Computer Intensive Methods in Mathematical Statistics

Johan Westerborn

Department of mathematics
KTH Royal Institute of Technology
johawes@kth.se

Lecture 14
More on the bootstrap
11 May 2017

Plan of today's lecture

1. Last time: Introduction to bootstrap (Ch. 7)
2. More on the bootstrap (Ch. 7)
   - Example: law schools
   - Parametric bootstrap
   - Semi-parametric bootstrap
3. MC methods for hypothesis testing (Ch. 8)
   - Preliminaries
   - MC tests
   - Permutation tests

Outline

Statistical problem under consideration

- We assume that we have at hand
  - observations y
  - and a (possibly parametric) model $\mathcal{P}$ for the data.
- In this setting we want to make inference about some property (estimand) $\tau = \tau(\mathbb{P}_0)$ of the distribution $\mathbb{P}_0$ that generated the data.
- For instance,

$$\tau(\mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0}(Y) = \int x f_0(x)\,dx, \quad \text{(mean)}$$

  where f_0 is the density of $\mathbb{P}_0$.
- The estimand τ is estimated using a statistic $t(y)$.

Statistical problem under consideration

- We assume that we have at hand
  - observations y
  - and a (possibly parametric) model $\mathcal{P}$ for the data.
- In this setting we want to make inference about some property (estimand) $\tau = \tau(\mathbb{P}_0)$ of the distribution $\mathbb{P}_0$ that generated the data.
- For instance,

$$\tau(\mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0}(Y) = \int x f_0(x)\,dx, \quad \text{(mean)}$$

where f_0 is the density of $\mathbb{P}_0$.
- The estimand τ is estimated using a statistic $t(y)$.

Statistical problem under consideration

- We assume that we have at hand
  - observations y
  - and a (possibly parametric) model $\mathcal{P}$ for the data.
- In this setting we want to make inference about some property (estimand) $\tau = \tau(\mathbb{P}_0)$ of the distribution $\mathbb{P}_0$ that generated the data.
- For instance,

$$\tau(\mathbb{P}_0) = \mathbb{E}_{\mathbb{P}_0}(Y) = \int x f_0(x)\, dx, \quad \text{(mean)}$$

  where f_0 is the density of $\mathbb{P}_0$.
- The estimand τ is estimated using a statistic $t(y)$.

Uncertainty of estimators

- It is important to always keep in mind that the estimate $t(y)$ is an observation of a random variable $t(Y)$. If the experiment was repeated, resulting in a new vector y of random observations, the estimator would take another value.
- In the same way, the error $\Delta(y) = t(y) - \tau$ is a realization of the random variable $\Delta(Y) = t(Y) - \tau$.
- To assess the uncertainty of the estimator we thus need to analyze the distribution of the error $\Delta(Y)$ (error distribution).

Bootstrap in a nutshell

- Using the bootstrap algorithm we deal with this matter by
  1. replacing $\mathbb{P}_0$ by an data-based approximation $\widehat{\mathbb{P}}_0$ and
  2. analyzing the variation of $\Delta(Y)$ by MC simulation from the approximation $\widehat{\mathbb{P}}_0$.
- A generic way to obtain the approximation $\widehat{\mathbb{P}}_0$ is to use the empirical distribution.

The empirical distribution (ED)

- The empirical distribution (ED) $\widehat{\mathbb{P}}_0$ associated with the data $y = (y_1, y_2, \ldots, y_n)$ gives equal weight $(1/n)$ to each of the y_is (assuming that all the y values are distinct).
- Consequently, if $Z \sim \widehat{\mathbb{P}}_0$ is a random variable, then Z takes the value y_i with probability $1/n$.
- The empirical distribution function (EDF) associated with the data y is defined by

$$\widehat{F}_n(z) = \widehat{\mathbb{P}}_0(Z \leq z)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i \leq z\}} = \text{fraction of } y_i\text{'s that are less than } z.$$

The empirical distribution (ED)

