Estimation, Accuracy, and the Bootstrap

Bradley Efron

brad@stat.stanford.edu

Department of Statistics
Stanford University

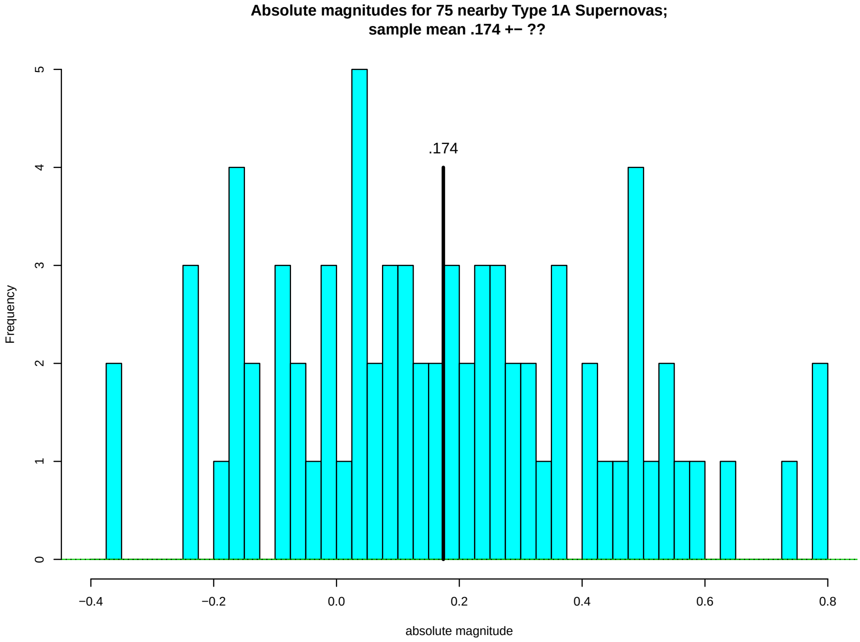
Estimates and Accuracy

- Unknown quantity of interest θ
- Data $\boldsymbol{x}$ relating to θ
- Estimation algorithm f
- Estimate: $\hat{\theta} = f(\boldsymbol{x})$

Question: How accurate is $\hat{\theta}$?

Amazing fact:

The same data that gives $\hat{\theta}$ can also give an estimate of its accuracy!


Absolute magnitudes for 75 nearby Type 1A Supernovas;
sample mean .174 +− ??
.174
Frequency
0
1
2
3
4
5
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
absolute magnitude

The Sample Mean

- *Observe* $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \qquad (n = 75)$
- $\theta =$ true mean
- $\hat{\theta} = \bar{x} = \sum x_i/n \quad (= 0.174)$
- **Famous formula:** The estimated "standard error" of $\hat{\theta}$ (its root mean square variability) is

$$\hat{\sigma} = \left\{ \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ \left[n(n-1) \right] \right\}^{1/2} \qquad (= 0.031)$$

- To a first order of approximation,

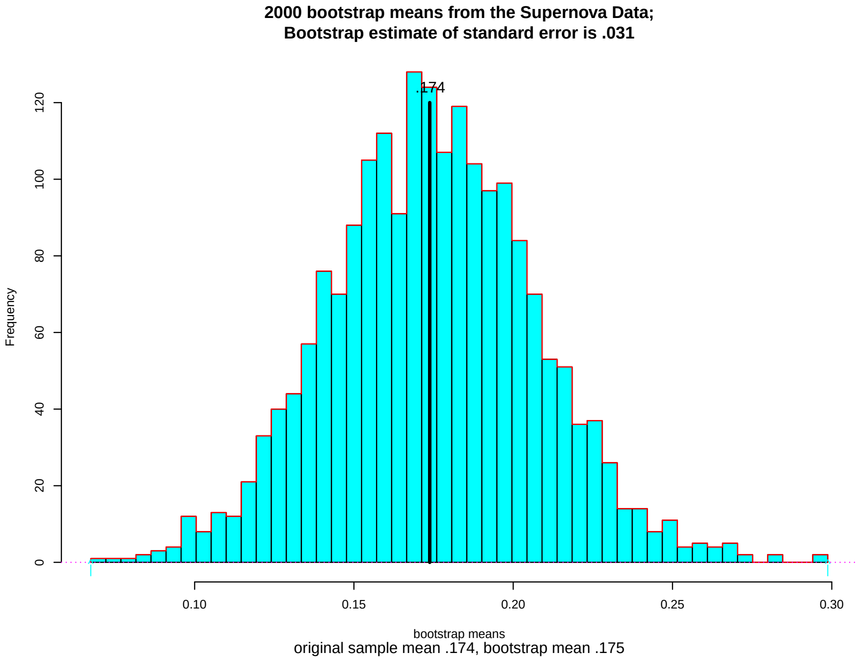
$$\Pr\left\{ \theta \in \hat{\theta} \pm 1.64\hat{\sigma} \right\} = 0.90$$

