Bi7740: Scientific computing

Resampling methods: bootstrapping and permutations

Vlad Popovici
popovici@iba.muni.cz

Institute of Biostatistics and Analyses
Masaryk University, Brno

Supplemental bibliography

- Efron, Tibshirani: *An Introduction to the Bootstrap*. 1993. Chapman & Hall
- Good: *Resampling Methods. A Practical Guide to Data Analysis*. 2006. Birkhäuser, 3rd Ed

Outline

Outline

1. Bootstrapping
    - Introduction
    - Empirical distribution and the plug-in principle
    - Improved bootstrap confidence intervals
    - Bootstrapping for hypothesis test
2. Permutation tests
    - Introduction
    - Example/exercise

Introduction

- resampling technique for statistical inference: assess uncertainty
- especially useful when no assumptions can be made on the underlying model
- confidence intervals without distributional assumptions
- there are many versions of bootstrapping

Example (from Efron, Tibshirani, 1993):

Group	Heart attacks	Subjects
aspirin	104	11037
placebo	189	11034

The odds ratio:

$$\hat{\theta} = \frac{104/11037}{189/11034} = 0.55$$

so it seems that aspirin reduced the incidence of heart attacks.

Log-odds can be approximated by the normal distribution, so we use it to construct a 95% CI. Standard error is

$$SE(\log(OR)) = \sqrt{1/104 + 1/189 + 1/11037 + 1/11034} = 0.1228$$

giving a 95% CI for $\log\theta$:

$$\log\hat{\theta} \pm 1.96 \times SE(\log(OR)) = (-0.839, -0.357)$$

with a corresponding 95% for θ obtained by exponentiating: $(0.432, 0.700)$.

At the same time, aspirin seems to have a detrimental effect on strokes

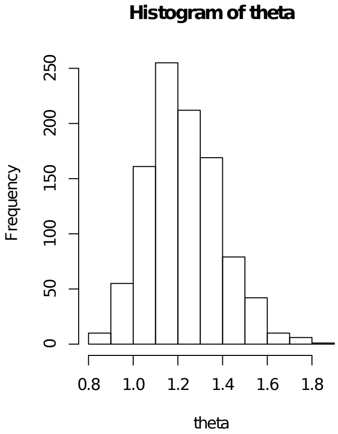
Group	Heart attacks	Subjects
aspirin	119	11037
placebo	98	11034

which leads to an odds ratio of $\hat{\theta} = 1.21$ with a 95% CI of $(0.925, 1.583)$.

...and how bootstrap would proceed to infering the CI:

- create a sample for the treatment (s_1) and one for the placebo (s_2) group as vectors containing as many 1s as case there are
- draw *with replacement* a random sample from s_1 and from s_2, of the same size as the groups
- compute the odds ratios based on the drawn samples
- repeat the process a number of times and record all the odds ratios computed
- using their empirical distribution, estimate the CI of interest

```
set.seed(1)
n1 = 11037; n1.cases = 119; n2 = 11034; n2.cases = 98
s1 = c(rep(1, n1.cases), rep(0, n1-n1.cases))
s2 = c(rep(1, n2.cases), rep(0, n2-n2.cases))
B = 1000;          # number of bootstrap samples
p = n2 / n1
theta = rep(0, B)
for (i in 1:B) {
  theta[i] = p * sum(sample(s1, n1, replace = TRUE)) /
    sum(sample(s2, n2, replace = TRUE))
}
hist(theta)
quantile(theta, probs=c(.025, .975))
     2.5%     97.5%
0.9365309 1.5711275
```



- the CI estimate by the quantiles is not the most precise nor efficient that can be obtained by bootstrapping
- it works for symmetric, close to normal distributions of the bootstrap replicate

Outline

The empirical distribution

- the underlying probability distribution F generates the observed sample:

$$F \rightarrow (x_1, \dots, x_n) = \mathbf{x}$$

- the empirical distribution $\hat{F}$ is the *discrete* distribution that puts probability $1/n$ at each value $x_i, i = 1, \dots, n$
- $\hat{F}$ assigns to a set A in the sample space of x its empirical probability:

$$\widehat{Prob}\{A\} = \frac{\#\{x_i \in A\}}{n} = Prob_{\hat{F}}\{A\}$$

- a *parameter* is a functional of the distribution function, $\theta = t(F)$. Example: the mean

$$\mu(F) = \int x dF(x)$$

- a *statistic* is a function of the sample x. Example: the sample average,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- the plug-in estimate of a parameter $\theta = t(F)$ is defined to be

$$\hat{\theta} = t(\hat{F})$$

(sometimes called summary statistics, estimates or estimator)

Bootstrap estimate of the standard error

- bootstrap sample: $\hat{F} \rightarrow (x_1^*, \ldots, x_n^*) = \mathbf{x}^*$ (resampling with replacement)
- let $\hat{\theta} = s(\mathbf{x})$ be an estimate for the parameter of interest
- the question is: what is the standard error of the estimate?
- bootstrap replication of $\hat{\theta}$ is