- The empirical distribution (ED) $\widehat{\mathbb{P}}_0$ associated with the data $y = (y_1, y_2, \ldots, y_n)$ gives equal weight ($1/n$) to each of the y_is (assuming that all the y values are distinct).
- Consequently, if $Z \sim \widehat{\mathbb{P}}_0$ is a random variable, then Z takes the value y_i with probability $1/n$.
- The empirical distribution function (EDF) associated with the data y is defined by

$$\widehat{F}_n(z) = \widehat{\mathbb{P}}_0(Z \leq z)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i \leq z\}} = \text{fraction of } y_i\text{'s that are less than } z.$$

The empirical distribution (ED)

- The empirical distribution (ED) $\widehat{\mathbb{P}}_0$ associated with the data $y = (y_1, y_2, \ldots, y_n)$ gives equal weight $(1/n)$ to each of the y_is (assuming that all the y values are distinct).
- Consequently, if $Z \sim \widehat{\mathbb{P}}_0$ is a random variable, then Z takes the value y_i with probability $1/n$.
- The empirical distribution function (EDF) associated with the data y is defined by

$$\widehat{F}_n(z) = \widehat{\mathbb{P}}_0(Z \leq z)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i \leq z\}} = \text{fraction of } y_i\text{'s that are less than } z.$$

Properties of the EDF

- It holds that

$$\lim_{z \to -\infty} \widehat{F}_n(z) = \lim_{z \to -\infty} F(z) = 0,$$
$$\lim_{z \to \infty} \widehat{F}_n(z) = \lim_{z \to \infty} F(z) = 1.$$

- In addition, $n\widehat{F}_n(z) \sim \text{Bin}(n, F(z))$.
- By the LLN (as $n \to \infty$),

$$\widehat{F}_n(z) \to F(z) \quad \text{(a.s.)}$$

- and by the CLT,

$$\sqrt{n}(\widehat{F}_n(z) - F(z)) \xrightarrow{\text{d.}} \text{N}(0, \sigma^2(z)),$$

where

$$\sigma^2(z) = F(z)(1 - F(z)).$$

Generating samples from the ED

- Consequently a sample Y^* of size n from the empirical distribution $\widehat{\mathbb{P}}_0$ associated with the observations $y = (y_1, y_2, \ldots, y_n)$ is generated by
  1. drawing indices $I_1, I_2, \ldots, I_n$ independently from the uniform distribution on the integers $\{1, 2, \ldots, n\}$, and
  2. letting $Y^* = (y_{I_1}, y_{I_2}, \ldots, y_{I_n})$.
- Note that this algorithm simply draws n values from the set $\{y_1, y_2, \ldots, y_n\}$ with replacement.

The bootstrap

- Having access to data y, we may now replace $\mathbb{P}_0$ by $\widehat{\mathbb{P}}_0$.
- Any quantity involving $\mathbb{P}_0$ can now be approximated by plugging $\widehat{\mathbb{P}}_0$ into the quantity instead. In particular,

$$\tau = \tau(\mathbb{P}_0) \approx \widehat{\tau} = \tau(\widehat{\mathbb{P}}_0),$$

which, e.g., in the case of the mean, becomes

$$\tau = \mathbb{E}_{\mathbb{P}_0}(Y) \approx \widehat{\tau} = \mathbb{E}_{\widehat{\mathbb{P}}_0}(Y) = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The bootstrap (cont.)

- Moreover, the uncertainty of $t(y)$ can be analyzed by drawing repeatedly $Y^* \sim \widehat{\mathbb{P}}_0$ and look at the variation (histogram) of $\Delta(Y^*) = t(Y^*) - \tau \approx t(Y^*) - \widehat{\tau}$.
- In the case of the empirical distribution, simulation from $\widehat{\mathbb{P}}_0$ is carried through by simply drawing, with replacement, among the values $y_1, \ldots, y_n$.

