Motivation

Recall pivotal quantity, $h(\theta; \boldsymbol{D})$,

(i) which involves only the parameter θ of interest (but no other unknown parameter) and data $\boldsymbol{D} \equiv \{X_1, \ldots, X_n\}$, [easy]

(ii) whose distribution is completely known. [hard]

Question Can we overcome the difficulty with a numeric trick of some sort? [Yes!]

The Bootstrap

- if repeatedly sample with replacement from data $\boldsymbol{D} \Rightarrow$ can get many slightly different data samples

$$\boldsymbol{D}^{*(1)}, \boldsymbol{D}^{*(2)}, \ldots, \boldsymbol{D}^{*(B)}$$

  of the <u>same size</u> n;

- can approximate the distribution of $h(\theta; \boldsymbol{D})$ with the empirical distribution (i.e., the histogram) of

$$\left\{ h(\widehat{\theta}; \boldsymbol{D}^{*(1)}), h(\widehat{\theta}; \boldsymbol{D}^{*(2)}), \ldots, h(\widehat{\theta}; \boldsymbol{D}^{*(B)}) \right\},$$

  where $\widehat{\theta}$ estimated from $\boldsymbol{D}$

Volume 47 • Issue 8

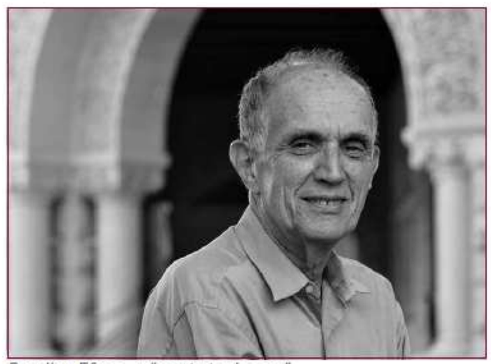
IMS Bulletin

December 2018

CONTENTS

International Prize in Statistics

The International Prize in Statistics is awarded to Bradley Efron, Professor of Statistics and Biomedical Data Science at Stanford University, in recognition of the "bootstrap," a method he developed in 1977 for assessing the uncertainty of scientific results that has had extraordinary and enormous impact across many scientific fields.



Bradley Efron, a "statistical poet"

With the bootstrap, scientists were able to learn from limited data in a simple way that enabled them to assess the uncertainty of their findings. In essence, it became possible to simulate a potentially infinite number of datasets from an original dataset, and in looking at the differences, measure the uncertainty of the result from the original data analysis. Made possible by computing, the bootstrap

