We take R samples of size N from $\hat{F}$.

There are two types of errors that arise:

- Statistical errors because $\hat{F} \neq F$.
- Sampling errors because we use a finite number of bootstrap samples.

We can make the latter arbitrarily small.

Notes

Bias of $\hat{\theta}$

The LLN may tell us that our estimator is consistent, but it may be biased.

It may be difficult to derive analytically the bias $\mathbb{E}_F\left[\hat{\theta}\right] - \theta$.

But bootstrapping gives us

- Multiple realizations of $\hat{\theta}$ from $\hat{F}$. This gives us $\hat{\mathbb{E}}_{\hat{F}}\left[\hat{\theta}\right]$, which converges to $\mathbb{E}_{\hat{F}}\left[\hat{\theta}\right]$.
- Since we have the full sample, we know the population value of θ, $\hat{\theta}$.

This enables us to estimate the bias for the sample, then we assume that this is the bias in the population.

Notes

Bias correction

How do we correct for bias? We have:

$$\hat{\theta} - \theta \approx \hat{\mathbb{E}}_{\hat{F}}\left[\hat{\theta}\right] - \hat{\theta}$$
$$\theta \approx 2\hat{\theta} - \hat{\mathbb{E}}_{\hat{F}}\left[\hat{\theta}\right].$$

In practice, if the bias is small (less than 0.25 standard errors), then it is not corrected; doing so tends to introduce more noise than it is worth.

Notes

Variance of $\hat{\theta}$

To calculate the variance of our estimator, we take our R estimates $\hat{\theta}_r^*$, find their average, $\bar{\theta}^*$, then calculate

$$\text{Var}\left(\hat{\theta}\right) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}_r^* - \bar{\theta}^*\right)^2.$$

Notes

Quantiles

For hypothesis testing and confidence intervals, we'll be interested in the quantiles of the bootstrapped distribution.

For R draws of $Y_i \overset{iid}{\sim} K$, then the jth order statistic $Y_{(j)}$ has

$$\mathbb{E}\left[Y_{(j)}\right] = K^{-1}\left(\frac{j}{R+1}\right).$$

If Y_i is a test statistic, R is the number of bootstraps, $j/(R+1)$ is the quantile of interest α, and j is the rank at that quantile of interest, which ought to be an integer. That is, $\alpha(R+1)$ should be an integer. This is why you may use $R = 999$ bootstraps.

Notes

Asymptotic refinement

Bootstrapping can be seen as an alternative to the asymptotic results that we have discussed in classical hypothesis testing.

But bootstrapping only works because $\hat{F} \xrightarrow{P} F$, which is an asymptotic result itself. Hence, bootstrapped hypothesis tests and confidence intervals are also only valid asymptotically.

Suppose that we can bootstrap a quantity that is asymptotically pivotal. Then bootstrapping approaches its asymptotic results faster than classical asymptotic approximations. This is called *asymptotic refinement*.

Since our estimator of variance can't be written as an asymptotic pivot, it is not a refinement over standard estimators.

Notes

Regular confidence intervals

Recall the asymptotic version of the confidence interval:

$$\left[\hat{\theta} - z^c\sqrt{\text{Var}\left(\hat{\theta}\right)}, \hat{\theta} + z^c\sqrt{\text{Var}\left(\hat{\theta}\right)}\right].$$

The simplest bootstrapping application would be to use the same formula, but use the bootstrapped standard error.

Notes

Basic bootstrap confidence intervals

We could instead use the basic bootstrapped confidence interval:

$$\left[2\hat{\theta} - \hat{\theta}^*_{((R+1)(1-\alpha/2))}, 2\hat{\theta} - \hat{\theta}^*_{((R+1)\alpha/2)}\right],$$

that is, $\hat{\theta}$ minus the distance from it to the $\alpha/2$th quantile and $\hat{\theta}$ plus the distance from it to the $(1-\alpha/2)$th quantile of the distribution of the $\hat{\theta}^*$.