The bootstrap: algorithm

- Construct the ED $\widehat{\mathbb{P}}_0$ from the data y.
- Simulate B new data sets Y_b^*, $b \in \{1, 2, \ldots, B\}$, where each Y_b^* has the size of y, from $\widehat{\mathbb{P}}_0$. Each Y_b^* is obtained by drawing, with replacement, n times among the y_is.
- Compute the values $t(Y_b^*)$, $b \in \{1, 2, \ldots, B\}$, of the estimator.
- By setting in turn $\Delta_b^* = t(Y_b^*) - \widehat{\tau}$, $b \in \{1, 2, \ldots, B\}$, we obtain values being approximately distributed according to the error distribution. These can be used for uncertainty analysis.

Etymology

Figure: The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps. — *"The Surprising Adventures of Baron Munchausen"* by Rudolph Eric

Bootstrap confidence bounds

- Recall that a confidence bound for τ on the level $1-\alpha$ is given by

$$I_\alpha = \left(\widehat{\tau} - F_\Delta^{-1}(1-\alpha/2), \widehat{\tau} - F_\Delta^{-1}(\alpha/2)\right),$$

  where F_Δ is the error distribution function.
- Having bootstrapped errors $(\Delta_b^*)_{b=1}^B$ being approximately distributed according to F_Δ, we may use the approximation

$$F_\Delta^{-1}(p) \approx \Delta^*_{(\lceil Bp \rceil)}, \quad p \in (0,1),$$

  where $(\Delta^*_{(1)}, \ldots, \Delta^*_{(B)})$ are the ordered errors.

Bootstrap confidence bounds (cont.)

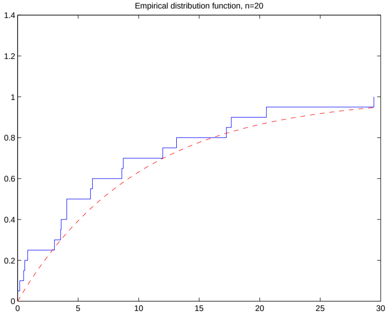
- This gives the bootstrap confidence bound

$$\mathrm{I}_\alpha = \left(\widehat{\tau} - \Delta^*_{(\lceil B(1-\alpha/2) \rceil)}, \widehat{\tau} - \Delta^*_{(\lceil B\alpha/2 \rceil)} \right),$$

  for τ on the level $1-\alpha$.
- One-sided intervals are obtained analogously (just let $\alpha/2 \leftarrow \alpha$).

A toy example

- We let $y = (y_1, \ldots, y_{20})$ be i.i.d. with unknown mean θ (and unknown distribution). As estimator we take, as usual, $t(y) = \bar{y}$.



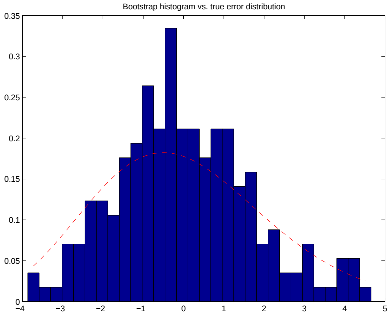
A toy example: MATLAB implementation

- In MATLAB:

```
n = 20;
B = 200;
tau_hat = mean(y);
boot = zeros(1,B);
for b = 1:B, % bootstrap
    I = randsample(n,n,'true',ones(1,n));
    boot(b) = mean(y(I));
end
delta = sort(boot - tau_hat);
alpha = 0.05; % CB level
L = tau_hat - delta(ceil((1 - alpha/2)*B));
U = tau_hat - delta(ceil(alpha*B/2));
```

A toy example: exponential distribution

- This histogram of $(\Delta_b^*)_{b=1}^{200}$ looks like follows:



- The associated confidence bound is (in fact, the y_is were observations of $Y_i \sim \text{Exp}(\theta)$, simulated under $\theta_0 = 10$)

$$I_{.05} = (7.3, 15.7).$$

Outline

Non-parametric Bootstrap

- The bootstrap algorithm considered above is non-parametric in the sense that we have no assumptions on the distribution $\mathbb{P}_0$ apart from the samples being i.i.d.; in particular, we do not assume that $\mathbb{P}_0$ belongs to a certain parametric family.
- Our approximation $\widehat{\mathbb{P}}_0$ of $\mathbb{P}_0$ is the empirical distribution function.
- The simulation step boils down to drawing from the empirical distribution, i.e., drawing from the data with replacement.

