An Introduction to the Bootstrap

Hannelore Liero

Introduction: What is Bootstrap?

- **Simulation-based statistical analysis**

- Aim: Replace complicated and often inaccurate approximations to bias, variances, confidence intervals and distributions by computer simulations

  key idea: „**resample from the original data**“

- create replicate data sets from which the variability of the quantities of interest can be assessed

  directly: samples from the original data

  indirectly: samples from a fitted model

Introduction: What is Bootstrap?

- Aim

  - **explanation of the basic ideas**
  - application to simple problems
  - **realization in R**

  R is a system for statistical analysis and graphics, it is both software and a language (http://cran.r-project.org)

- References

  Davison, A. C., Hinkley, D. V.: Bootstrap Methods and their Application, Cambridge University Press 1997

  Efron, B. and Tibshirani R. J.: An Introduction to the Bootstrap, Chapman & Hall 1993

Introduction: Estimation of the mean

- simplest statistical problem:

  estimation of a mean μ of a r. v. by the arithmetic mean

- strong mathematical formulation:

  $\boldsymbol{x} = (x_1, \ldots, x_n)$ be a sample from a population with cumulative distribution function (cdf) F

  population mean $\quad \mu = \int z \, \mathrm{d}F(z) = t(F)$

  estimate $\quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \int z \, \mathrm{d}\hat{F}_n(z) = t(\hat{F}_n)$

  $\hat{F}_n$ is the empirical distribution function (edf) of the data $\boldsymbol{x}$;
  $t(\cdot)$ is a statistical function

Introduction: The empirical distribution function

- $$\hat{F}_n(z) = \frac{\text{number of } x_i \leq z}{n} = \frac{1}{n}\sum_{i=1}^{n} 1(x_i \leq z)$$

- Note that for an arbitrary function g we have

$$\mathsf{E}g(x) = \int g(z)\,\mathrm{d}F(z) \qquad\qquad \int g(z)\,\mathrm{d}\widehat{F}_n(z) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

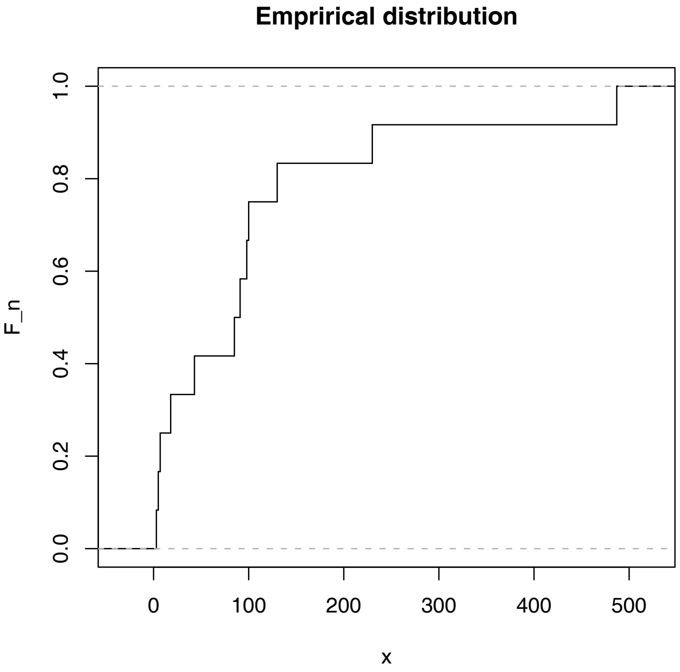
- Many estimates (not only the mean) are constructed by the **plug-in principle**:

  The parameter $\theta = t(F)$ is estimated by $\quad \hat{\theta} = t(\hat{F}_n)$.

Introduction: The empirical distribution function

Numerical example: *The following table gives times of $n = 12$ failures of air-conditioning equipment in a Boing 720 jet.*

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	3	5	7	18	43	85	91	98	100	130	230	487



$$\overline{x} = 108.08$$

Introduction: Properties of the arithmetic mean

- Arithmetic mean is an **unbiased estimator** of μ: $\quad \mathsf{E}_F \bar{x} = \mu$
- We can compute the **variance** of $\bar{x}$, but it is unknown:

$$\mathsf{Var}_{\boldsymbol{F}}\, \boldsymbol{\bar{x}} = \boldsymbol{\sigma^2/n,}$$

where

$$\sigma^2 = \mathsf{Var}_F\, x_1 = \int z^2\, \mathrm{d}F(z) - \left(\int z\, \mathrm{d}F(z) \right)^2 =: \sigma^2(F)$$

- Estimate of $\mathsf{Var}_{\boldsymbol{F}}\, \boldsymbol{\bar{x}}$: $\quad \boldsymbol{s^2/n}$

well-known **empirical variance:** $\quad s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Introduction: Properties of the arithmetic mean

- theoretical point of view:

  $\boldsymbol{s^2}$ is the biased-corrected plug-in-estimate: $\boldsymbol{F \longleftarrow \hat{F}_n}$

$$
\begin{aligned}
\sigma^2_F &= \int z^2 \,\mathrm{d}F(z) - \left(\int z \,\mathrm{d}F(z)\right)^2 = \sigma^2(F) \\
\widehat{\sigma^2_F} &= \int z^2 \,\mathrm{d}\hat{F}_n(z) - \left(\int z \,\mathrm{d}\hat{F}_n(z)\right)^2 = \sigma^2(\hat{F}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sigma^2_{\hat{F}_n}
\end{aligned}
$$

- In other words: Let $\boldsymbol{x}$ be **fixed** and $\boldsymbol{x^*}$ be distributed according to the **distribution function $\hat{F}_n$**, then $\boldsymbol{\sigma^2_{\hat{F}_n}}$ can be considered as the **variance of x^***.

Mean in a parametric model

- Let data x_i be drawn from an **exponentially** distributed population

$$x_i \sim \mathrm{Exp}(1/\mu), \quad \text{i.e. } F = F_\mu; \quad F_\mu(z) = 1 - \exp(-z/\mu)$$

$$\mathsf{E}_{F_\mu} x_1 = \mu = \int z \,\mathrm{d}F_\mu(z) \qquad\qquad \hat{\mu} = \bar{x}$$

- unknown variance $\mathsf{Var}_{\boldsymbol{F_\mu}} \boldsymbol{\bar{x} = n^{-1}\mu^2}$ is estimated by $\boldsymbol{n^{-1}\bar{x}^2}$

- In the air-condition example: Suppose that the failure times are exponentially distributed, then

  $\overline{x} = 108, 08, \qquad s^2/n \approx 1546 \qquad \overline{x}^2/n \approx 973$

  The variance of the estimator is estimated by 1546, if we have no parametric assumption, and by 973 by assuming the exponential model.

Mean in a parametric model

- theoretical point of view:

  $\bar{\boldsymbol{x}}^{\mathbf{2}}$ is the plug -in-estimate: $\boldsymbol{F_{\mu}} \longleftarrow \boldsymbol{F_{\bar{x}}}$

  $\boldsymbol{F_{\bar{x}}}$ is the **fitted distribution function**

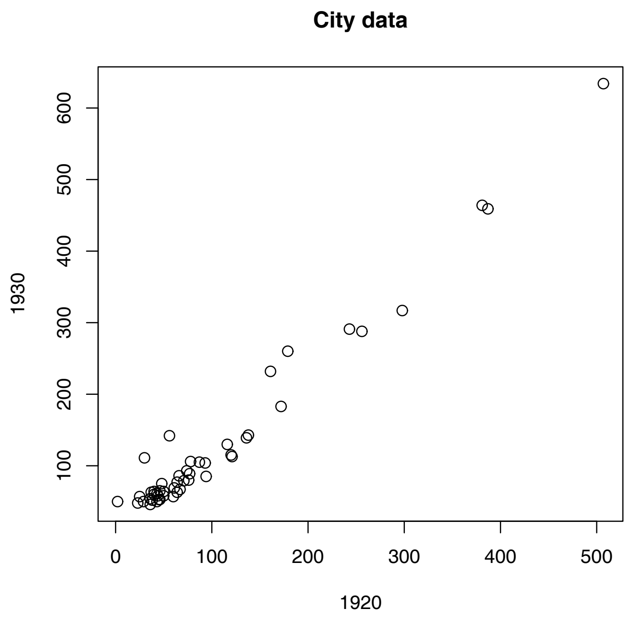
$$\begin{aligned} \sigma^2_{F_\mu} &= \mu^2 = \sigma^2(F_\mu) = \mathsf{Var}_{F_\mu} x \\ \widehat{\sigma^2_{F_\mu}} &= \bar{x}^2 = \sigma^2(F_{\bar{x}}) = \mathsf{Var}_{F_{\bar{x}}} x^* \end{aligned}$$

  Let $\boldsymbol{x}$ be **fixed**, then $\boldsymbol{\sigma^2(F_{\bar{x}})}$ is the **variance** of a r.v. $\boldsymbol{x}^*$, which is **exponentially distributed with expectation** $\bar{\boldsymbol{x}}$.

- until now: no new results, only another point of view

Estimation of the ratio of two means

- Example: City Data Given 49 data pairs, each corresponding to a US city, the pair being the 1920 and 1930 populations of the city. Interest here is the ratio of means, because this would enable us to estimate the total population of the US in 1930 from the 1920 figure.



Estimation of the ratio of two means

- Let the data be realizations of i.i. d. random pairs (u_i, v_i) with cdf F and marginal cdf F_1 and F_2, then

$$\boldsymbol{\theta} = \frac{\mathsf{E}_{\boldsymbol{F}} \boldsymbol{v}_1}{\mathsf{E}_{\boldsymbol{F}} \boldsymbol{u}_1} \quad \text{and} \quad \hat{\boldsymbol{\theta}} = \frac{\bar{\boldsymbol{v}}}{\bar{\boldsymbol{u}}} \qquad \text{plug-in-estimate}$$

- Expectation and variance of $\dfrac{\bar{\boldsymbol{v}}}{\bar{\boldsymbol{u}}}$??

  Theoretical calculations are possible only under additional model assumptions (if at all!!!)

  Solution: Application of the central limit theorem which gives an asymptotic variance

  Justified?

City data

	u	v
1	138	143
2	93	104
3	61	69
4	179	260
5	48	75
6	37	63
7	29	50
8	23	48
9	30	111
10	2	50
11	38	52
12	46	53
13	71	79
14	25	57
15	298	317
16	74	93
17	50	58
18	76	80
19	381	464
20	387	459
21	78	106
22	60	57
23	507	634
24	50	64
25	77	89
26	64	77
27	40	60
28	136	139
29	243	291
30	256	288
31	94	85
32	36	46
33	45	53
34	67	67
35	120	115
36	172	183
37	66	86
38	46	65
39	121	113
40	44	58
41	64	63
42	56	142
43	40	64
44	116	130
45	87	105
46	43	61
47	43	50
48	161	232
49	36	54

Bootstrap estimates of bias and variance

- General approach:

  Data $\boldsymbol{x} = (x_1, \ldots, x_n)$ from a population with cdf F are observed and wewish to estimate the parameter of interest $\boldsymbol{\theta = t(F)}$ on the basis of $\boldsymbol{x}$.