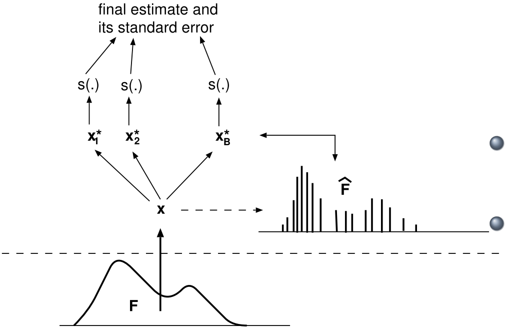
$$\hat{\theta}^* = s(\mathbf{x}^*)$$

- *ideal bootstrap estimate* of SE:

$$se_{\hat{F}}(\hat{\theta}^*)$$

  i.e. the standard error of $\hat{\theta}$ for data sets of size n randomly sampled from $\hat{F}$
- unfortunately, close-form formulas exist only for some estimators

General form of the bootstrap method



- by resampling with replacement from **x** one samples from the empirical distribution $\hat{F}$
- $\mathbf{x}_b^*$ are the bootstrap samples of size n
- $s(\mathbf{x}_b^*) = \hat{\theta}_b^*$ are the bootstrap replications of the parameter of interest θ

Bootstrap estimation of standard errors

1. select B independent bootstrap samples $\mathbf{x}_1^*, \ldots, \mathbf{x}_B^*$
2. evaluate the bootstrap replicate of each bootstrap sample $\hat{\theta}_b^* = s(\mathbf{x}_b^*)$, $b = 1, 2, \ldots, B$
3. estimate the standard error $se_{\hat{F}}(\hat{\theta})$ by the sample standard deviation of the B replications:

$$\widehat{se}_B = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\theta}_b^* - \hat{\theta}_0^*\right]^2}$$

where $\hat{\theta}_0^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*$

Implement the previous procedure in R:

- write a function `bstrap.nonparam(x, B, s, ...)` which will generate `B` bootstrap samples $\mathbf{x}_b^*$ and for each of them will compute the bootstrap replicate of the parameter: $\hat{\theta}_b^* = s(\mathbf{x}_b^*, \cdots)$
- write a function `bstrap.theta0(T)` which computes the bootstrap estimate of the parameter, given the bootstrap replicates in the vector `T` ($\hat{\theta}_0^*$)
- write a function `bstrap.`**`se`**`(T)` which computes the bootstrap estimate of the standard error of the parameter, given the bootstrap replicates in the vector `T` ($\widehat{se}_B$)
- use the `Rainfall` data set to compute the bootstrap estimate of the mean, median and corresponding standard errors
- HOMEWORK: compare with textbook results! (discuss!)

Outline

1. Bootstrapping
   - Introduction
   - Empirical distribution and the plug-in principle
   - Improved bootstrap confidence intervals
   - Bootstrapping for hypothesis test

2. Permutation tests
   - Introduction
   - Example/exercise

Bias–corrected and accelerated CI

- the quantile-based CI is not tight enough nor robust
- idea: better exploit the quantiles of the empirical distribution by:
    - correcting the bias
    - improving convergence
- simple bootstrap quantile-based CI: for an $(1-2\alpha)$ coverage, the bounds of the CI are given by $(\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)})$ where $\hat{\theta}^{*(q)}$ is the q–th quantile of the bootstrap replicates

The BCa CI is given by $(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)})$ where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right)$$

where

- $\Phi(\cdot)$ is the standard normal CDF
- $z^{(q)}$ is the q–th quantile of standard normal distribution
- $\hat{a}$ and $\hat{z}_0$ are cleverly chosen

The parameters of BCa CIs:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^*_b < \hat{\theta}\}}{B}\right)$$

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right]^{3/2}}$$

where

- $\hat{\theta}_{(i)}$ is the value of the parameter computed on the vector **x** with the *i*–th component removed (*jackknife values* of the parameter)
- $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$

Exercise: implement the BCa procedure in R:

- write a function `bstrap.bca(x, B, s, ..., alpha=c(0.025, 0.05))` that returns the low and upper bounds of the CI computed by BCa method
- you can use (call) the previous function `bstrap.nonparam`
- compute the 90% and 95% BCa CIs for the mean of `Rainfall` data: `bstrap.bca(Rainfall, 2000, mean)`

Important properties of BCa CIs

- *transformation respecting*: the bounds of the CIs transform correctly if the parameter is changed by some function: e.g. the CIs for $\sqrt{\cdot}$-transformed parameter are obtained by taking $\sqrt{}$ of the bounds of the parameter itself
- *second order accurate*: convergence rate of $1/n$ to true coverage