Non-parametric Bootstrap

- The bootstrap algorithm considered above is non-parametric in the sense that we have no assumptions on the distribution $\mathbb{P}_0$ apart from the samples being i.i.d.; in particular, we do not assume that $\mathbb{P}_0$ belongs to a certain parametric family.
- Our approximation $\widehat{\mathbb{P}}_0$ of $\mathbb{P}_0$ is the empirical distribution function.
- The simulation step boils down to drawing from the empirical distribution, i.e., drawing from the data with replacement.

Non-parametric Bootstrap

- The bootstrap algorithm considered above is non-parametric in the sense that we have no assumptions on the distribution $\mathbb{P}_0$ apart from the samples being i.i.d.; in particular, we do not assume that $\mathbb{P}_0$ belongs to a certain parametric family.
- Our approximation $\widehat{\mathbb{P}}_0$ of $\mathbb{P}_0$ is the empirical distribution function.
- The simulation step boils down to drawing from the empirical distribution, i.e., drawing from the data with replacement.

Outline

Example: law schools

- We have average test scores (LSAT and GPA) from 15 american law schools and want to investigate if the two scores are correlated, i.e. τ is the correlation between the two datasets.
- Our data consists of pairs $(x, y) = ((x_1, y_1), \ldots, (x_{15}, y_{15}))$.
- When estimating the correlation we take a nonparametric approach as follows.

Example: law schools (cont.)

1. Estimate the correlation of the data using the sample correlation

$$\widehat{\tau} = t(x, y) = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\sqrt{n \sum_i x_i^2 - (\sum_i x_i)^2} \sqrt{n \sum_i y_i^2 - (\sum_i y_i)^2}} \approx 0.776.$$

2. Create bootstrap samples $(X, Y)_b^*$, $b \in \{1, 2, \ldots, B\}$, where each sample $(X, Y)_b^*$ is generated by drawing 15 times with replacement from the pairs (x_i, y_i), $i \in \{1, \ldots, 15\}$.
3. Calculate the correlation $t((X, Y)_b^*)$ for each random sample.

Example: law schools (cont.)

- Given the $(X, Y)^*_b$s we create variables $\Delta^*_b = t((X, Y)^*_b) - \widehat{\tau}$, $b \in \{1, 2, \ldots, B\}$, being approximately distributed according to the error distribution.
- This gives that the bias of our estimate is approximately $\mathbb{E}(\Delta(X, Y)) \approx \overline{\Delta^*} = -0.0057$.
- The bias-corrected estimate is $t(x, y) - \overline{\Delta^*} = 0.783$.
- A one-sided 95%-confidence interval for the correlation is, consequently,

$$\begin{aligned} I_{0.05} &= \left(\widehat{\tau} - F_\Delta^{-1}(0.95), 1\right) \\ &\approx \left(\widehat{\tau} - \Delta^*_{(\lceil 0.95B \rceil)}, 1\right) \\ &= (0.614, 1). \end{aligned}$$

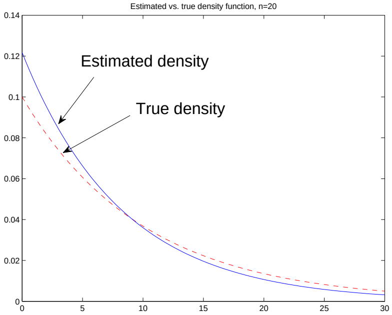
Outline

Parametric bootstrap

- In the non-parametric bootstrap we had no assumptions on the distribution function apart from the observed data y being i.i.d.
- In the parametric bootstrap we assume that data comes from a distribution $\mathbb{P}_0 = \mathbb{P}_{\theta_0} \in \{\mathbb{P}_\theta; \theta \in \Theta\}$ belonging to some parametric family.
- Instead of using the ED, we find an estimate $\widehat{\theta} = \widehat{\theta}(y)$ of θ_0 from our observations and
  1. generate new bootstrapped samples Y_b^*, $b \in \{1, 2, \ldots B\}$, from $\widehat{\mathbb{P}}_0 = \mathbb{P}_{\widehat{\theta}}$.
  2. After this, we form, as usual, bootstrap estimates $\widehat{\theta}(Y_b^*)$ and errors $\Delta_b^* = \widehat{\theta}(Y_b^*) - \widehat{\theta}$, $b \in \{1, 2, \ldots B\}$.