- We compute the estimate $\boldsymbol{\hat{\theta} = s(x)}$.

  ($s(\boldsymbol{x})$ may be the plug-in estimate $t(\hat{F}_n)$, but doesn't have to be.)

  How accurate is $\boldsymbol{\hat{\theta}}$?

- Accuracy is measured by the bias and the variance:

$$\text{Bias} \quad = \boldsymbol{\beta} = \mathrm{E}(\boldsymbol{\hat{\theta}}|\boldsymbol{F}) - \boldsymbol{t(F)} \qquad\qquad \boldsymbol{\nu} = \mathrm{Var}(\boldsymbol{\hat{\theta}}|\boldsymbol{F})$$

We write $\mathrm{E}(\hat{\theta}|F)$ to mean that the r.v.'s from which $\hat{\theta}$ is calculated have cdf F.

Bootstrap estimates of bias and variance

- $\boldsymbol{\beta} = \mathsf{E}(\hat{\boldsymbol{\theta}}|\boldsymbol{F}) - \boldsymbol{t}(\boldsymbol{F}) = \boldsymbol{b}(\boldsymbol{F}) \qquad \boldsymbol{\nu} = \mathsf{Var}(\hat{\boldsymbol{\theta}}|\boldsymbol{F}) = \boldsymbol{v}(\boldsymbol{F})$

  Dependence of β and ν from F is expressed by the functionals b and v.

- Let $\tilde{\boldsymbol{F}}$ be an estimate of $\boldsymbol{F}$ (empirical df or parametric fit), then estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ are given by

$$\boldsymbol{B} = \boldsymbol{b}(\tilde{\boldsymbol{F}}) \qquad \boldsymbol{V} = \boldsymbol{v}(\tilde{\boldsymbol{F}})$$

- B and V are called **ideal bootstrap estimates** of β and ν.

Bootstrap estimates of bias and variance

- This is equivalent to: We have a sample $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)$ with distribution $\tilde{F}$ (which is completely known!) and compute the estimate

  $$\boldsymbol{B} = \boldsymbol{b}(\tilde{\boldsymbol{F}}) = \mathsf{E}(\hat{\boldsymbol{\theta}}^*|\tilde{\boldsymbol{F}}) - \boldsymbol{t}(\tilde{\boldsymbol{F}}) \qquad \boldsymbol{V} = \boldsymbol{v}(\tilde{\boldsymbol{F}}) = \mathsf{Var}(\hat{\boldsymbol{\theta}}^*|\tilde{\boldsymbol{F}}).$$

  The star notation indicates that $\boldsymbol{x}^*$ is not the actual data set $\boldsymbol{x}$, but rather a resampled version.

  $\hat{\boldsymbol{\theta}}^* = \boldsymbol{s}(\boldsymbol{x}^*)$ is the bootstrap replication of $\hat{\theta}$; it is the result of applying the same function $s(\cdot)$ to $\boldsymbol{x}^*$ as was applied to $\boldsymbol{x}$.

- In the example of estimating the mean we know the functions b and v. Thus we can compute these ideal bootstrap estimates.

Estimates of bias and variance **Bootstrap algorithm**

- But, unfortunately, for virtually other estimate $\hat{\theta}$ other than the mean, there is no formula for the functionals b and v; thus we cannot compute B and V exactly.

  $\Rightarrow$ these quantities are computed by simulations!

  A bootstrap algorithm is a computational way of obtaining good approximations to the numerical values of B and V.

- Simulations: Given our **original sample** x of sample size n we **generate** R **samples** $x^{*r} = (x_1^{*r}, \dots, x_n^{*r})$ of size n, where each component is distributed **according to** $\tilde{F}$, $\quad r = 1, \dots, R$.

Estimates of bias and variance **Bootstrap algorithm**

- Generate $\boldsymbol{x}^{*1} = (x_1^{*1}, \ldots, x_n^{*1})$ from $\tilde{F}$ compute $\hat{\theta}^*(1) = s(\boldsymbol{x}^{*1})$

- Generate $\boldsymbol{x}^{*2} = (x_1^{*2}, \ldots, x_n^{*2})$ from $\tilde{F}$ compute $\hat{\theta}^*(2) = s(\boldsymbol{x}^{*2})$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

- Generate $\boldsymbol{x}^{*R} = (x_1^{*R}, \ldots, x_n^{*R})$ from $\tilde{F}$ compute $\hat{\theta}^*(R) = s(\boldsymbol{x}^{*R})$

- we obtain

$$\boldsymbol{\hat{\theta}^*(1),\ \hat{\theta}^*(2),\ \ldots,\ \hat{\theta}^*(R)}$$

The ideal bootstrap bias $\boldsymbol{B} = \mathsf{E}(\boldsymbol{\hat{\theta}^*}|\boldsymbol{\tilde{F}}) - \boldsymbol{t}(\boldsymbol{\tilde{F}})$ is approximated by

$$\boldsymbol{\hat{B}_R = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}^*(r) - t(\tilde{F}) = \hat{\theta}^*(\cdot) - t(\tilde{F})}$$

Bootstrap algorithm

- The ideal bootstrap variance $V = \text{Var}(\hat{\theta}^* | \tilde{F})$ is approximated by

$$\hat{V}_R = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\theta}^*(r) - \hat{\theta}^*(\cdot))^2 \qquad \hat{\theta}^*(\cdot) = \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}^*(r)$$

- Choice of R will be discussed later.

Parametric Simulations

- Suppose we have a **parametric model**, i.e. $\boldsymbol{x}$ is drawn from a population with cdf $\boldsymbol{F_\gamma}$.

- Based on (the original) $\boldsymbol{x}$ we construct an estimate $\hat{\gamma}$ for γ.

- The fitted $\boldsymbol{F_{\hat{\gamma}}}$ is completely known. Set $\boldsymbol{\tilde{F} = F_{\hat{\gamma}}}$.

- We consider now 3 examples of parametric simulation. The first one is only for demonstration purposes, the second is somewhat artificial and the third one is (hopefully) of real interest.

Parametric Simulations: Air-condition”-Example 1

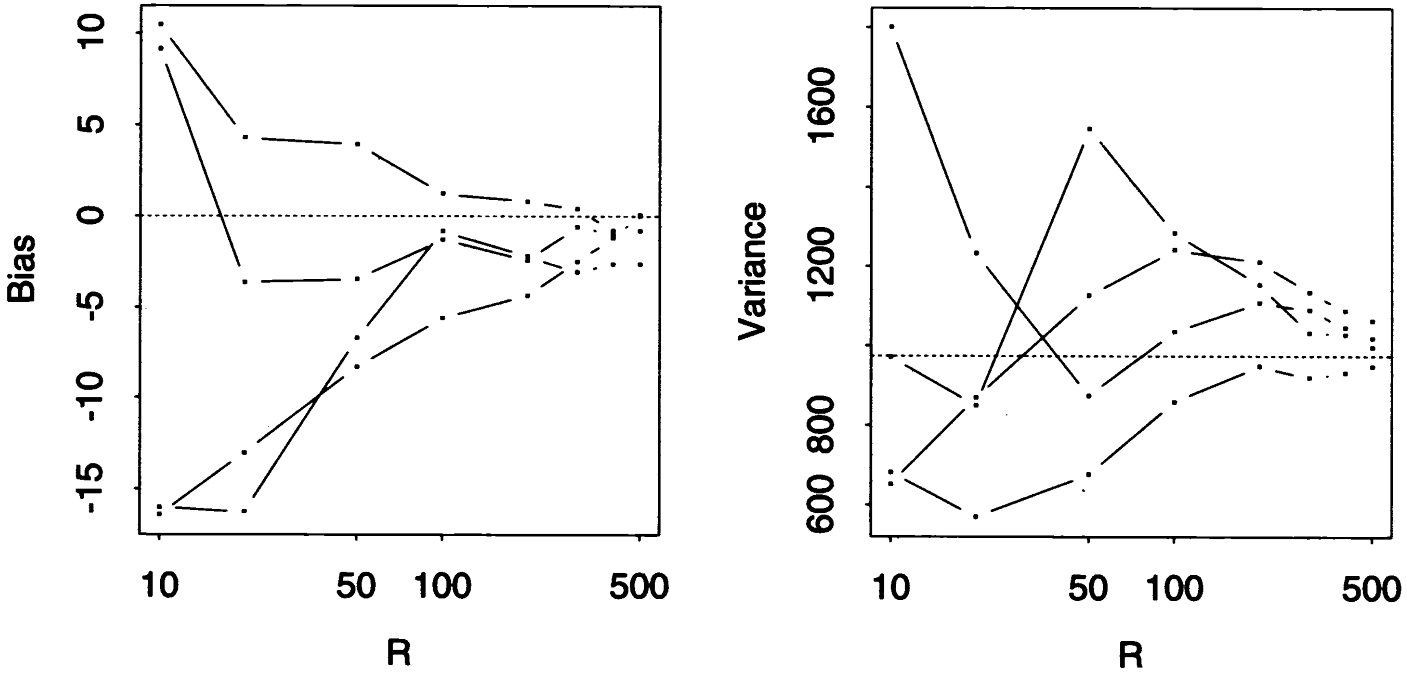
- Here $\gamma = \mu = \theta$ is the population mean. The estimate of the parameter is $\bar{x} = 108.08$. In this case we know the ideal bootstrap estimates: $B = 0$ and $V = \bar{x}^2/n = 973.5$.

- For demonstration purposes we compute the approximations: We generate $\boldsymbol{R = 50}$ samples of **exponentially distributed random numbers with expectation** 108.08. For each sample we calculate the mean $(\boldsymbol{\bar{x}^*(r)})$. The arithmetic mean of these 50 estimates is 101.1 and the resulting approximation is:

  $\hat{B}_{50} = -7 \qquad \hat{V}_{50} = 22.27^2 \approx 496.$ The R-procedure is given in the Appendix.

- On the following page you see a copy from the book of Davison/Hinkley, showing the results of four repetitions of parametric simulations with increasing R.

Figure 2.1 Empirical biases and variances of $\bar{Y}^*$ for the air-conditioning data from four repetitions of parametric simulation. Each line shows how the estimated bias and variance for $R = 10$ initial simulations change when further simulations are successively added. Note how the variability decreases as the simulation size increases, and how the simulated values converge to the exact values under the fitted exponential model, given by the horizontal dotted lines.