Outline

Bootstrapping for tests

- consider two possibly different distributions F and G,

$$\begin{aligned} F &\rightarrow \mathbf{z} = (z_1, \dots, z_n) \\ G &\rightarrow \mathbf{y} = (y_1, \dots, y_m) \end{aligned}$$

- hypotheses:

$$\begin{aligned} H_0 &: F = G \\ H_1 &: F \neq G \end{aligned}$$

- $F = G \Leftrightarrow Prob_F\{A\} = Prob_G\{A\}$ for all sets A
- *observe* a test statistic $\hat{\theta}$ (e.g. mean difference)
- *achieved significance level (ASL)*: probability of observing that large a value under H_0:

$$ASL = Prob_{H_0}\{\hat{\theta}^* \geq \hat{\theta}\}$$

Bootstrapping hypothesis testing procedure

1. choose a test statistic (not necessary a parameter): $t(\mathbf{x})$ (for example: $t(\mathbf{x}) = \bar{\mathbf{z}} - \bar{\mathbf{y}}$)
2. draw B samples of size $n + m$ from $\mathbf{x} = (\mathbf{z}, \mathbf{y})$ and call the first n observations $\mathbf{z}^*$ and the remaining m $\mathbf{y}^*$
3. evaluate $t(\cdot)$ for each sample: $t(\mathbf{x}_b^*)$
   (for example
   $$t(\mathbf{x}_b^*) = \bar{\mathbf{z}}_b^* - \bar{\mathbf{y}}_b^*$$
)
   for $b = 1, 2, \ldots, B$
4. approximate ASL_{boot} by
   $$\widehat{ASL}_{boot} = \#\{t(\mathbf{x}_b^*) \geq t(\mathbf{x})\}/B$$

Exercise:

- consider the data vectors `mouse.c` and `mouse.t` for the *control* and *treatment* arms of an experiment (some clinical variable)
- implement the bootstrap hypothesis testing procedure
- use the test statistic

$$t(\mathbf{x}) = \frac{\bar{z} - \bar{y}}{\bar{\sigma}\sqrt{1/n + 1/m}}$$

where

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(z_i - \bar{z})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2}{(n+m-2)}}$$

Implementations in R

- many R packages implement various bootstrapping procedures
- `bootstrap` package contains data and functions accompanying the book by Efron and Tibshirani
- `boot` package contains *a lot* of well tested and documented functions

Outline

1. Bootstrapping
   - Introduction
   - Empirical distribution and the plug-in principle
   - Improved bootstrap confidence intervals
   - Bootstrapping for hypothesis test
2. Permutation tests
   - Introduction
   - Example/exercise

Outline

1. Bootstrapping
   - Introduction
   - Empirical distribution and the plug-in principle
   - Improved bootstrap confidence intervals
   - Bootstrapping for hypothesis test

2. Permutation tests
   - Introduction
   - Example/exercise

Permutation tests

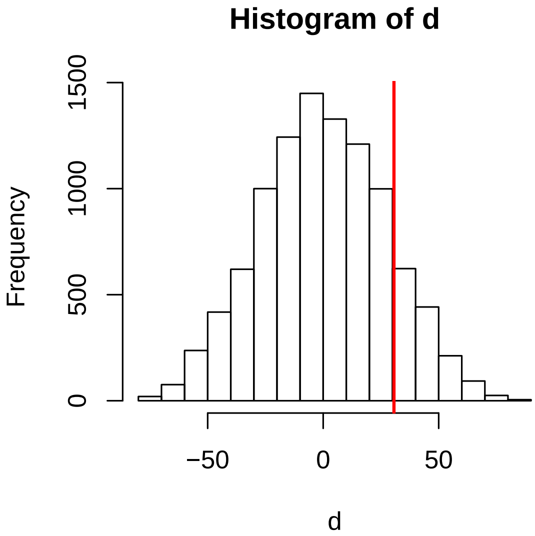
- nonparametric testing procedure
- allow testing hypotheses when the properties of the test statistic under the null hypothesis are not known
- do not make assumptions on the data
- work on small data sets
- idea: generate the distribution of the test statistic under the null hypothesis *from the data*

- exact permutation tests: for (very) small data sets, generate *all* permutations and compute the corresponding test statistics
- random test: for large data sets, generate a number of random permutations, for which compute the test statistic
- test procedure: count how many times the test statistic from the permutations is more extreme than the real test statistic and reject H_0 if the proportion is below the predefined α–level

Outline

Example - two populations tests

- consider the data vectors `mouse.c` and `mouse.t` for the *control* and *treatment* arms of an experiment (some clinical variable)
- implement a permutation testing procedure for testing
  H_0 : there is no significant difference in the clinical variable between control and treatment
  vs
  H_1 : there is a significant difference in the clinical variable between control and treatment
- which test statistic? what to permute? how many permutations?
- what should be changed if the test was about superiority of treatment vs control?


Histogram of d
Frequency
0
500
1000
1500
−50
0
50
d