A toy example: exponential distribution

- We let $y = (y_1, \ldots, y_{20})$ be i.i.d. observations of $Y_i \sim \text{Exp}(\theta_0)$, with unknown mean θ_0. The MLE of θ_0 is $\widehat{\theta}(y) = \bar{y}$ and following plot displays $\text{Exp}(\widehat{\theta}(y))$ vs. $\text{Exp}(\theta_0)$.

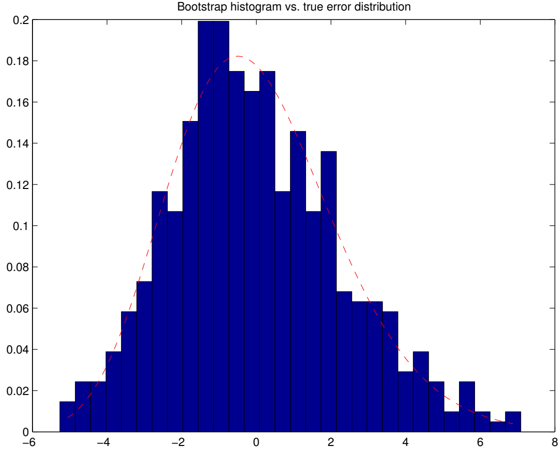


A toy example: exponential distribution (cont.)

In Matlab:

```
n = 20;
B = 500;
theta_hat = mean(y);
boot = zeros(1,B);
for b = 1:B, % bootstrap
    y_boot = exprnd(theta_hat,1,n);
    boot(b) = mean(y_boot);
end
delta = sort(boot - theta_hat);
alpha = 0.05; % CB level
L = theta_hat - delta(ceil((1 - alpha/2)*B));
U = theta_hat - delta(ceil(alpha*B/2));
```

A toy example: exponential distribution (cont.)



Outline

Semi-parametric bootstrap: regression

- We assume a parametric model for the data, for instance

$$Y_i = kx_i + m + \varepsilon_i, \quad i \in \{1, 2, \dots n\},$$

  and a non-parametric model for the residuals ε_i.
- Our only assumption on the residuals are that these are i.i.d.
- Given data $y = (y_1, \dots, y_n)$ we want construct estimators $\widehat{k}(y)$ and $\widehat{m}(y)$ of the parameters k and m and assess the uncertainty of the same.

Semi-parametric bootstrap: regression (cont)

- To do this we want to generate bootstrap samples Y_b^* and parameter estimates $\widehat{k}(Y_b^*)$ and $\widehat{m}(Y_b^*)$ and study the variation of, e.g., $\Delta_b^* = \widehat{k}(Y_b^*) - \widehat{k}(y)$.
- A confidence interval for k is then given by

$$\left(\widehat{k}(y) - \Delta^*_{(\lceil B(1-\alpha/2)\rceil)}, \widehat{k}(y) - \Delta^*_{(\lceil B\alpha/2\rceil)}\right).$$

- Bootstrap samples Y_b^* are obtained by bootstrapping the residuals and adding these to the line. We proceed as follows.

Semi-parametric bootstrap: regression (cont.)

1. Find estimators $\widehat{k} = \widehat{k}(y) = S_{xy}/S_{xx}$ and $\widehat{m} = \widehat{m}(y) = \bar{y} - \widehat{k}\bar{x}$ for the parameters using least squares.
2. Estimate the residuals as

$$\widehat{\varepsilon}_i = y_i - \widehat{k}x_i - \widehat{m}, \quad i \in \{1, 2, \ldots, n\}.$$

3. Now, the $\widehat{\epsilon}_i$'s approximately form an i.i.d. sample from an unknown distribution. Hence, for $b = 1, 2, \ldots, B$,
   - (i) generate, by resampling, $\epsilon_b^* = (\epsilon_1, \ldots \epsilon_n)_b^*$ and
   - (ii) use the bootstrapped residuals to generate bootstrapped observations

$$(Y_i)_b^* = \widehat{k}x_i + \widehat{m} + (\epsilon_i)_b^*.$$

   - (iii) Given the bootstrapped observations, estimate the parameters to obtain $\widehat{k}(Y_b^*)$ and $\widehat{m}(Y_b^*)$.