Parametric Simulations: Air-condition”-Example 2

- The parameter of interest is not the mean, but $\boldsymbol{\theta = \log \mu}$.

  We have:

  $\theta \;=\; \log \int z \, \mathrm{d}F_{\mu}(z) \;=\; t(F_{\mu}) \qquad \hat{\theta} \;=\; \log \int z \, \mathrm{d}F_{\bar{x}}(z) \;=\; t(F_{\bar{x}}) \;=\; \log \bar{x}$

  $\hat{\theta}$ is not unbiased; the ideal bootstrap estimate of the bias is

$$\boldsymbol{B \;=\; \int \cdots \int \log(\frac{1}{n}\sum_{i=1}^{n} z_i) \, \mathrm{d}F_{\overline{x}}(z_1) \cdots \mathrm{d}F_{\overline{x}}(z_n) \;-\; \log \overline{x},}$$

  which is difficult to compute. Hence we will approximate it by

$$\boldsymbol{\hat{B}_R = \frac{1}{R}\sum_{r=1}^{R} \hat{\theta}^*(r) \;-\; \log \overline{x}} \qquad \text{with} \quad \boldsymbol{\hat{\theta}^*(r) \;=\; \log \overline{x}^*(r)}.$$

Parametric Simulations: Air-condition"-Example 2

The variance is estimated by

$$\hat{V}_R = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\theta}^*(r) - \hat{\theta}^*(\cdot))^2$$

$$\hat{\theta}^*(r) = \log \overline{x}^*(r), \qquad \hat{\theta}^*(\cdot) = \frac{1}{R} \sum_{r=1}^{R} \log \overline{x}^*(r)$$

The corresponding R-procedure is given in the Appendix.

Parametric Simulations: Correlation in a normal distribution

- Let $(\boldsymbol{x}, \boldsymbol{y}) = ((x_1, y_1), \ldots, (x_n, y_n))$, be a sample from a **bivariate normal distribution** with parameter

$$\gamma = (\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1^2}, \boldsymbol{\sigma_2^2}, \boldsymbol{\rho}),$$

where μ_j are the expectations, σ_j^2 are the variances and ρ is the correlation of the components. The parameter of interest is the correlation ρ. The plug- in estimator based on the original data is

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \ \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Again, we replace the ideal bootstrap estimates for the bias and the variance by bootstrap sampling.

Parametric Simulations: Correlation in a normal distribution

- The bootstrap samples $(\boldsymbol{x}^*, \boldsymbol{y}^*)^r$ are generated according a bivariate normal distribution with expectations $\bar{x}$ and $\bar{y}$, variances s_x^2 and s_y^2 and correlation $\hat{\rho}$.

- From each sample we calculate the correlation $\hat{\boldsymbol{\rho}}^*(\boldsymbol{r})$. The ideal bootstrap estimates are approximated by

$$\hat{B}_R = \frac{1}{R} \sum_{r=1}^{R} \hat{\rho}^*(r) - \hat{\rho} \qquad \text{and}$$

$$\hat{V}_R = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\rho}^*(r) - \hat{\rho}^*(\cdot))^2, \qquad \hat{\rho}^*(\cdot) = \frac{1}{R} \sum_{r=1}^{R} \hat{\rho}^*(r)$$

The corresponding R-procedure is given in the Appendix.

Nonparametric Simulations

- We do not suppose a parametric model. The **bootstrap samples** are generated according to the empirical distribution $\hat{F}_n$ of the original data $(\tilde{F} = \hat{F}_n)$.

- That is, the data points $x_1^*, x_2^*, \ldots, x_n^*$ of a bootstrap sample $\boldsymbol{x}^*$ are a random sample of size n **drawn with replacement** from the population of n objects $x_1, x_2, \ldots, x_n$, i.e. for all i and j: $\mathsf{P}(x_j^* = x_i) = 1/n$

- The bootstrap data set $(x_1^*, x_2^*, \ldots, x_n^*)$ consists of **members of the original data set**, some appearing zero times, some appearing once, some appearing twice, etc.

- As in the parametric approach we generate R bootstrap samples to calculate the approximations $\hat{B}_R$ and $\hat{V}_R$ of the ideal bootstrap estimates B and V, respectively.

Nonparametric Simulations: City-Example

- For simplicity of presentation let us consider only the first 10 pairs of the observations. Also, for simplicity we draw only $R = 9$ bootstrap samples.

- The following **table** shows the frequencies with which each original data pair appears in each of the nine nonparametric bootstrap samples. Furthermore, the estimates $\hat{\theta}^*(r)$ are given.

- For example: In the 7th bootstrap sample the data pairs 1,6,7,8,10 are drawn once, the pair 4 is drawn twice and pair 9 is drawn three times. The pairs 2, 3 and 5 do not appear in this bootstrap sample. The ratio of the means in this sample is 1.783.

  As an estimate for the bias we obtain $\hat{B}_9 = 0.03$, and for the variance $\hat{V}_R = 0.3^2$.

Nonparametric Simulations: City-Example

Number of times each pair sampled $R = 9 \quad n = 10$

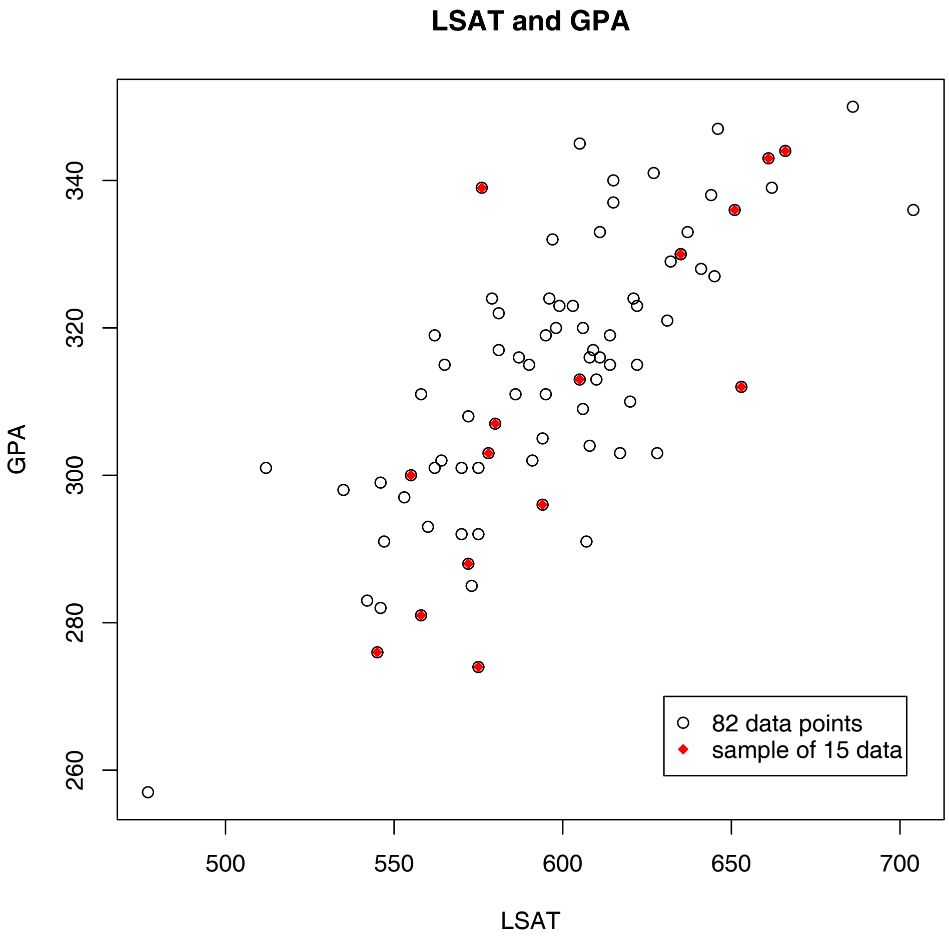
Data	1	1	1	1	1	1	1	1	1	1	$\hat{\theta} = 1.520$
Replicate r											$\hat{\theta}^*$
1	0	0	0	1	3	0	1	3	2	0	1.873
2	2	2	0	0	1	1	2	1	1	0	1.354
3	4	1	0	1	0	2	1	0	1	0	1.278
4	0	0	2	0	2	1	0	1	1	3	2.102
5	2	2	0	2	0	0	2	1	1	0	1.367
6	1	4	1	0	1	2	0	1	0	0	1.225
7	1	0	0	2	0	1	1	1	3	1	1.783
8	0	3	2	0	0	3	1	1	0	0	1.307
9	1	0	2	1	1	1	0	1	3	0	1.664

Nonparametric Simulations: Correlation

- The following example is taken from the book of Efron and Tibshirani: *The law school data. A random sample of size $\boldsymbol{n = 15}$ was taken from the collection of $\boldsymbol{N = 82}$ American law schools participating in a large study of admission practices. Two measurements were made on the entering classes of each school in 1973:* ***LSAT****, the average score for the class on a national law test, and* ***GPA****, the average undergraduate grad-point average for the class.*

- This example is artificial because these data are available for the entire population. So we can check how good our analysis based on the sample of size $n = 15$ is.