Nonparametric Bootstrap SE

R package `bootstrap`

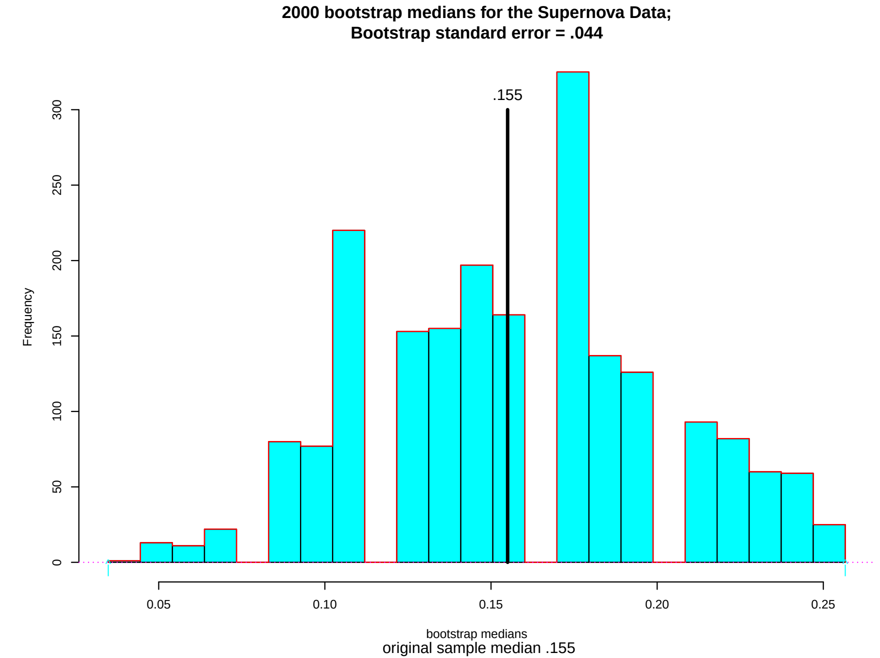
- *Bootstrap sample* n draws with replacement from $\{x_1, x_2, \ldots, x_n\} \Longrightarrow \boldsymbol{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$
- Bootstrap replication of $\hat{\theta} = f(\boldsymbol{x})$: $\hat{\theta}^* = f(\boldsymbol{x}^*)$
- Bootstrap standard error:
  - B replications $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$
  - $\widehat{\text{se}} = \left[\sum_{b=1}^{B} \left(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\right)^2 \Big/ (B-1)\right]^{1/2}$

    where $\hat{\theta}^*(\cdot) = \sum \hat{\theta}^*(b)/B$


2000 bootstrap means from the Supernova Data;
Bootstrap estimate of standard error is .031
.174
Frequency
0
20
40
60
80
100
120
0.10
0.15
0.20
0.25
0.30
bootstrap means
original sample mean .174, bootstrap mean .175

More Complicated Statistics

- Estimates $\hat{\theta} = f(\boldsymbol{x})$ can be much more complicated than the sample mean
- Individual observations x_i can be vectors, time series, images, ...
- No "magic formula"
- *Supernova example* $\hat{\theta} = f(\boldsymbol{x})$ sample median $= 0.155 \pm ??$


2000 bootstrap medians for the Supernova Data;
Bootstrap standard error = .044
.155
Frequency
0
50
100
150
200
250
300
0.05
0.10
0.15
0.20
0.25
bootstrap medians
original sample median .155

Why Does the Bootstrap Work?

