Ch 11. Bootstrap

Ana-Maria Staicu

11/21/2019

What is bootstrap?

- A resampling technique that does not require distn assn
- Introduced by Efron 1979 as an alternative methog to jackknife to estimate the accuracy of an estimator
- Used in estimating standard error, constructing confidence intervals, approximating p-value

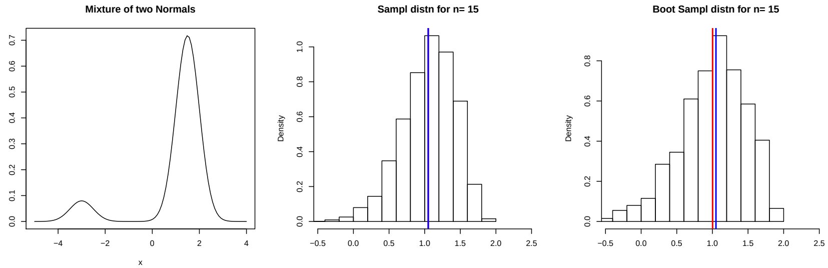
Illustration: Sample mean

Interest: estimate population mean using sample mean from an IID sample. What is the accuracy of your estimator? What is the sampling variability of the estimator?

- $Y_1, \ldots, Y_n$ IID sample from unknown distn with (μ, σ^2)
- Estimator: $\widehat{\mu} = \bar{Y}$. Mean/variance: $E[\widehat{\mu}] = \mu$; $Var(\widehat{\mu}) = \sigma^2/n$
- Accuracy: estimate $Var(\widehat{\mu})$ using the sample standard deviation $\{\sum_{i=1}^{n}(y_i - \bar{y})^2/n\}^{1/2}$
- Statistical inference: need sampling distn of $\widehat{\mu}$! CLT-based confidence intervals require large n!
- Reliable inference requires n sufficiently large! Also how to estimate the sampling variability of the estimator, if it's not readily available?

Illustration: Sample mean (cont'd)

Left: underlying population distn. Middle: Sampling distn of the sample mean. Right: Approx sampling distn using bootstrap



In most cases we only have access to a sample from the pop distn. How to approximate the sampling distn of the estimator, when the sample size is moderate?

General intuition

Setting: $Y_1, \ldots, Y_n$ IID from distn F and θ a para attached to F. Mathematically, describe it via using functional $t(\cdot)$

$$\theta = t(F).$$

Example: mean $\mu = \int y dF(y) = E_F[Y_1]$;

variance $\sigma^2 = \int (y-\mu)^2 dF(y) = E_F[(Y_1-\mu)^2]$; etc.

Goal: Find an estimator for θ and calculate its standard deviation.

Plug-in estimator: $\widehat{\theta} = t(\widehat{F}_n)$, where $\widehat{F}_n$ is the empirical CDF defined

$$\widehat{F}_n(y) = \frac{|\{Y_i : Y_i \leq y,\ i = 1, \ldots, n\}|}{n}.$$

What is the plug in estimator for μ and σ^2? Eg $\widehat{\mu} = E_{\widehat{F}_n}[Y_1]$.

Remark: The para θ for F is viewed similarly to how $\widehat{\theta}$, for a given sample, is for $\widehat{F}_n$, for that sample!

Real world and Bootstrap world

Real world:

- unknown pop distn F and $\theta = t(F)$
- $Y_1, \ldots, Y_n$ is IID sample drawn from F
- $\widehat{\theta} = s(\boldsymbol{Y})$ is estimator/statistic of interest; $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$

Sampling distn $\widehat{\theta}$: 1) draw B data sets of size n from F; 2) compute $\widehat{\theta}^b$ for each data set b; 3) Approx dist of $\widehat{\theta}$ by distn of $\{\widehat{\theta}^1, \ldots, \widehat{\theta}^B\}$.

Essentially use Monte Carlo simulation to get sampling distn of $\widehat{\theta}$.

Remark: The sampling distn, due to its dependence on F, is not always accessible!

Real world and Bootstrap world (cont'd)

Bootstrap world (always accessible). Say you observe data $\boldsymbol{y}$:

- estimate F by empirical distn $\widehat{F}_n$; $\widehat{\theta}_{\boldsymbol{y}} = t(\widehat{F}_n)$
- $Y_1^*, \dots, Y_n^*$ is sample drawn from $\widehat{F}_n$.
- $\widehat{\theta}^* = s(\boldsymbol{Y}^*)$ - based on bootstrap sample $\boldsymbol{Y}^* = (Y_1^*, \dots, Y_n^*)$

The statistic $\widehat{\theta}^*$ is called "bootstrap replication".