Notes

A pivot

Consider

$$t_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\sqrt{\text{Var}\left(\hat{\theta}_b^*\right)}}.$$

This is a pivot. Rearranged, we have

$$\hat{\theta}_b^* = \hat{\theta} - t_b^* \sqrt{\text{Var}\left(\hat{\theta}_b^*\right)}.$$

Notes

Studentized bootstrap confidence interval

Choosing the appropriate percentiles gives a confidence interval of

$$\left[\hat{\theta} - t^*_{((R+1)(1-\alpha/2))} \sqrt{\text{Var}\left(\hat{\theta}\right)}, \hat{\theta} - t^*_{((R+1)(\alpha/2))} \sqrt{\text{Var}\left(\hat{\theta}\right)}\right],$$

is known as the *studentized bootstrap confidence interval* and is an improvement (theoretically) over the basic bootstrap CI.

Notes

Advantages of percentile-based methods

Bootstrap confidence intervals have the following properties:

- Asymmetric
- Range preserving
- Transformation preserving
  For monotonic transformations $h(\theta)$, a 95% CI for $h(\theta)$ will be the transformation applied to the endpoints of the CI for θ itself.

Notes

Other considerations

The standard choice of these interval need not be the shortest possible, however.

Further enhancements, such as the BC_a *bias corrected and accelerated* or ABC confidence intervals, are possible.

Notes

Monte carlo p-values

Suppose that the true sample gives a test statistic of t. After performing R bootstraps, k of the bootstrapped test statistics equal or exceed t (perhaps in absolute value). Then

$$\Pr(T \geq t; H_0) = \frac{k+1}{R+1}$$

is the *Monte carlo p-value*.

The important point is that the bootstrapping *must be done from the null distribution*.

Notes

Parametric bootstrapping the null

One way of sampling under the null is to use the parametric bootstrap. Suppose that $X_i \stackrel{iid}{\sim} F_{\theta,\lambda}$ and we consider a null hypothesis that $\theta = \theta_0$.

We estimate $\hat{\lambda}_0$ by constrained MLE on our true sample.

We draw bootstrapped samples from $F_{\theta_0,\hat{\lambda}_0}$.

Notes

Non-nested LRT

As an example, suppose that our hypotheses are

$$H_0 : X_i \overset{iid}{\sim} F_\theta$$
$$H_a : X_i \overset{iid}{\sim} F_\lambda;$$

that is, the X_i come from one parametric family or a totally distinct parametric family.

The asymptotic version of the LRT allows us to test *nested models*, where $\Omega_0 \in \Omega_1$. Since that is not the case here, we don't know what the distribution of our test statistic would be.

Bootstrapping finds the distribution of that test statistic.

Notes

Recentering

Suppose that we want to test the null that $\mu = \mu_0$—the mean takes on a particular value.

Under the null distribution, our X_i should have mean μ_0.

But they actually have mean $\bar{x}$.

Let $x_i^* = x_i - \bar{x} + \mu_0$, which now has mean μ_0.

We can bootstrap under the null by resampling from $\{x_i^*\}$.

Notes

Recentering for equality of means

Suppose that we have two samples, y_i and z_j. We want to test that $\mu_Y = \mu_Z$.

Under the null distribution, $\mu_Y = \mu_Z = \mu$.

But actually we have means $\bar{y}$, $\bar{z}$, and $\bar{x}$, where $x = \{y, z\}$.

Let $y_i^* = y_i - \bar{y} + \bar{x}$ and $z_i^* = z_i - \bar{z} + \bar{x}$. Now both samples have the same mean.

Bootstrapping proceeds by resampling from the $\{y_i^*\}$ and $\{z_i^*\}$.

Notes

Two samples with the same distribution

Keeping two samples, suppose that we have a stronger null hypothesis: $F_Y = F_Z$.

Let $\hat{F}_x$ be the ECDF of the set $\{y, z\}$.

If the true sample has n observations from Y and m from Z, then bootstrapping proceeds by drawing n observations from $\hat{F}_x$ and calling them y^* and m more observations and calling them z^*.

The standard Wald test statistic can be used to test the null.

Notes

Notes

Notes