- What happened in the real world:

$$\begin{array}{ccccc} F & \xrightarrow{\text{iid sampling}} & x_1, x_2, \ldots, x_n & \xrightarrow{f(\cdot)} & \hat{\theta} \\ \uparrow & & \uparrow & & \uparrow \\ \text{true distribution} & & \text{observed sample} & & \text{estimate} \end{array}$$

- Just one sample and one estimate

Why Does the Bootstrap Work?

- What happens in the bootstrap world:

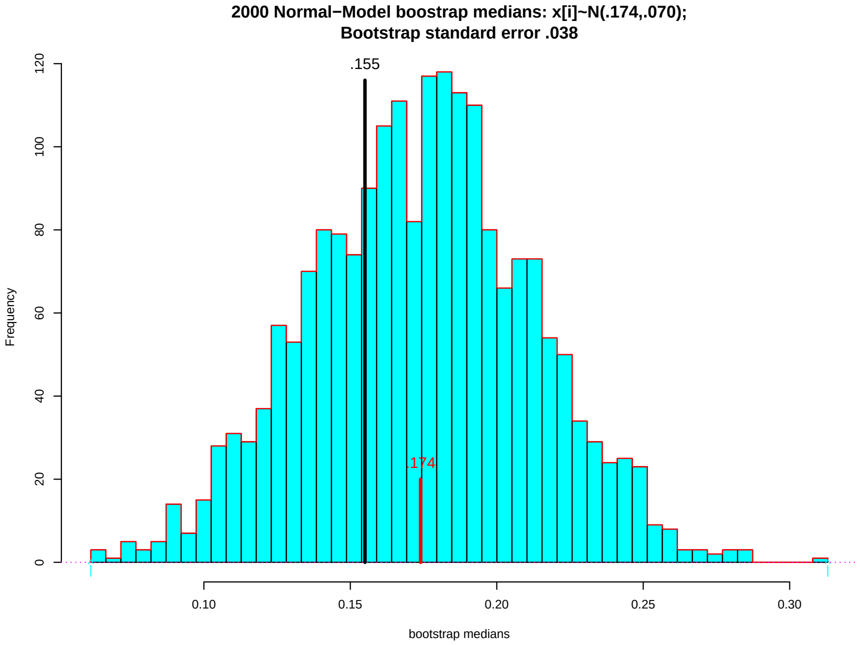
$$
\widehat{F} \quad \xrightarrow{\text{iid}} \left\{ \begin{array}{ccc}
x_1^*(1), x_2^*(1), \ldots, x_n^*(1) & \xrightarrow{f(\cdot)} & \hat{\theta}^*(1) \\
x_1^*(2), x_2^*(2), \ldots, x_n^*(2) & \xrightarrow{f(\cdot)} & \hat{\theta}^*(2) \\
\vdots \quad \vdots \quad \vdots \quad \vdots & \vdots & \vdots \\
x_1^*(B), x_2^*(B), \ldots, x_n^*(B) & \xrightarrow{f(\cdot)} & \hat{\theta}^*(B)
\end{array} \right.
$$

$$
\begin{array}{ccc}
\uparrow & \uparrow & \uparrow \\
\text{estimated distribution} & \text{bootstrap samples} & \text{bootstrap estimates}
\end{array}
$$