- How many different boot samples of size n can draw from $\widehat{F}_n$?

$$\binom{2n-1}{n-1}$$

- This gets large quickly! Instead use reasonable large B number of samples from $\widehat{F}_n$. For standard error estimation: $B \approx 200$
- Distn of $(\widehat{\theta} - \theta)$ is approx by the distn of the boot replicates $(\widehat{\theta}^{*b} - \widehat{\theta}_{\boldsymbol{y}})$, $b = 1, \dots, B$!

More intuition

More generally, assume $\widehat{\theta}$ is $AN(\theta, \sigma^2/n)$. A one-term Edgeworth expansion (expansion of the CDF using its cumulants) for $\widehat{\theta}$ gives

$$P\left\{\sqrt{n}(\widehat{\theta} - \theta) \leq x\right\} = \Phi(x/\sigma) + \frac{c}{\sqrt{n}} + o(n^{-1/2}) \quad \text{for each } x,$$

where Φ is the CDF of N(0,1).

Analogously, in the bootstrap world we have

$$P^*\left\{\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \leq x \middle| \boldsymbol{Y}\right\} = \Phi(x/\sigma_n^*) + \frac{c_n}{\sqrt{n}} + o_p(n^{-1/2}); \quad \text{for each } x,$$

where $\sigma_n^* \to_p \sigma$ and $c_n \to_p c$ as $n \to \infty$.

More intuition (cont'd)

Now suppose that $\sigma_n^* - \sigma = O_p(n^{-1/2})$. It follows (Taylor series)

$$\begin{aligned} P\left\{\sqrt{n}(\widehat{\theta} - \theta) \le x\right\} &- P^*\left\{\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \le x|\boldsymbol{Y}\right\} \\ &= \Phi(x/\sigma) + \frac{c}{\sqrt{n}} - \left\{\Phi(x/\sigma_n^*) + \frac{c_n}{\sqrt{n}}\right\} + o_p(n^{-1/2} \\ &= O_p(n^{-1/2}). \end{aligned}$$

It follows that the bootstrap distn of $\sqrt{n}(\widehat{\theta}^* - \widehat{\theta})$ is within $O_p(n^{-1/2})$ of the distn of $\sqrt{n}(\widehat{\theta} - \theta)$.

Bootstrap alg. for estimating standard errors

- Denote sample from the unknown distn F by $Y_1, \ldots, Y_n$
- Select B independent bootstrap samples $\boldsymbol{Y}^{*1}, \ldots, \boldsymbol{Y}^{*B}$ each consisting of n data values drawn with replacement from $Y_1, \ldots, Y_n$.
- Evaluate the bootstrap replication corresponding to each bootstrap sample

$$\widehat{\theta}^{*b} = s(\boldsymbol{Y}^{*b}), \qquad b = 1, \ldots, B$$

- Estimate the standard error $se_F(\widehat{\theta})$ by the sample standatd deviation of the B bootstrap replications

$$\widehat{se}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{\theta}^{*b} - \bar{\theta}^{*})^2 \right\}^{1/2}$$

where $\bar{\theta}^{*} = \sum_{b=1}^{B} \widehat{\theta}^{*b} / B$.

Bootstrap alg. for estimating standard errors (cont'd)

- The bootstrap estimate of $se_F(\widehat{\theta})$ is a plug-in estimate that uses $\widehat{F}_n$ in place of the unknown F and is defined by $se_{\widehat{F}_n}(\widehat{\theta}^*)$

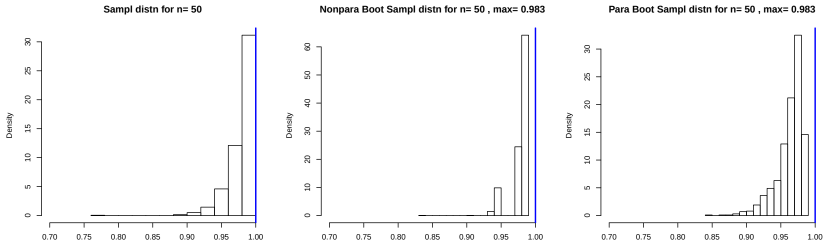
$$se_{\widehat{F}_n}(\widehat{\theta}^*) = \lim_{B\to\infty} \widehat{se}_B$$

- We refer to $B = \infty$ by "ideal bootstrap"; $\widehat{se}_\infty$ is the "ideal bootstrap estimate of standard error".