Example: Gaussian residuals

- Assume that $Y_i = kx_i + m + \varepsilon_i$, with standard Gaussian residuals.
- To test the semi-parametric bootstrap we simulate a data set with $m = 3$ and $k = 4$.
- Given data, the parameters are estimated using least squares estimation.
- For comparison, we know from the theory of linear regression that an exact confidence interval for k is

$$\mathrm{I}_\alpha = \left(\widehat{k} - t_{\alpha/2}(n-2)s_b, \widehat{k} + t_{\alpha/2}(n-2)s_b\right),$$

where

$$s_b^2 = \frac{\frac{1}{n-2}\sum_i \widehat{\epsilon}_i^2}{\sum_i (x_i - \bar{x})^2}.$$

Example: Gaussian residuals (cont.)

- Applying this to the given data set yields the exact bound

$$I_{0.05} = (3.84, 4.79) .$$

- For a comparison we applied semi-parametric as well as parametric bootstrap to the same data set.
  - Using semi-parametric bootstrap, where we resample the estimated residuals as above, we obtain the interval

$$I_{0.05} = (3.85, 4.78) .$$

  - Instead using parametric bootstrap, where we draw new residuals from $N(0, \hat{\sigma}^2)$, we obtain

$$I_{0.05} = (3.86, 4.77) .$$

Summary: Different types of bootstrap

- Non-parametric bootstrap
  - makes no assumptions on the distribution apart from i.i.d.
  - needs more data than parametric.
- Parametric bootstrap
  - assumes that data comes from a parametric family of distributions.
  - needs less data to obtain good estimates due to stronger assumptions.
  - may however be sensitive to assumptions.
- Semi-parametric bootstrap
  - assumes a parametric model, coupled with non-parametric nuisance variables, often residuals.
  - is typically used for regression.

Outline

1. Last time: Introduction to bootstrap (Ch. 7)
2. More on the bootstrap (Ch. 7)
   - Example: law schools
   - Parametric bootstrap
   - Semi-parametric bootstrap
3. MC methods for hypothesis testing (Ch. 8)
   - Preliminaries
   - MC tests
   - Permutation tests

Outline

1. Last time: Introduction to bootstrap (Ch. 7)
2. More on the bootstrap (Ch. 7)
   - Example: law schools
   - Parametric bootstrap
   - Semi-parametric bootstrap
3. MC methods for hypothesis testing (Ch. 8)
   - Preliminaries
   - MC tests
   - Permutation tests

Statistical hypotheses

- A statistical hypothesis is a statement about the distributional properties of data.
- The goal of a hypothesis test is to see if data agrees with the statistical hypothesis.
- Rejection of a hypothesis indicates that there is sufficient evidence in the data to make the hypothesis unlikely.
- Strictly speaking, a hypothesis test does not accept a hypothesis; it fails to reject it.

Testing hypotheses

- The basis of a hypothesis test consist of
  - a null hypothesis $\mathcal{H}_0$ that we wish to test.
  - a test statistic $t(y)$, i.e., a function of the observed data y.
  - a critical region R.
- If the test statistic falls into the critical region, then we reject the null hypothesis $\mathcal{H}_0$.

Important concepts

Significance	The probability (risk) that the test incorrectly rejects the null hypothesis.
Power	The probability that the test rejects correctly the null hypothesis. Is a function of the true parameter.
p-value	The probability, given the null hypothesis, of observing a result at least as extreme as the test statistic.
Type I error	Incorrectly rejecting the null hypothesis.
Type II error	Failing to reject the null hypothesis.