- As many samples and estimates as we want

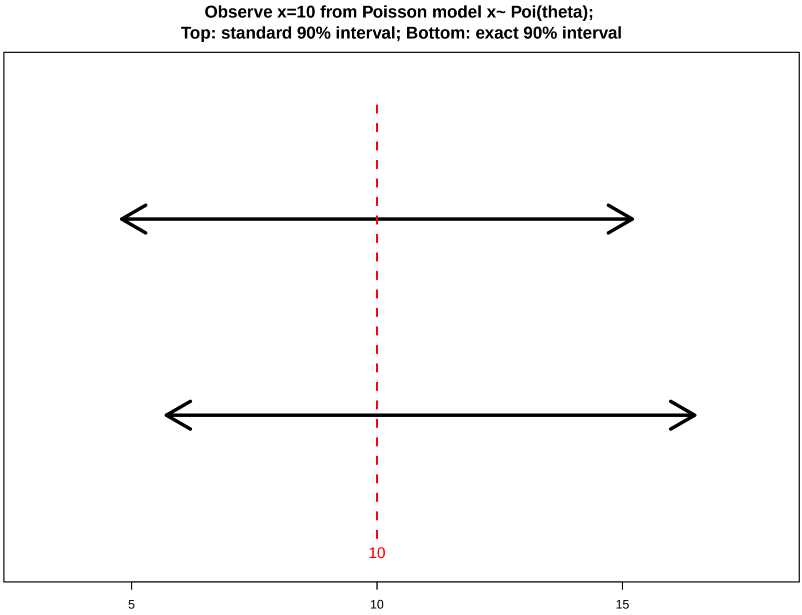
Estimating the Distribution F

- Nonparametric bootstrap:
  - $\widehat{F} =$ empirical distribution
  - probability $1/n$ on each x_i
- Parametric bootstrap:
  - $\widehat{F}$ maximum likelihood estimate from some parametric model
  - e.g., assume $x_i \sim \mathcal{N}(\mu, \nu)$, $\widehat{F} \sim \mathcal{N}(\hat{\mu}, \hat{\nu})$


2000 Normal−Model boostrap medians: x[i]~N(.174,.070);
Bootstrap standard error .038
.155
.174
Frequency
0
20
40
60
80
100
120
0.10
0.15
0.20
0.25
0.30
bootstrap medians

Bootstrap Confidence Intervals

- **Standard intervals** $C_\alpha(\text{stan}) = \hat{\theta} \pm c_\alpha \hat{\sigma}$
  - $c_\alpha = 1.645$ for $\alpha = 0.90$; $c_\alpha = 1.96$ for $\alpha = 0.95$
  - Accuracy $\Pr\left\{\theta \in C_\alpha(\text{stan})\right\} = \alpha + O\left(1/\sqrt{n}\right)$
- **Bootstrap intervals** R package `bcaboot`
  - Second-order accuracy $\Pr\left\{\theta \in C_\alpha(\text{bca})\right\} = \alpha + O(1/n)$
  - Typically requires $B \approx 2000$


Observe x=10 from Poisson model x~ Poi(theta);
Top: standard 90% interval; Bottom: exact 90% interval
10
5
10
15

Bca Intervals

"bias corrected and accelerated"

- **Standard intervals** assume

$$\hat{\theta} \sim \mathcal{N}(\theta, \mathrm{se}^2)$$

normal $\nearrow$ $\uparrow$ $\nwarrow$ constant se

unbiased

- **Bca intervals** make three corrections:
  1. for non-normality — using bootstrap percentiles
  2. for bias — using proportion $\hat{\theta}^* < \hat{\theta}$
  3. for changing se — "acceleration"