The type of bootstrap discussed here is called "non-parametric bootstrap" because it uses NO information about the underlying distn. This is in contrast to the "parametric bootstrap" which uses a different estimate for F based on assumed parametric model.

Example of bootstrap failure

Setting: $Y_1, \ldots, Y_n \sim$ IID $Unif(0, \theta)$. The MLE of θ is $Y_{(n)}$ - the largest sample value.



Left: Sampling distn of $Y_{(n)}$. Middle: Nonpara Boot approx sampling distn. Right: Para Boot approx

- What happens with the nonparam bootstrap? The empirical distn $\widehat{F}_n$ is not a good estimate of the true, in the extreme tail
- In general the nonparam bootstrap fails if the parameter is non a smooth functional (Bickel and Freedman 1981, Shao 1994)
- In this case, more knowledge of F is required to remedy matters. What is param bootstrap?

Parametric bootstrap

Underlying distn depends on para η, say $F = F(\cdot;\eta)$ and let θ be para of interest. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ be IID sample from F. Let $\widehat{\theta} = s(\boldsymbol{Y})$ be estimator of θ as before.

Bootstrap world (always accessible):

- estimate distn $\widehat{F}(\cdot) = F(\cdot, \widehat{\eta})$, for $\widehat{\eta}$ based on observed data $\boldsymbol{y}$
- $Y_1^*, \ldots, Y_n^*$ is sample drawn from $\widehat{F}$.
- $\widehat{\theta}^* = s(\boldsymbol{Y}^*)$ - based on bootstrap sample $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_n^*)$; called "bootstrap replication".
- How many different boot samples of size n can draw from $\widehat{F}$?
- Use reasonable large B number of samples from $\widehat{F}$. For standard error estimation: $B \approx 200$
- Distn of $(\widehat{\theta} - \theta)$ is approx by the distn of boot replicates $(\widehat{\theta}^{*b} - \widehat{\theta}_{\boldsymbol{y}})$, $b = 1, \ldots, B$!

Golden rule of bootstrapping

Bootstrap statistics are to the original sample statistic

as

the original sample statistic is to the population parameter

Application of bootstrap (I)

Let X and Y designate the yield return of two financial assets of interest. Denote by α the fraction of our money to be invested in X; $(1-\alpha)$ fraction is invested in Y. The yield return is

$$\alpha X + (1-\alpha)Y$$

The optimal α is the value that minimizes the risk of our investments (variance of the investments),
$\alpha_{opt} = \arg\min_{\alpha\in(0,1)} Var\{\alpha X + (1-\alpha)Y\}$.

Algebra gives (under some assn)

$$\alpha_{opt} = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

$\sigma_X^2 = VarX$, $\sigma_Y^2 = VarY$, and $\sigma_{XY} = Cov(X, Y)$.

Application of bootstrap (I, cont'd)

Suppose the data consists of 50 pairs $\{(x_i, y_i) : i = 1, \ldots, 50\}$. Estimate α_{opt} and its variability!

Compute estimates for of the co/variances , say $\widehat{\sigma}_X^2$, $\widehat{\sigma}_Y^2$ and $\widehat{\sigma}_{XY}$ and get a plug-in estimate of α_{opt},

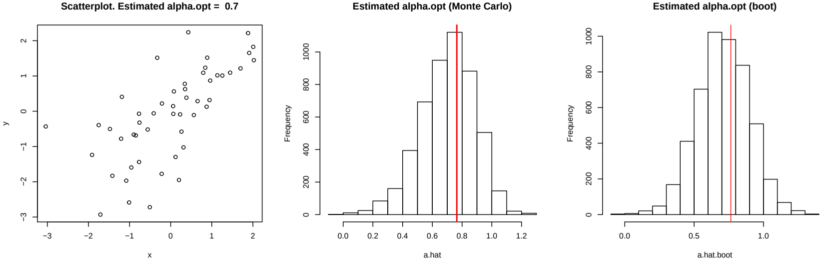
$$\widehat{\alpha}_{opt} = \frac{\widehat{\sigma}_Y^2 - \widehat{\sigma}_{XY}}{\widehat{\sigma}_X^2 + \widehat{\sigma}_Y^2 - 2\widehat{\sigma}_{XY}}.$$

Remarks:

- What is the sampling distn of $\widehat{\alpha}_{opt}$?
- How to calculate the accuracy/precision of the estimator $\widehat{\alpha}_{opt}$?
- Suppose parametric assumption about the distribution for (X, Y) are not reasonable!