Testing simple hypotheses

- A simple hypothesis specifies completely a single distribution for the data, e.g., $Y \sim \mathrm{N}(\theta, 1)$ with $\mathcal{H}_0 : \theta = 0$.
- We construct/define a test statistic $t(y)$ such that large values of $t(y)$ indicate evidence against $\mathcal{H}_0$.
- The p-value of the test is now $p(y) = \mathbb{P}(t(Y) \geq t(y) \| \mathcal{H}_0)$.
- The rejection region is $\mathrm{R} = \{y : p(y) \leq \alpha\}$, where α is the level of the test.
- Thus, to evaluate the p-value we need to be able to compute probabilities under the distribution of $t(Y)$ under $\mathcal{H}_0$.
- This can be tricky if this distribution is complex. Use MC!

Outline

1. Last time: Introduction to bootstrap (Ch. 7)
2. More on the bootstrap (Ch. 7)
   - Example: law schools
   - Parametric bootstrap
   - Semi-parametric bootstrap
3. MC methods for hypothesis testing (Ch. 8)
   - Preliminaries
   - MC tests
   - Permutation tests

MC test of a simple hypothesis

- An MC-based algorithm for testing simple hypotheses goes as follows:
  1. Draw N samples, $Y_1, \ldots, Y_N$, from the distribution specified by $\mathcal{H}_0$.
  2. Calculate the test statistic $t_i = t(Y_i)$ for each sample.
  3. Estimate the p-value using MC integration by letting

$$\widehat{p}(y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{t_i \geq t(y)\}}.$$

  4. If $\widehat{p}(y) \leq \alpha$, reject $\mathcal{H}_0$.

Outline

1. Last time: Introduction to bootstrap (Ch. 7)

2. More on the bootstrap (Ch. 7)
   - Example: law schools
   - Parametric bootstrap
   - Semi-parametric bootstrap

3. MC methods for hypothesis testing (Ch. 8)
   - Preliminaries
   - MC tests
   - Permutation tests

Permutation tests

- The random variables of a set Y is said to be exchangeable if they have the same distribution for all permutations.
- The conditional distribution of Y given the ordered sample is then the uniform distribution on the set of all permutations of Y.
- Conditioning on the ordered variables leads to permutation tests.
- Permutation tests can be very efficient in testing an exchangeable null-hypothesis against a non-exchangeable alternative, e.g. for testing if two samples differ in some way.

MC permutation test

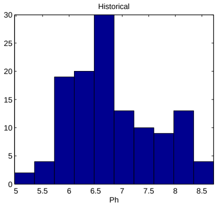
- An MC-based permutation test can be implemented as follows.
  1. Draw N permutations, $Y_1, \ldots, Y_N$, of the vector y.
  2. Calculate the test statistic $t_i = t(Y_i)$ for each permutation.
  3. Estimate the p-value using MC integration according to

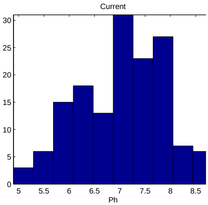
$$\widehat{p}(y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{t_i \geq t(y)\}}.$$

  4. If $\widehat{p}(y) \leq \alpha$, reject $\mathcal{H}_0$.

Example: pH data

- We have 273 historical and current pH-measurements of 149 lakes in Wisconsin and want to test if the pH-levels have increased.
- We assume that all measurements are independent and that historical measurements have a distribution F_0 and that new measurements have a distribution G_0.
- We want to test $\mathcal{H}_0 : F_0 = G_0$ against $\mathcal{H}_1 : F_0 \neq G_0$.





Example: pH data (cont.)

- Assume that the distribution for current data can be written as $G_0(y) = F_0(y - \theta)$. That is, the mean of the current data is the mean of the historical data plus θ.
- We now want to test $\mathcal{H}_0 : \theta = 0$ against $\mathcal{H}_1 : \theta > 0$.
- Under $\mathcal{H}_0$, all data are i.i.d. and thus exchangeable.
- We use the difference in the sample means as a test statistic.
- A permutation test gives $p = 0.0198$. Reject $\mathcal{H}_0$!