- The parameter of interest is the **correlation** $\boldsymbol{\rho}$ between LSAT and GPA. The estimate for this parameter based on the sample is $\hat{\rho} = 0.776$


LSAT and GPA
GPA
LSAT
260
280
300
320
340
500
550
600
650
700
○ 82 data points
◆ sample of 15 data

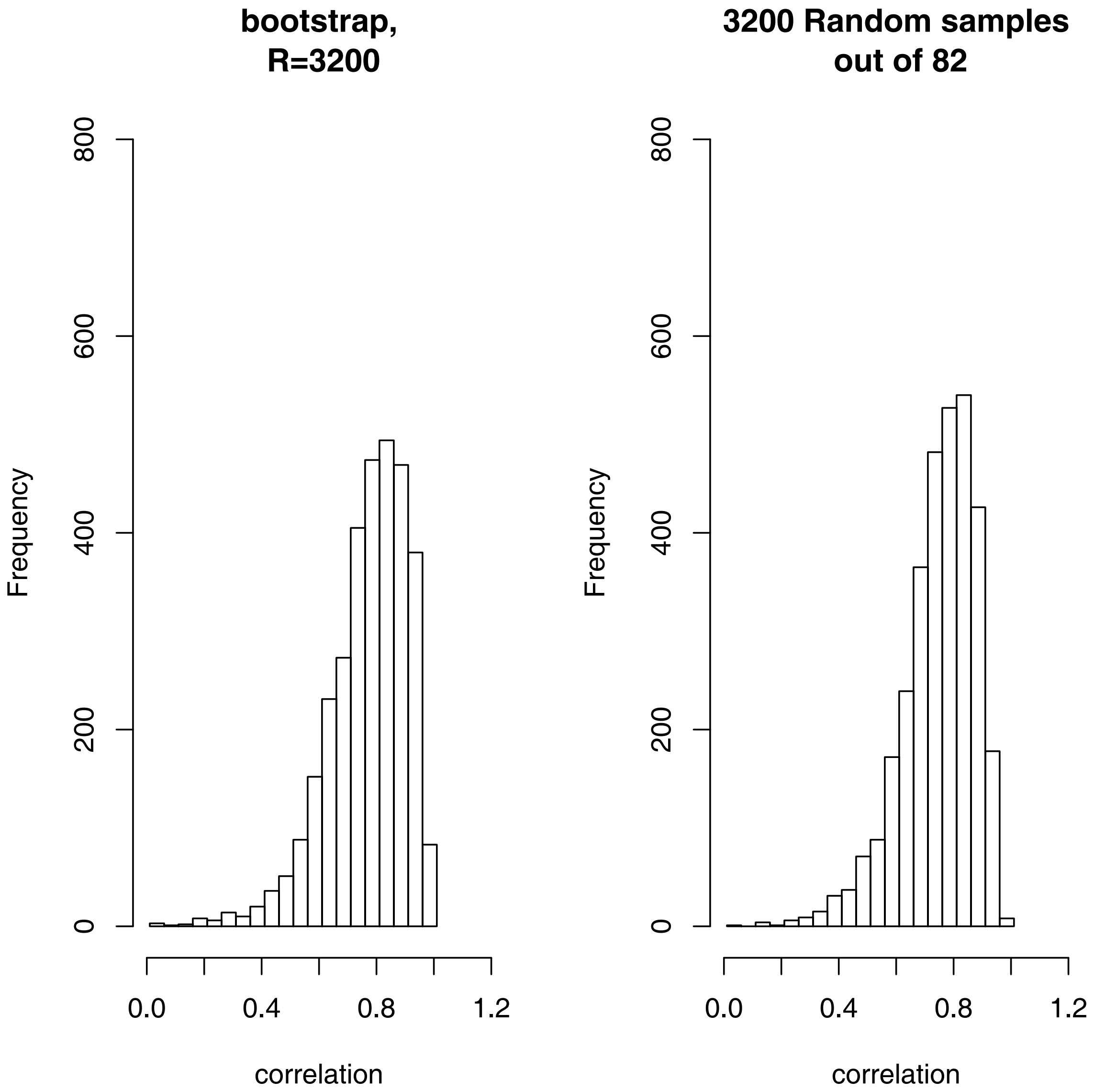
Nonparametric Simulations: Correlation

- The following table gives **bootstrap approximations for the standard error** of this estimate: $\widehat{se}_R = \sqrt{\hat{V}_R}$

R	25	50	100	200	400	800	1600	3200
$\widehat{se}_R$	0.140	0.142	0.151	0.143	0.141	0.137	0.133	0.132

- In the following picture you see two histograms: The first one shows the frequencies of the 3200 replications of the bootstrap versions of the correlation estimate $\hat{\rho}^*(r)$.

  The histogram is non-normal, having a long tail toward the left.


bootstrap,
R=3200
3200 Random samples
out of 82
Frequency
0
200
400
600
800
correlation
0.0
0.4
0.8
1.2

Nonparametric Simulations: Correlation

- In this example we have the entire population of 82 points. The right histogram shows the frequencies of $\hat{\rho}$ for 3200 samples of size $n = 15$ drawn from the entire population.

  The standard deviation of the 3200 $\hat{\rho}$-values was 0.131, so $\widehat{se}_R$ is a good estimate.

- Further, the bootstrap histogram on the left strongly resembles the population histogram on the right.

  Remember, in a real problem we would have the information on the left, from which we would be trying to infer the situation on the right.

The number of bootstrap replications

- Ideal bootstrap estimate B and V corresponds to $R = \infty$, in other words: For fixed $\boldsymbol{x}$

  $\lim_{R\to\infty} \hat{B}_R = B$ and $\lim_{R\to\infty} \hat{V}_R = V$

- computer time, which depends mainly on how long it takes to evaluate the bootstrap replicates, increases with R

  time constraints may dictate a small R

- experience (**rules of thumb**): For the calculation of the bootstrap approximations $\hat{B}_R$ and $\hat{V}_R$ the choice $R = 50$ is often sufficient; seldom more than $R = 200$ are needed

- much bigger values are necessary for the construction of bootstrap confidence intervals

A "bootstrap look" at the distribution

- Very often the unknown distribution of a statistic is approximated by the normal distribution

  Central limit theorem (CLT): For $S_n = \sum_{i=1}^{n} x_i$ we have

  $$\mathsf{P}\left(\frac{S_n - \mathsf{E}S_n}{\sqrt{\mathsf{Var}S_n}} \leq z\right) \to \Phi(z) \qquad \text{shortly} \qquad \frac{S_n - \mathsf{E}S_n}{\sqrt{\mathsf{Var}S_n}} \overset{app}{\sim} N(0,1)$$

- But this approach is not always justified!

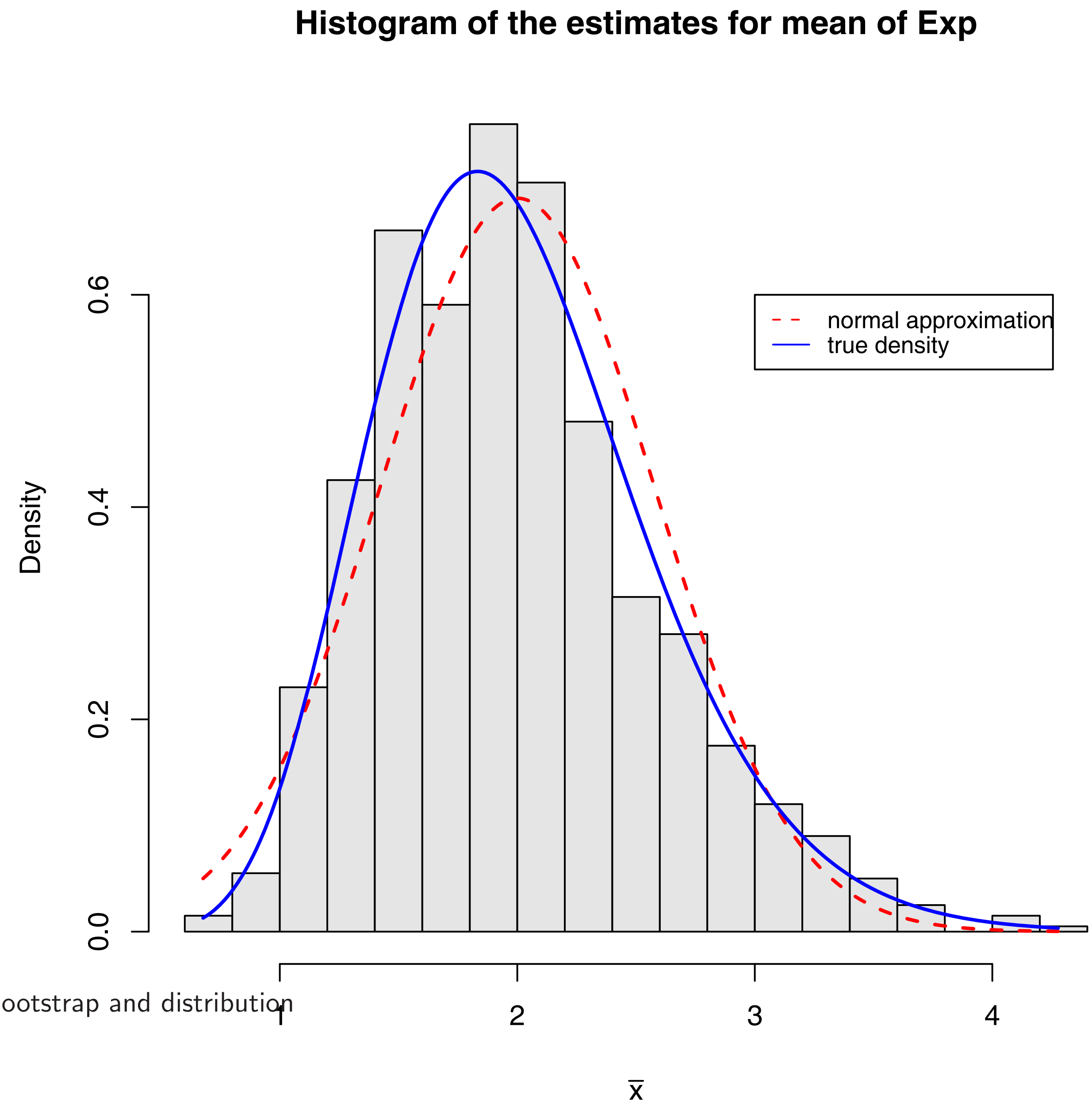
A "bootstrap look" at the distribution

- Example: Data $\boldsymbol{x} = (x_1, \ldots, x_n)$ are drawn from $\mathrm{Exp}(1/\mu)$.
  In our example we choose $\boldsymbol{\mu = 2}$ and $\boldsymbol{n = 12}$. The CLT says:

$$\frac{\sqrt{n}(\overline{x} - \mu)}{\overline{x}} \overset{app}{\sim} N(0,1) \qquad \text{or} \qquad \overline{x} \overset{app}{\sim} N(\mu, \overline{x}^2/n)$$

