Monte Carlo experiments and resampling methods

Statistical Programming with R

Universiteit Utrecht

Simulation, computer intensive statistical methods

- classical statistical methods are largely based on idealized assumptions (e.g. linear relations, normal distribution)
- the application of statistical methods to nonlinear, non-gaussian problems required the adoption of alternative statistical tools
- the advances in computational power of todays computers have enabled newer and more complicated statistical methods
- a new perspective in statistics: hard theoretical analysis is replaced by computationally intensive but simpler methods
- Monte Carlo simulation and resampling techniques are increasing in popularity.

Why use simulation?

- if classical inference theory fails, statistical inference is still possible using simulation to obtain standard errors, confidence intervals, and test hypotheses (example: median)
- to compare the quality of estimators
- to evaluate performance of models, algorithms
- to validate inferential and resampling methods
- to assess and select models

Monte Carlo as an experimental design

Monte Carlo method is a designed experiment

Criterion:

- Fit measure

Explanatory variables:

- structure of the data
- methods, algorithms

	F1			
F2		1	2	3
	1	r	r	r
	2	r	r	r
	3	r	r	r
	3	r	r	r

Monte Carlo simulation vs Resampling

Monte Carlo simulation

Sample from:

- distribution
- true model

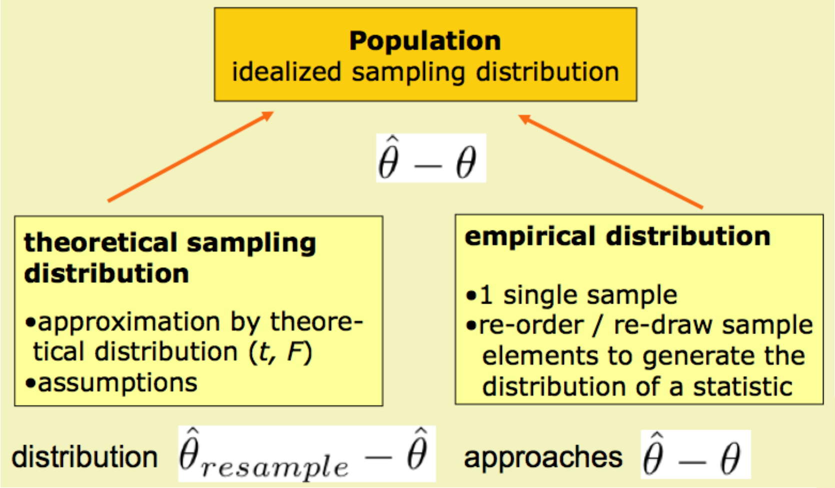
Strong assumptions on distribution of the data

Resampling

- **Re**sample from original sample
- no information on population characteristics or distribution

No assumptions on distribution of the data

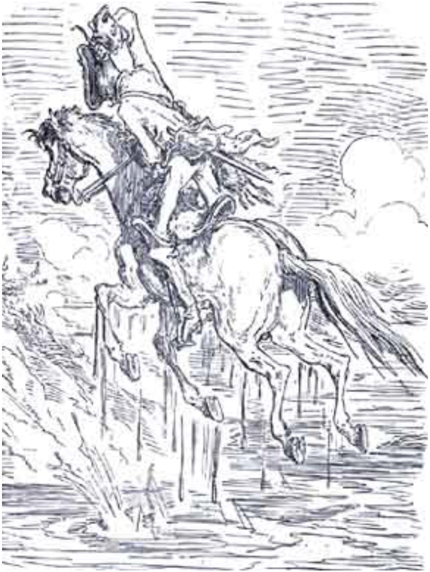
Paper Rodgers (1999)



Key concepts

- estimator
- bias and variance of an estimator (mean squared error)
- hypothesis testing
- confidence interval
- model
- model assessment
- model selection

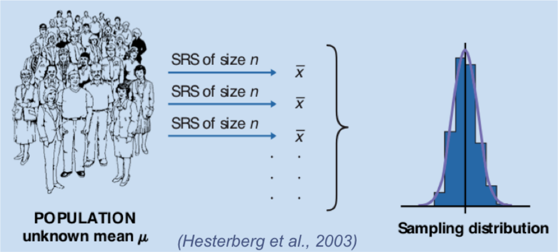
A resampling technique: the bootstrap

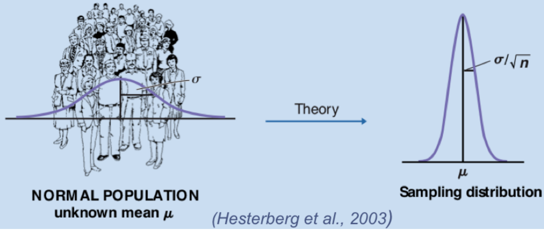


Baron Münchausen has fallen in a lake and pulls himself and his horse up by his bootstrap (or hair, depending on the version)

Rudolph Erich Raspe, *Münchausen* (1763)

Probability theory



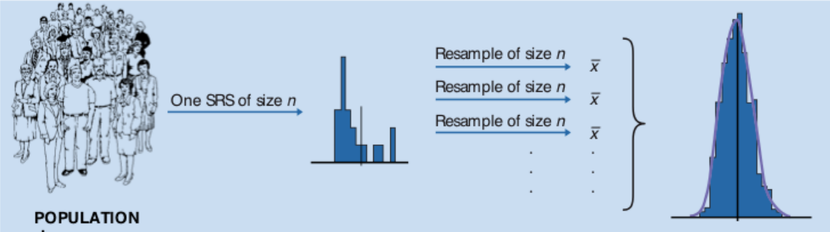


Take many random samples from population and compute mean (sampling distribution of the mean)