Application of bootstrap (I, cont'd)

Left: Scatterplot of data ($\widehat{\alpha}_{opt,data} = 0.7$). Middle: Sampling distn of $\widehat{\alpha}_{opt}$ as approx by Monte Carlo simulation (true mean ≈ 0.76). Right: Sampling distn of $\widehat{\alpha}_{opt}$ by resampling the pairs (*bootstrap*).



Application of bootstrap (II): linear regression

Consider data (X_i, Y_i) for $i = 1, \ldots, n$ and assume the linear model $Y_i = \alpha + \beta X_i + \epsilon_i$ where $\epsilon_i \sim (0, \sigma_i^2)$.

Least squares estimators for α and β are:

$$
\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
& \quad Var(\widehat{\beta}) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sigma_i^2}{SS_X^4} \\
& \quad SS_X^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \\
\widehat{\alpha} &= \bar{Y} - \widehat{\beta}\bar{X}
\end{aligned}
$$

Linear regression (cont'd)

Classical bootstrap: resample the residuals

- Estimate the residuals $e_i = Y_i - \widehat{\alpha} - \widehat{\beta} X_i$
- Draw $e_1^*, \ldots, e_n^*$ from $\{\widehat{e}_1, \ldots, \widehat{e}_n\}$ where $\widehat{e}_i = e_i - n^{-1} \sum_{i=1}^n e_i$
- Form bootstrap (X_i^*, Y_i^*) where $X_i^* = X_i$ and $Y_i^* = \widehat{\alpha} + \widehat{\beta} X_i + e_i^*$
- Fit linear regression model and estimate $\widehat{\beta}^*$ and $\widehat{\alpha}^*$. Obtain

$$
\begin{aligned}
\widehat{\beta}^* &= \widehat{\beta} + \frac{\sum_{i=1}^n (X_i - \bar{X})(e_i^* - \bar{e}^*)}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
\widehat{\alpha}^* &= \widehat{\alpha} + (\widehat{\beta} - \widehat{\beta}^*)\bar{X} + \bar{e}^*
\end{aligned}
$$

Repeat the procedure B times

* $Var_B(\widehat{\beta}^*) = E_B[(\widehat{\beta}^* - \widehat{\beta})^2] \approx Var(\widehat{\beta})$ is efficient when $\sigma_i^2 = \sigma^2$

* $Var_B(\widehat{\beta}^*)$ does not approximate $Var(\widehat{\beta})$ when $\sigma_i^2 \neq \sigma^2$ (inconsistent when the errors are heteroscedastic).

Linear regression (cont'd)

Bootstrap of the pairs: resample the pairs

- Resample the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$
- Let $(X_1^*, Y_1^*), \ldots, (X_n^* Y_n^*)$ be the bootstrap sample
- Fit linear regression model and estimate $\widehat{\beta}^*$ and $\widehat{\alpha}$. Obtain

$$\begin{aligned}\widehat{\beta}^* &= \frac{\sum_{i=1}^{n}(X_i^* - \bar{X}^*)(Y_i^* - \bar{Y}^*)}{\sum_{i=1}^{n}(X_i^* - \bar{X}^*)^2} \\ \widehat{\alpha}^* &= \bar{Y}^* - \widehat{\beta}^* \bar{X}^*\end{aligned}$$

Repeat the procedure B times

* $Var_B(\widehat{\beta}^*) \approx Var(\widehat{\beta})$ even when $\sigma_i^2 \neq \sigma^2$

Thus *pair bootstrap* is robust to heteroscedasticity.

When does bootstrap work well?

- Sample Means
- Sample Variances
- Sample Coefficient of Variation
- Maximum Likelihood Estimators
- Least Squares Estimators
- Correlation Coefficients
- Regression Coefficients
- Smooth transforms of these statistics

Remarks

- Bootstrap is based on *resampling the original data*. Characteristics about the generating distn, that are present in the data, are expected to be present in resamples of the data.

- The study of bootstrap has been expanded beyond IID

- Resampling is a Monte Carlo method of *simulating datasets* from a given data, *but without assumptions about the underlying distn*.

- Resampling procedures are supported by solid theoretical foundations.

- Key monographs on bootstrap (IID): Efron and Tibshirani (1993), Hall (1992), Davison and Hinkley (1997), Chernik (2007), Chernik and Labudde(2011) etc + (dependent data) Lahiri, (2003), etc.