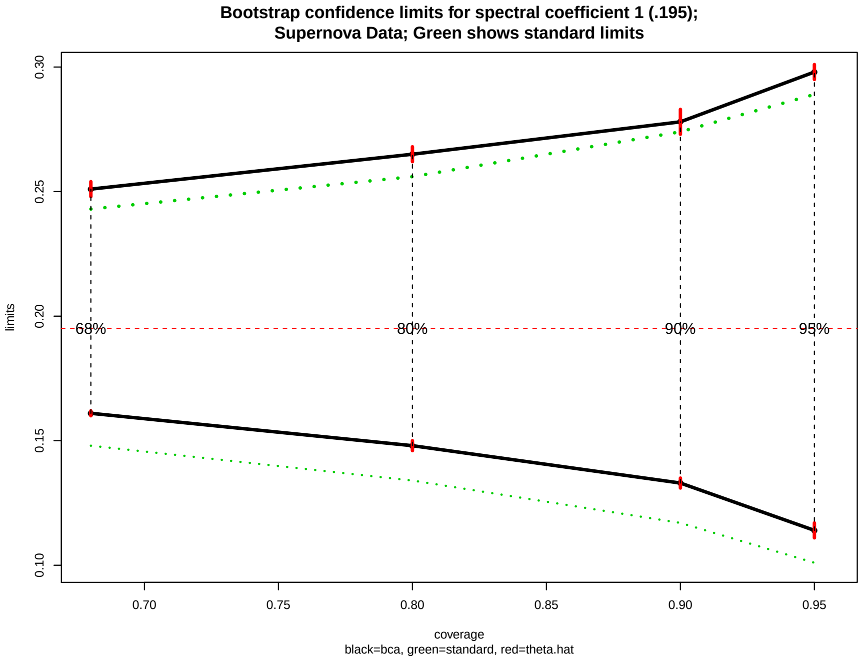
The Supernova Prediction Problem

- Data $\{v_i = (\boldsymbol{s}_i, x_i) \text{ for } i = 1, 2, \ldots, 75\}$
  - $\boldsymbol{s}_i = 12$ spectral energy measurements
  - $x_i =$ absolute magnitude
- **Goal:** Predict x_i from vector $\boldsymbol{s}_i$
- Model $\boldsymbol{x} = \boldsymbol{S}\beta + \boldsymbol{\epsilon}$
  - $\underset{75\times 12}{\boldsymbol{S}}$ rows $\boldsymbol{s}_i$
  - $\boldsymbol{\epsilon}$ normal noise

$\longrightarrow \hat{\beta}_1$ (coef of iron) $= 0.194$ strongly significant

Nonparametric Bootstrap Confidence Intervals

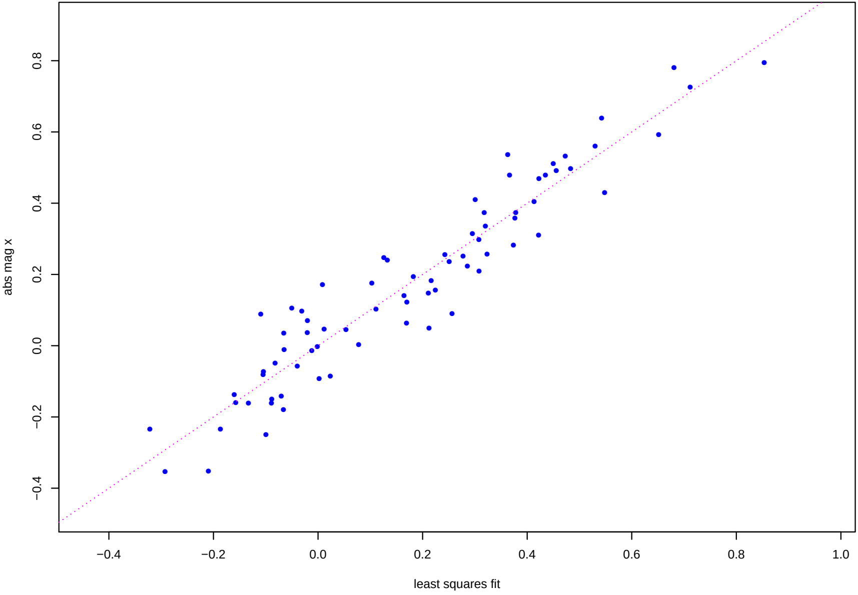
- Need:
  - $V =$ supernova data matrix (each $(\boldsymbol{s}_i, x_i)$ one row)
  - $B =$ number bootstrap replications (≈ 2000)
  - $f =$ statistic of interest ($\hat{\theta} = f(V)$, i.e., $\hat{\beta}_1$)
- R package `bcaboot`: `bcajack`(V, B, f)


Bootstrap confidence limits for spectral coefficient 1 (.195);
Supernova Data; Green shows standard limits
limits
0.10
0.15
0.20
0.25
0.30
68%
80%
90%
95%
0.70
0.75
0.80
0.85
0.90
0.95
coverage
black=bca, green=standard, red=theta.hat

	bca limits	standard limits	jackknife sd's
.025	.114	.101	.003
.05	.133	.117	.002
.16	.161	.148	.001
.5	.206	.195	.001
.9	.265	.256	.003
.95	.278	.274	.005
.975	.298	.289	.003

	theta	boot se	bias	accel
estimate	.195	.048	.262	.023
jack sd	.000	.001	.020	.000