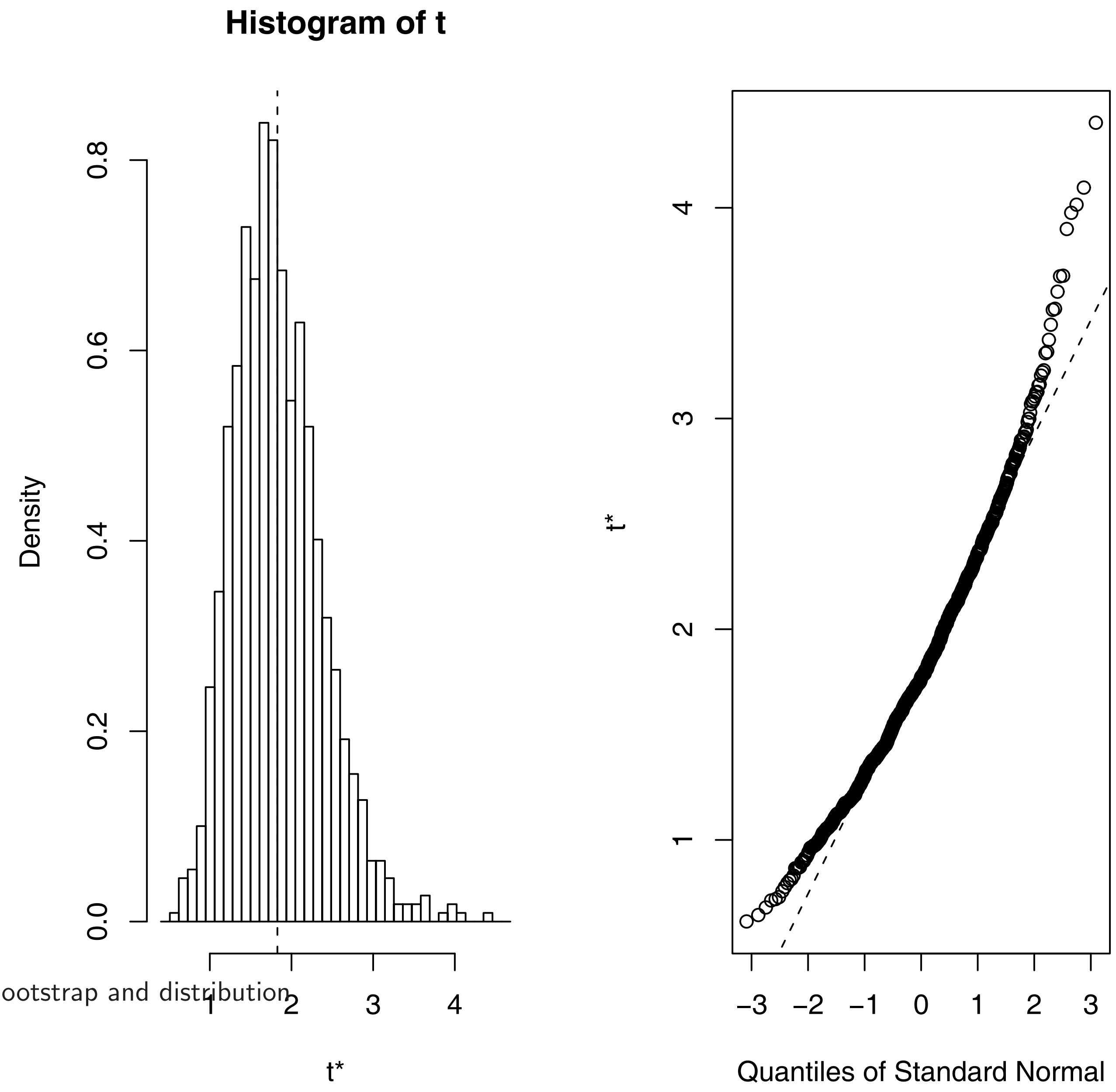
- Look at the following histograms. The first one shows the frequencies of 999 estimates for the mean; they are computed from 999 different samples from the exponentially distributed population.

- In this case we know the distribution of $\overline{x}$. The sum of exponential distributed r.v.'s is Gamma-distributed. The true density of $\overline{x}$ for our example is the (blue) solid line.

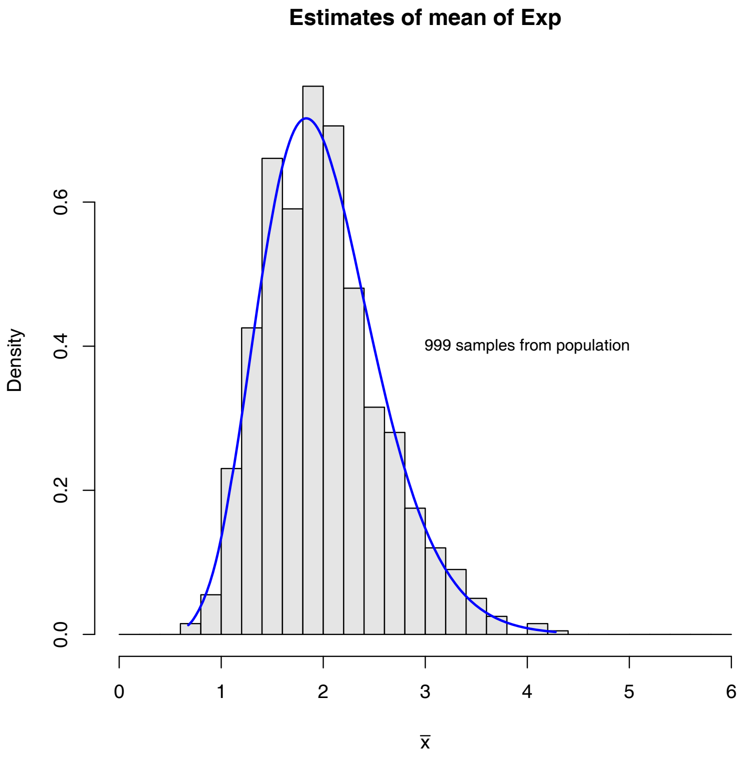

Histogram of the estimates for mean of Exp
Density
0.0
0.2
0.4
0.6
1
2
3
4
$\bar{x}$
normal approximation
true density

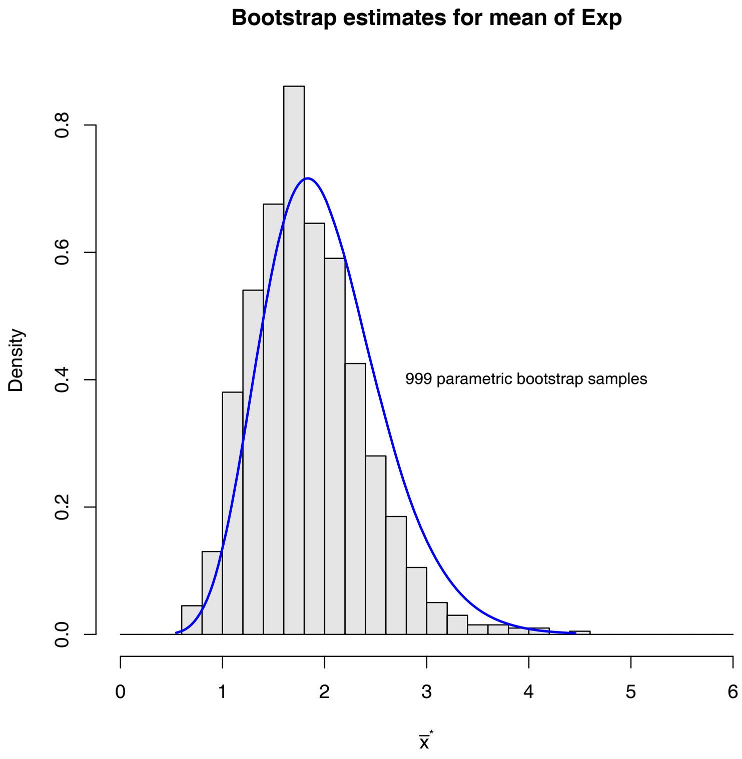
A ”bootstrap look” at the distribution

- The next figure is the **bootstrap output** from the R procedure. We see the histogram of $R = 999$ parametric bootstrap estimators for the mean and a **normal QQ-plot**. Both pictures show that the estimates are not normally distributed.


Histogram of t
Density
0.0
0.2
0.4
0.6
0.8
1
2
3
4
t*
t*
1
2
3
4
−3
−2
−1
0
1
2
3
Quantiles of Standard Normal

A ”bootstrap look” at the distribution





Here we see, that the bootstrap histogram resembles the histogram of the samples.

A "bootstrap look" at the distribution

- For more complicated statistics we use the delta method: Let $\hat{\theta} = s(\boldsymbol{x})$[1] be a function of k sums of i.i.d. r.v.'s, say $\hat{\theta} = h(S_{n1}, \ldots, S_{nk})$

  Suppose a multivariate CLT holds, i.e.

  $$\sqrt{n}(\boldsymbol{S}_n - \mathsf{E}(\boldsymbol{S}_n|F)) \overset{app}{\sim} N_k(\mathbf{0}, \boldsymbol{C}),$$

  where $\boldsymbol{S}_n$ is the vector of the sums S_{nj} and $\boldsymbol{C}$ is a non-singular $k \times k$-matrix.

  Let ∇h be the vector of partial derivatives of h, then we have

  $$\sqrt{n}(\hat{\theta} - h(\mathsf{E}(\boldsymbol{S}_n|F)) \overset{app}{\sim} N_1(0, \tau^2)$$

  with $\tau^2 = \nabla h^t C \nabla h$ at the point $\mathsf{E}(\boldsymbol{S}_n|F)$.

[1] For simplicity we consider one dimensional parameters θ

Bootstrap and distribution : City - Example

- Let us apply this to our city-data example: We have

$$\theta = \frac{\mathsf{E}v_1}{\mathsf{E}u_1} \qquad \hat{\theta} = \frac{\sum_i v_i}{\sum_i u_i} = \frac{S_{n1}}{S_{n2}} = h(S_{n1}, S_{n2})$$

- Further $\quad h(v,u) = \frac{v}{u} \qquad \nabla h(v,u) = (u^{-1}, -vu^{-2})^t$ and

$$\boldsymbol{C} = \begin{pmatrix} \sigma_v^2 & \sigma_{vu} \\ \sigma_{vu} & \sigma_u^2 \end{pmatrix}$$

where σ_v^2, σ_u^2 and σ_{vu} are the variances and covariance, respectively.