Probability theory shortcut:
assume population follows normal distribution
sampling distribution of the mean approaches normal

Universiteit Utrecht

Basic idea bootstrap resampling

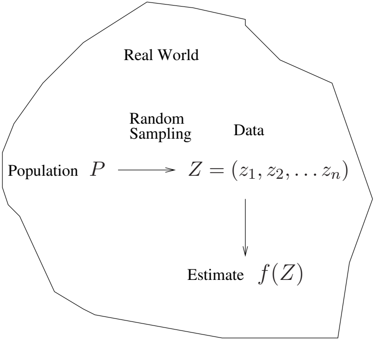


(Hesterberg et al., 2003)

The bootstrap idea

When there is **only one sample** and theory fails:

- the sample stands for the population
- the bootstrap distribution stands for sampling distribution
- substitute computing power for theoretical analysis


Real World
Random Sampling
Data
Population P
$Z = (z_1, z_2, \dots z_n)$
Estimate $f(Z)$
Bootstrap World
Random Sampling
Bootstrap dataset
Estimated Population $\hat{P}$
$Z^* = (z_1^*, z_2^*, \dots z_n^*)$
Bootstrap Estimate $f(Z^*)$

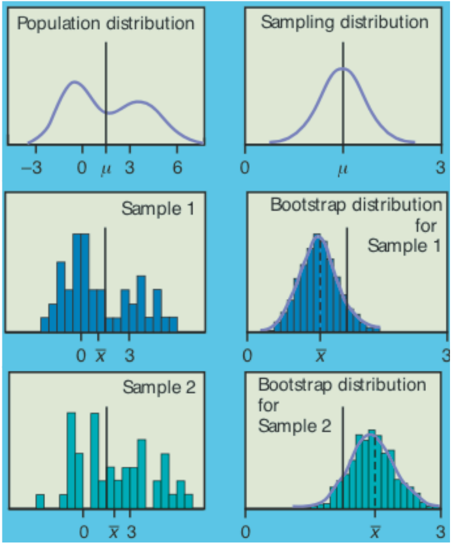
Basic idea bootstrap resampling

Use resampling to estimate how the sample statistic of a sample of size *n* from a certain population varies because of random sampling.

The bootstrap distribution does not reveal the center of the sampling distribution directly, but reveals the bias.

A statistic is biased when its sampling distribution is not centered at the true value of the parameter

How resampling works



The sampling distribution is centered at population mean μ

The bootstrap distributions are centered close to value of the sample mean $\bar{x}$

Almost all the variation in bootstrap distributions comes from selection of original sample

The shape and spread of bootstrap distributions resemble those of sampling distribution

Universiteit Utrecht

Bias and random sampling error

Sampling distribution of a statistic:

centered at true value + bias

bias = mean sampling distribution - true value

$= \hat{\theta} - \theta$

$= \bar{x} - \mu$

Bootstrap distribution of a statistic:

centered at sample value + bias

bias = mean bootstrap distribution - mean original sample

$= \hat{\theta}_{\text{resampling}} - \hat{\theta}$

$= \frac{1}{B} \sum_{b=1}^{B} \bar{x}^{\star} - \bar{x}$

($\bar{x}^{\star}$ = B replications of the statistic)

The two biases are similar even though the centers are not

Universiteit Utrecht

Use bootstrap distribution

The *bootstrap* is used for *estimation* to obtain the following measures of a statistic:

- bias ($\hat{\theta}_{\text{resampling}} - \hat{\theta}$):

$$\frac{1}{B}\sum_{b=1}^{B} \bar{x}^{\star} - \bar{x}$$

- standard error = standard deviation of bootstrap distribution

$$\hat{\sigma}_b(\hat{\theta}) = \hat{\sigma}_{b,\bar{x}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\bar{x}^{\star} - \frac{1}{B}(\bar{x}^{\star})\right)^2}$$

- confidence interval (percentile method)
  quantile(bootstrap distribution, p = c(0.025, 0.975))

RMSE: bias and variance combined

The *root mean squared error* of a statistic is a variability measure that combines bias and variance (*cf.* Efron & Gong, 1983):

$$\begin{aligned}\sqrt{\text{MSE}} &= \sqrt{E_F[(\hat{\theta} - \theta)]} \\ &= \sqrt{bias^2 + st.dev^2} \\ &= \sqrt{(\hat{\theta} - \theta)^2 + \hat{\sigma}_b(\hat{\theta})^2}\end{aligned}$$

Also useful to compare the distributions of a statistic obtained with several methods

Universiteit Utrecht

Resampling methods

Resampling methods can be used for different goals:

1. estimation (bias, standard errors, confidence intervals)
   - bootstrap
   - jacknife
2. hypothesis testing
   - bootstrap
   - permutation
3. model assessment and selection
   - cross-validation

Bootstrap vs Cross-validation

- In cross-validation, each of the K validation folds is distinct from the other $K-1$ folds used for training: *there is no overlap*. This is crucial for its success.
- To estimate prediction error using the bootstrap, we could think about using each bootstrap dataset as our training sample, and the original sample as our validation sample.
- But each bootstrap sample has significant overlap with the original data. About two-thirds of the original data points appear in each bootstrap sample.
- This will cause the bootstrap to seriously underestimate the true prediction error.