Observed absolute magnitudes versus spectral msmnts LS fit; Adjusted R2 = .919



Parametric Bootstrap Confidence Intervals

- *Supernova model:* $\boldsymbol{x} = \boldsymbol{S}\beta + \boldsymbol{\epsilon}$
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- $\boldsymbol{S}$ fixed, only $\boldsymbol{\epsilon}$ random
- Parameter (β, σ^2)
- $\hat{\beta}$ gives $\hat{\boldsymbol{\mu}} = \boldsymbol{S}\hat{\beta}$

$$R^2 = 1 - \frac{\sum_1^{75}(x_i - \hat{\mu}_i)^2}{\sum_1^{75}(x_i - \bar{x})^2} = 0.933$$

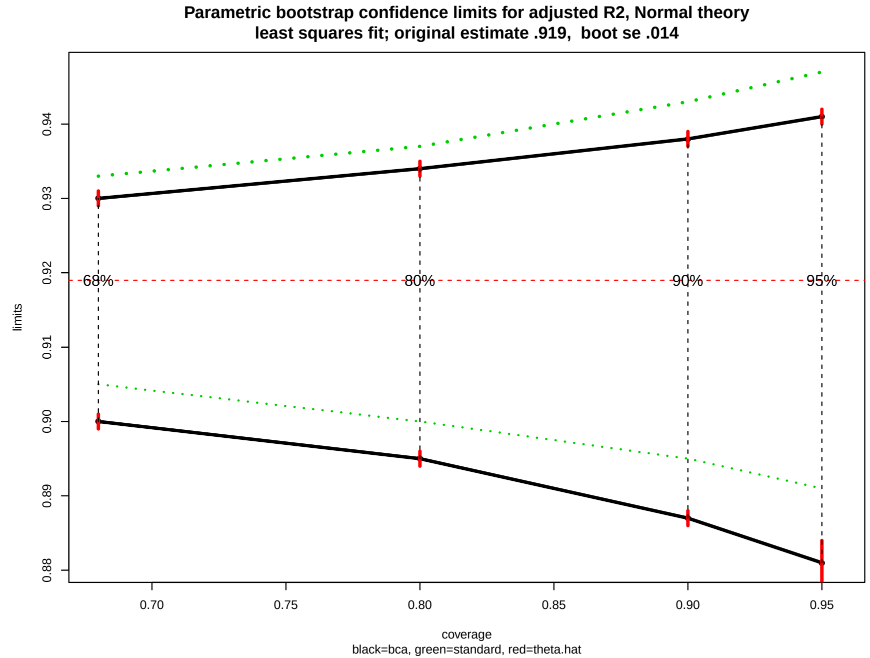
- "Adjusted R^2" $= 0.919 \pm ??$

Program `bcapar`

from R package bcaboot

- Parametric bootstrap sample $\boldsymbol{x}^* \sim \mathcal{N}_{75}(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2)$
- Rerun linear model with data $\Big(\underset{75\times 12}{\boldsymbol{S}}, \boldsymbol{x}^*\Big)$
  gives $\hat{\theta}^* =$ adjusted R^2
- Also need "sufficient statistic" $b^* = (\boldsymbol{S}^\top \boldsymbol{x}^*, \|\boldsymbol{x}^*\|^2)$
- B *replications* $\Big(\hat{\theta}^*(j), b^*(j)\Big)$, $j = 1, 2, \ldots, 2000$

$$\texttt{bcapar}(\underset{0.919}{\hat{\theta}}, \underset{2000}{\hat{\theta}^*}, \underset{2000\times 13}{\boldsymbol{b}^*})$$


Parametric bootstrap confidence limits for adjusted R2, Normal theory
least squares fit; original estimate .919, boot se .014
limits
0.88
0.89
0.90
0.91
0.92
0.93
0.94
68%
80%
90%
95%
0.70
0.75
0.80
0.85
0.90
0.95
coverage
black=bca, green=standard, red=theta.hat

References

1. Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York. *Chapters 6–8 discuss bootstrap standard errors in various settings.*
2. Efron, B. and Narasimhan, B. (2020). "The automatic construction of bootstrap confidence intervals." *Journal of Computational and Graphical Statistics*, Taylor & Francis Online 1–32. *Gives a careful description of bootstrap confidence intervals, both nonparametric and parametric, and the package* `bcaboot`*.*