Bootstrap and distribution : City - Example

With $\nabla h^t \boldsymbol{C} \nabla h = u^{-4}\left(u\sigma_v^2 - 2vu\sigma_{uv} + v\sigma_u^2\right)$ we have

$$\sqrt{n}(\hat{\theta} - \theta) \overset{app}{\sim} N(0, \tau^2) \quad \text{with}$$

$$\tau^2 = (\mathsf{E}u_i)^{-4}\left(\mathsf{E}u_i\,\sigma_v^2 - 2\mathsf{E}v_i\,\mathsf{E}u_i\,\sigma_{uv} + \mathsf{E}v_i\,\sigma_u^2\right)$$

If the normal approximation can be justified, then the (ideal for the normal model) bootstrap estimate for the variance of the estimator is

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{(\overline{v}u_i - \overline{u}v_i)^2}{\overline{u}^4}\,(= V_{normal})$$

Bootstrap and distribution : City - Example

- in our numerical example we get the following values

  $n = 10$ nonparametric bootstrap approximation (R=999)

  $\hat{B}_R = 0.026 \qquad \hat{V}_R = 0.0467$

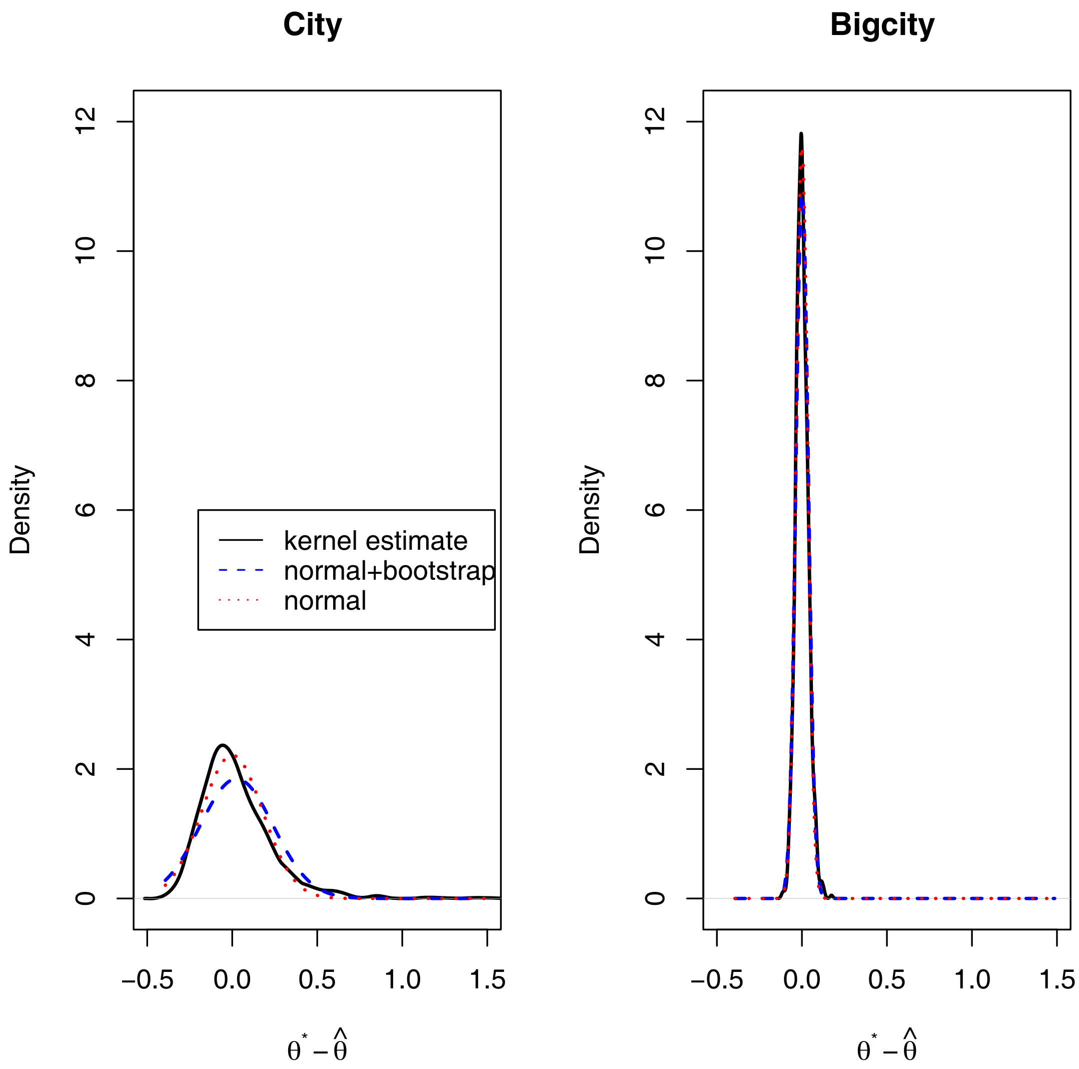
  normal approximation: $B_{normal} = 0 \qquad V_{normal} = 0.0325$

- $n = 49 \qquad \hat{B}_R = 0.0014 \qquad \hat{V}_R = 0.0013$

  normal approximation: $B_{normal} = 0 \qquad V_{normal} = 0.00116$

- If the sample size n is large, the normal approximation is o.k., but for small sample sizes bootstrap can be a helpful tool.

- See also the following figures:


City
Bigcity
Density
Density
0
2
4
6
8
10
12
−0.5
0.0
0.5
1.0
1.5
$\theta^* - \hat{\theta}$
kernel estimate
normal+bootstrap
normal

Bootstrap - Distribution - Quantiles

- For statistical inference about θ it is necessary to know the distribution

  $G(z) \;=\; \mathsf{P}(\hat\theta - \theta \le z)$

- Let $\hat\theta^*$ be a bootstrap version of $\hat\theta$, i.e. $\hat\theta = s(\boldsymbol{x})$ and $\hat\theta^* = s(\boldsymbol{x}^*)$

  The distribution of $\hat\theta^* - \hat\theta$ is denoted by G^*

  $G^*(z) \;=\; \mathsf{P}(\hat\theta^* - \theta \le z|\tilde F)$

- G^* is approximated by bootstrap simulations:

  from repeated bootstrap samples we get $\hat\theta^*(1), \ldots, \hat\theta^*(R)$, and set

$$\widehat{G}^*_R(z) = \frac{\#\{\hat\theta^*(r) - \hat\theta \le z\}}{R}$$

Bootstrap - Distribution - Quantiles

- Note, here we have two approximation errors:

  the bootstrap df G^* estimates G, the error $G - G^*$ depends on the underlying problem;

  the empirical df $\hat{G}^*_R$ approximates G^*, the error $G^* - \hat{G}^*_R$ depends on the simulation size R (can be made small)

- From $\hat{G}^*_R$ we calculate **approximations for the quantiles** of G.

  The p-quantile $G^{-1}(p)$ is estimated by $\hat{\theta}^*_{[(R+1)p]} - \hat{\theta}$, [2]

  where $\hat{\theta}^*_{[1]} \leq \hat{\theta}^*_{[2]} \leq \cdots \leq \hat{\theta}^*_{[R]}$ are the ordered bootstrap estimates

  Example: $R = 99$, $p = 0.95$ $G^{-1}(0.95)$ is estimated by $\hat{\theta}^*_{[95]} - \hat{\theta}$.

[2] We assume that R is chosen so that $(R+1)p$ is an integer.

Confidence intervals based on bootstrap

- Exact case:
  Suppose that the data $\boldsymbol{x} = (x_1, \ldots, x_n)$ are drawn from a $N(\mu, \sigma^2)$-distributed population. Then an exact confidence interval with coverage probability $1 - 2\alpha$ for the mean μ is given by

  $$[\overline{x} - t_{n-1;1-\alpha}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1;1-\alpha}\frac{s}{\sqrt{n}}].$$

  Here $t_{n-1;1-\alpha}$ is the $(1 - \alpha)$-quantile of the t- distribution (Student-distribution) with $n - 1$ degrees of freedom.

- In the following we consider three methods for the construction of approximative confidence intervals for the case that normality of the data is not assumed.

Bootstrap confidence intervals: Normal approximation

- Approximative interval based on normal approximation:
  Suppose that a limit statement of the form

$$\frac{\hat{\theta} - \mathsf{E}(\hat{\theta}|F)}{\sqrt{\mathsf{Var}(\hat{\theta}|F)}} \overset{app}{\sim} N(0,1) \quad \text{i.e.} \quad \frac{\hat{\theta} - \theta - \beta}{\sqrt{\mathsf{Var}(\hat{\theta}|F)}} \overset{app}{\sim} N(0,1),$$

  holds, where $\beta = \mathsf{E}(\hat{\theta}|F) - \theta$ is the bias.

- Replacing now the unknown quantities β and $\nu = \mathsf{Var}(\hat{\theta}|F)$ by the bootstrap approximations $\hat{B}_R$ and $\hat{V}_R$ we obtain as approximative confidence interval

$$\left[\hat{\theta} - \hat{B}_R - u_{1-\alpha}\sqrt{\hat{V}_R}, \quad \hat{\theta} - \hat{B}_R + u_{1-\alpha}\sqrt{\hat{V}_R}\right].$$

Bootstrap confidence intervals: Normal approximation

- Note, that here the t-quantile is replaced by the quantile $u_{1-\alpha}$ of the standard normal distribution. This is due to the fact that for large n the t- and the normal distribution are very similar.

- In the case that θ is the mean this approach leads to the usual one; here it is not necessary to construct R bootstrap samples because $B = 0$ and $V = \mathsf{Var}(\hat{\theta}^* | \tilde{F})$ is known.

Bootstrap confidence intervals: Basic bootstrap

- Starting point of a construction of a confidence interval $[l,\ u]$ for θ is:

$$\begin{aligned} 1-2\alpha &= \mathsf{P}(l \le \hat{\theta} - \theta \le u) = G(u) - G(l) \\ &= \mathsf{P}(\hat{\theta} - u \le \theta \le \hat{\theta} - l) \end{aligned}$$

  where G is the distribution of $\hat{\theta} - \theta$. The lower and the upper bound l and u, are given by the (unknown) quantiles

  $\boldsymbol{u = G^{-1}(1-\alpha)}$ and $\boldsymbol{l = G^{-1}(\alpha)}$

- they are approximated by the bootstrap quantiles $\Rightarrow$ Basic bootstrap intervals

$$\hat{\theta}^*_{[(R+1)(1-\alpha)]} - \hat{\theta} \qquad \text{and} \qquad \hat{\theta}^*_{[(R+1)\alpha]} - \hat{\theta}$$

Bootstrap confidence intervals: Basic bootstrap

- Thus the confidence interval is given by

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

- Example $R = 999$, $\alpha = 0.025$, coverage probability 0.95

$$\left[2\hat{\theta} - \hat{\theta}^*_{[975]}, \quad 2\hat{\theta} - \hat{\theta}^*_{[25]}\right]$$

Bootstrap confidence intervals: Studentized intervals

- The studentized deviation of the estimate from the parameter is

$$z = \frac{\hat{\theta} - \theta}{\widehat{se}},$$

where $\widehat{se}$ is a suitable estimate of the standard error of $\hat{\theta}$ based on the original data $\boldsymbol{x}$ (e.g. V or an estimate derived by the delta method)

- bootstrap versions of z are generated by R simulations:

$$z^*(r) = \frac{\hat{\theta}^*(r) - \hat{\theta}}{\widehat{se^*(r)}},$$

Here $\widehat{se^*(r)}$ is the estimator for the standard error constructed with the bootstrap sample $\boldsymbol{x}^{*r}$.

Bootstrap confidence intervals: Studentized intervals

- From the empirical distribution of the $z^*(1), z^*(2), \ldots, z^*(R)$ we calculate quantiles:

$$z^*_{[(R+1)(1-\alpha)]} \qquad \text{and} \qquad z^*_{[(R+1)\alpha]}$$

- the confidence interval is given by

$$\left[\hat{\theta} - \widehat{se} \cdot z^*_{[(R+1)(1-\alpha)]}, \quad \hat{\theta} - \widehat{se} \cdot z^*_{[(R+1)\alpha]}\right]$$

Bootstrap confidence intervals: Air-condition example

- Under the assumption that the data are from a exponentially distributed population one can compute an exact confidence interval (using the Gamma distribution);

  $\Rightarrow \qquad [65.9,\ 209.2]$ (coverage probability 0.95)

- the normal approximation $\quad \overline{x} - \hat{B}_R \pm u_{1-\alpha}\sqrt{\hat{V}_R}$

  $\overline{x} = \quad B = 0 \quad V = \frac{\overline{x}^2}{n} = 31.2^2$

  $\Rightarrow \qquad [46.93,\ 169.24]$

- basic bootstrap $\quad R = 999 \qquad \left(2\overline{x} - \overline{x}^*_{[975]}, \quad 2\overline{x} - \overline{x}^*_{[25]}\right)$

  in simulation $\overline{x}^*_{[975]} = 171.3$ and $\overline{x}^*_{[25]} = 55.7$

  $\Rightarrow \qquad [44.86,\ 160.46]$

Bootstrap confidence intervals: Air-condition example

- studentized bootstrap $\left(\overline{x} - \widehat{se} \cdot z^*_{((R+1)(1-\alpha))}, \overline{x} - \widehat{se} \cdot z^*_{((R+1)(\alpha))}\right)$

$$z = \sqrt{n}\,\frac{\overline{x} - \mu}{\overline{x}} = \sqrt{n}\,\left(1 - \frac{\mu}{\overline{x}}\right) \qquad z^* = \sqrt{n}\,\left(1 - \frac{\overline{x}}{\overline{x}^*}\right)$$

$z^*(1), \ldots, z^*(R) \qquad R = 999 \qquad \widehat{se} = \overline{x}/\sqrt{n} = 31.2$

in simulation $\overline{z}^*_{[975]} = 1.28$ and $\overline{z}^*_{[25]} = -3.26$

$\Rightarrow \qquad [68.19,\ 209.71]$

Bootstrap confidence intervals: City-Data Example

- the normal approximation $\hat{\theta} - \hat{B}_R \pm u_{1-\alpha}\sqrt{\hat{V}_R}$

  $\hat{\theta} = 1.52 \quad B_R = 0.048 \quad V_R = 0.235^2$

  $\Rightarrow \qquad [1.01,\ 1.93]$

- basic bootstrap $R = 999$ $\left(2\hat{\theta} - \hat{\theta}^*_{[975]}, \quad 2\hat{\theta} - \hat{\theta}^*_{[25]}\right)$

  in simulation $\hat{\theta}^*_{[975]} = 2.176$ and $\hat{\theta}^*_{[25]} = 1.251$

  $\Rightarrow \qquad [0.87,\ 1.79]$

Bootstrap confidence intervals: City-Data Example

- studentized bootstrap $\left(\hat{\theta} - \widehat{se_L} \cdot z^*_{((R+1)(1-\alpha))}, \hat{\theta} - \widehat{se_L} \cdot z^*_{((R+1)(\alpha))}\right)$

  in simulation $R = 999$ $\widehat{se_L} = 0.0325$

  $\overline{z}^*_{[975]} = 1.49$ and $\overline{z}^*_{[25]} = -3.64$

  $\Rightarrow$ $[1.25,\ 2.18]$

Bootstrap - Simple linear regression

- Observations (x_i, y_i) $i = 1, \ldots, n$ satisfy

$$\boldsymbol{y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\mu_i} + \varepsilon_i} \qquad (1) \qquad\qquad \mathsf{E}\boldsymbol{\varepsilon_i = 0} \quad \mathsf{Var}\boldsymbol{\varepsilon_i = \sigma^2} \qquad (2)$$

l.s.e for parameters β_0 und β_1

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} \qquad \widehat{\beta}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$$

Residuals: $e_i = y_i - \widehat{\mu}_i$ $\quad \widehat{\mu}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$ $\quad \widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$

optimal procedure Verfahren if (1) and (2) true

Bootstrap - Simple linear regression

- statistical inference (confidence intervals, tests) is based on

$$\boldsymbol{\varepsilon_i \sim \mathsf{N}(0, \sigma^2)} \qquad (3)$$

- important: **residuals** $e_i = y_i - \widehat{\mu}_i$

  since under (1) and (2) $\mathsf{E}\widehat{\beta}_0 = \beta_0$ und $\mathsf{E}\widehat{\beta}_1 = \beta_1$

  $\Rightarrow \quad \mathsf{E}_x e_i = 0 \qquad$ and $\qquad \mathsf{Var}_x e_i = \sigma^2(1 - h_i)$

  $h_i = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_j (x_j - \overline{x})^2}$ so-called leverages

- **modified residuals** $r_i = \frac{\mathrm{e}_i}{\sqrt{1 - h_i}} \qquad \mathsf{E}_x r_i = 0 \qquad \mathsf{Var}_x r_i = \sigma^2$

Simple linear regression **Model-based resampling**

- **Fixed design model**, x_i fixed quantities

  y_i are realizations of r.v.'s Y_i with distribution functions F_i (not identically distributed)

  $\mathsf{P}(Y_i \le t) = F_i(t) = F_{x_i}(t)$

  distribution of the error ε_i : $\boldsymbol{G(t) = \mathsf{P}(\varepsilon_i \le t)}$

$$\Rightarrow \qquad F_i(t) = \mathsf{P}(\varepsilon_i \le t - \mu_i) = G(t - \mu_i)$$

(i) from "parametric fit" approach: $G(t - \widehat{\mu}_i)$

(ii) no parametric assumption on G $\Rightarrow$ nonparametric estimation of G by empirical df??

Simple linear regression **Model-based resampling**

nonparametric estimation of G by empirical df??

but the ε_i's are not observable $\Rightarrow$ replace ε_i by modified residuals r_i or centered residuals $r_i - \overline{r}$

(not only $\mathrm{E}(r_i - \overline{r}) = 0$ but also $\sum(r_i - \overline{r}) = 0$)

the df G is estimated by the empirical distribution of $r_i - \overline{r}$

(iii) the same design as in the original data model, $x_i^* = x_i$

$\Rightarrow$ Bootstrap version of $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$\begin{aligned} \boldsymbol{y_i^*} &= \boldsymbol{\widehat{\beta}_0 + \widehat{\beta}_1 x_i^* + \varepsilon_i^*} \qquad \boldsymbol{\varepsilon_i^*} \text{ randomly sampled from } \widehat{G} \\ \boldsymbol{\widehat{G}(u)} &= \boldsymbol{\frac{\#\{r_i - \overline{r} \le u\}}{n}} \end{aligned}$$

Simple linear regression Model-based resampling

- **Algorithm:** (**Model-based resampling in linear regression**)
  For $r = 1, \ldots, R$

```
1. For $i = 1, \ldots, n$
    (a) set $x_i^* = x_i$
    (b) randomly sample $\varepsilon_i^*$ from $r_1 - \overline{r}, \ldots, r_i - \overline{r}$
    (c) set $y_i^* = \widehat{\beta}_0 + \widehat{\beta}_1 x_i^* + \varepsilon_i^*$
2. Fit least squares regression to $(x_1^*, y_1^*), \ldots, (x_n^*, y_n^*)$
    ⇒ $\boldsymbol{\widehat{\beta}^*_{0,r}, \widehat{\beta}^*_{1r}, \sigma_r^{*2}}$
```

Model-based resampling

- **Properties**: Because of

$$
\begin{aligned}
\widehat{\beta}^*_{1,r} &= \frac{\sum(x^*_{ir}-\overline{x}^*_r)y^*_{ir}}{\sum(x^*_{jr}-\overline{x}_r)^2} = \frac{\sum(x_i-\overline{x})y^*_{ir}}{\sum(x_j-\overline{x})^2} \\
& \frac{\sum(x_i-\overline{x})(\widehat{\beta}_0+\widehat{\beta}_1 x_i+\varepsilon^*_{ir})}{\sum(x_i-\overline{x})^2} = \frac{\sum(x_i-\overline{x})(\overline{y}-\widehat{\beta}_1\overline{x}+\widehat{\beta}_1 x_i+\varepsilon^*_{ir})}{\sum(x_i-\overline{x})^2} \\
&= \widehat{\beta}_1 + \frac{\sum(x_i-\overline{x})\varepsilon^*_{ir}}{\sum(x_i-\overline{x})^2}
\end{aligned}
$$

and $\mathsf{E}^*(\varepsilon^*_{ir}) = \frac{1}{n}\sum(r_j-\overline{r}) = 0$ we have

$$\mathsf{E}^*\widehat{\boldsymbol{\beta}}^*_{\boldsymbol{1,r}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}$$

Model-based resampling

Further

$$\begin{aligned}
\text{Var}^*(\widehat{\boldsymbol{\beta}}^*_{\boldsymbol{1,r}}) &= \frac{\sum(x_i - \overline{x})^2 \ \text{Var}^*(\varepsilon^*_{ir})}{(\sum(x_i - \overline{x})^2)^2} \\
&= \frac{\frac{1}{n}\sum(r_j - \overline{r})^2}{\sum(x_i - \overline{x})^2} \\
&\sim \frac{1}{n-2}\sum e_j^2 \\
&= \widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}}
\end{aligned}$$

- Aim

  Check of normality assumption

  Computing quantiles

Bootstrap - Simple linear regression **Example**

- **Mammals:** Average body weight (kg) and brain weight (g) for 62 species of mammals

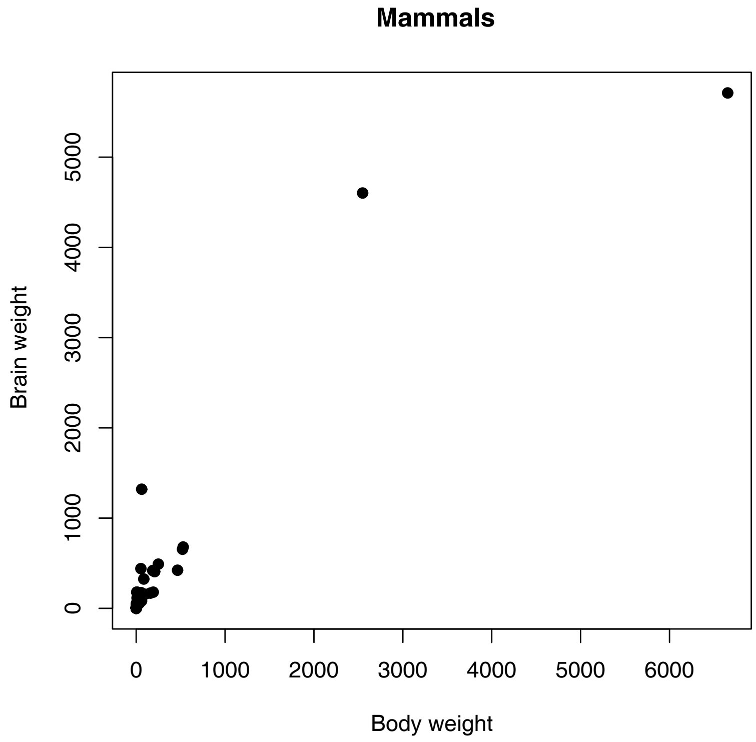
  $x_i = \log(\text{body weight}) \qquad y_i = \log(\text{brain weight})$

$$y_i \; = \; \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, \ldots, 62$$

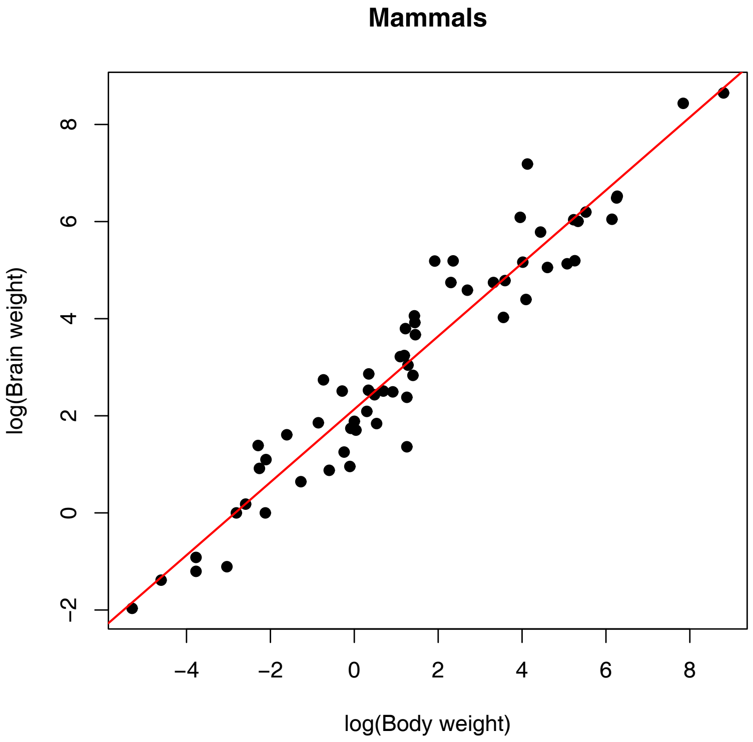
- *mam.lm← glm(log(brain) log(body),data=mammals)*

	Estimate	Std. Error	t value
$\hat{\beta}_0$	2.135	0.0960	22.23
$\hat{\beta}_1$	0.752	0.0285	26.41

Bootstrap - Simple linear regression **Example**



Mammals



Mammals

Bootstrap - Simple linear regression **Example**

- $R = 499$ bootstrap samples

ORDINARY NONPARAMETRIC BOOTSTRAP

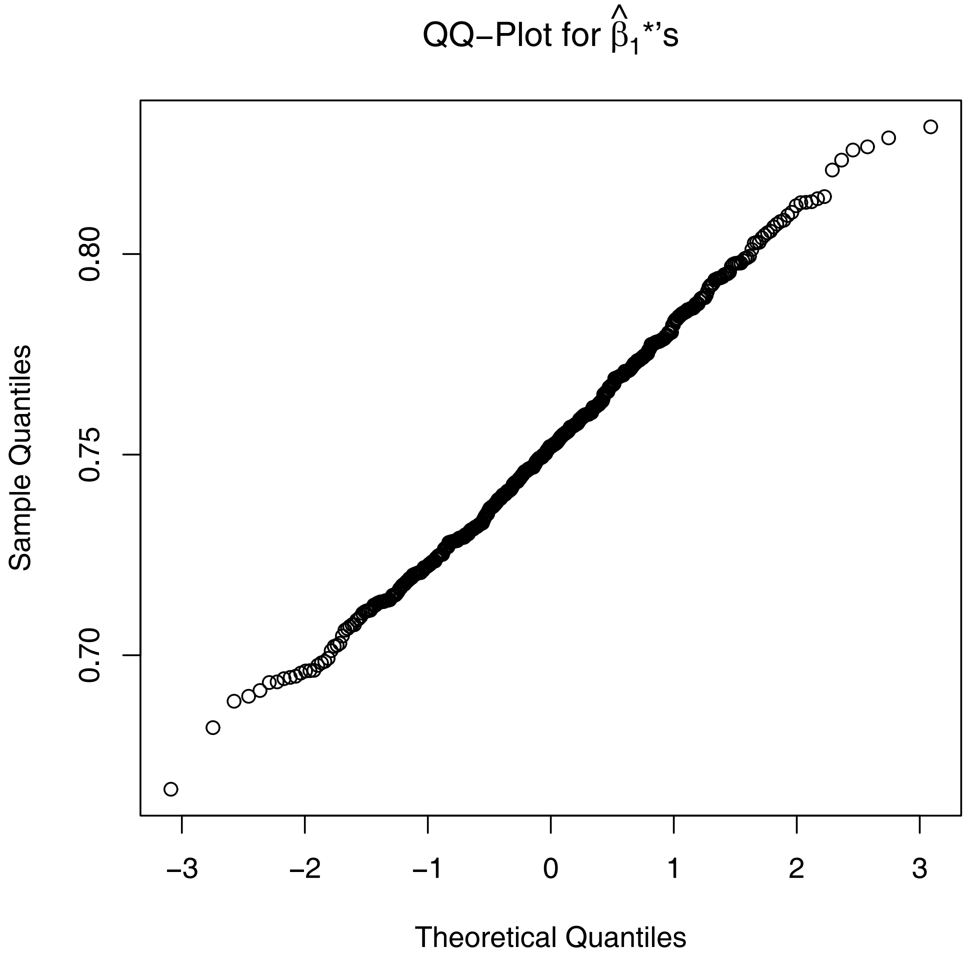
Call: *boot(data = mam, statistic = mam.fun, R = 499)*

Bootstrap Statistics :

	original	bias	std. error
t1*	2.135	0.0005	0.0940
t2*	0.752	0.0006	0.0291

compared to the ”theoretical values” 0.096 and 0.0285

Bootstrap - Simple linear regression **Example**



Bootstrap - Simple linear regression **Example**

- for normally distributed errors the **confidence interval** for β_1 is given by

$$\frac{\widehat{\beta}_1 - \beta_1}{\hat{se}(\widehat{\beta}_1)} \sim t_{n-1} \qquad \widehat{\beta}_1 \pm t_{n-1;1-\frac{\alpha}{2}} \cdot \hat{se}(\widehat{\beta}_1)$$

for $\alpha = 0.1$: $t_{61;0.05} = -1.67,\ t_{61;0.95} = 1.67$ yields [0.7041, 0.7991]

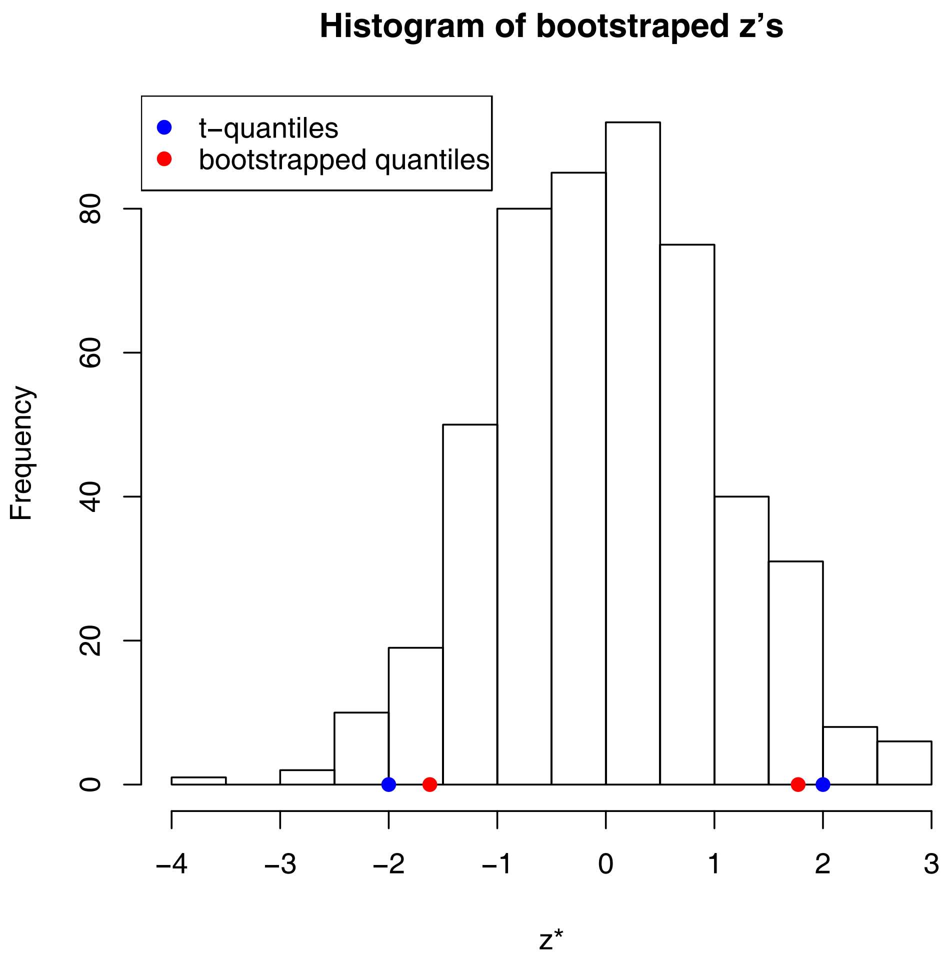
- bootstrap quantiles $r = 1, \ldots, R = 499$

$$\boldsymbol{z_r^* = \frac{\widehat{\beta}_{1,r}^* - \widehat{\beta}_1}{\hat{se}^*(\widehat{\beta}_{1r}^*)}} \qquad \text{ordered} \quad \boldsymbol{z_{(1)}^* \leq \cdots \leq z_{(499)}^*}$$

Quantiles: $z_{(25)}^* = -1.62 \qquad z_{(475)}^* = 1.77$

interval $[\widehat{\beta}_1 - 1.62 \cdot \hat{se}(\widehat{\beta}_1)\,,\ \widehat{\beta}_1 + 1.77 \cdot \hat{se}(\widehat{\beta}_1)] \ = \ [0.7055,\ 0.8020]$

Bootstrap - Simple linear regression Example


Histogram of bootstraped z's
● t–quantiles
● bootstrapped quantiles
Frequency
z*