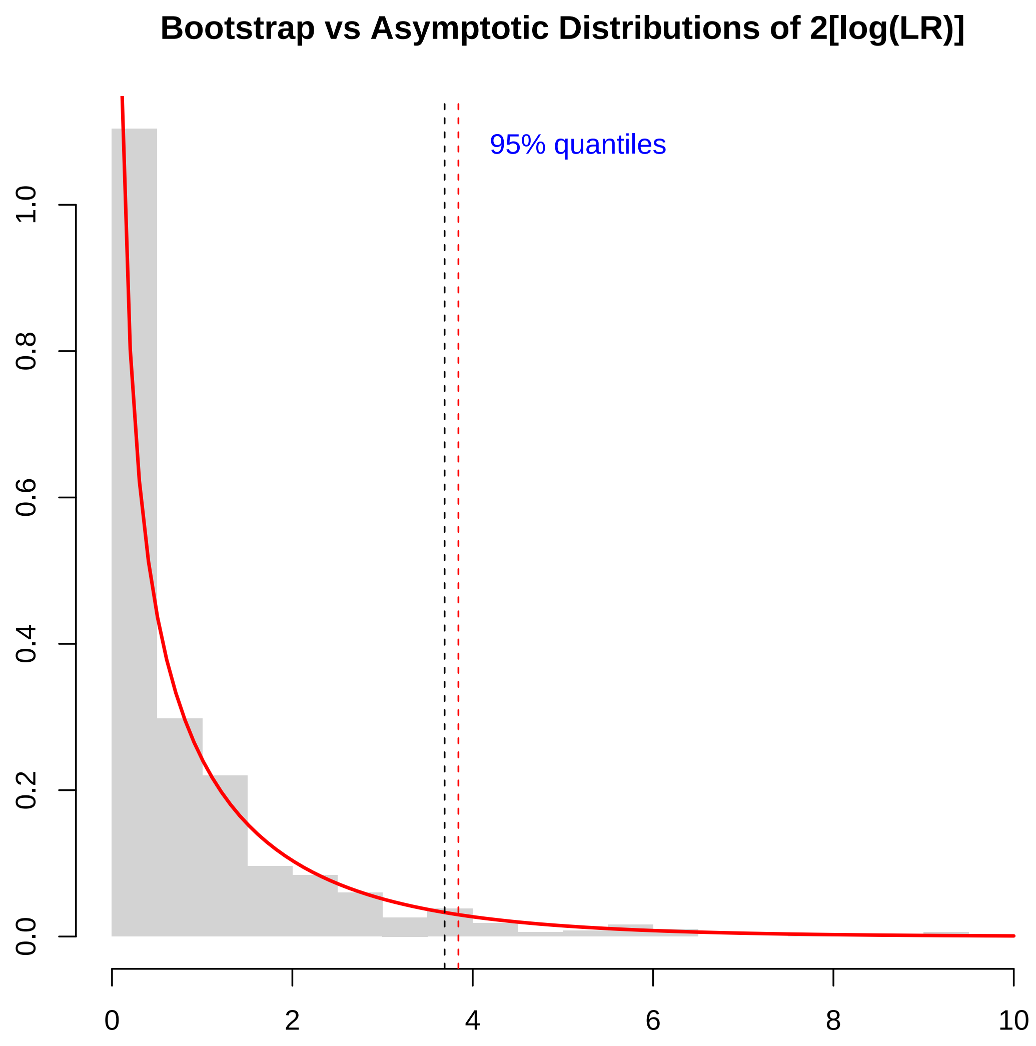
Exercise 7.5 (p. 131)

Let

$$h(\theta; \boldsymbol{D}) = 2[\ell(\widehat{\theta}; \boldsymbol{D}) - \ell(\theta; \boldsymbol{D})]$$

be the twice-log-transformed likelihood ratio, where $\widehat{\theta}$ is the MLE of θ based on $\boldsymbol{D}$. Verify empirically that the bootstrap is in agreement with the asymptotic result (last lecture) about how this quantity is distributed.

Example Use the log-likelihood $\ell(\theta)$ from the censored Poisson problem (where $X_i \sim \text{Poisson}(\theta v_i)$ but we observe only whether $X_i > 0$ or $X_i = 0$). Let $\widehat{\theta}^{*(b)}$ denote the MLE of θ based on the re-sampled data $\boldsymbol{D}^{*(b)}$. Examine the empirical distribution (i.e., histogram) of $h(\widehat{\theta}; \boldsymbol{D}^{*(b)}) = 2[\ell(\widehat{\theta}^{*(b)}; \boldsymbol{D}^{*(b)}) - \ell(\widehat{\theta}; \boldsymbol{D}^{*(b)})]$.



Bootstrap vs Asymptotic Distributions of 2[log(LR)]
95% quantiles
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

Confidence Set

$$\left\{h(\widehat{\theta};\boldsymbol{D}^{*(1)}),h(\widehat{\theta};\boldsymbol{D}^{*(2)}),\ldots,h(\widehat{\theta};\boldsymbol{D}^{*(B)})\right\}$$

$$\Downarrow$$

identify set C^*_α [e.g., empirical quantiles $(c^{*\text{lo}}_\alpha, c^{*\text{hi}}_\alpha)$]

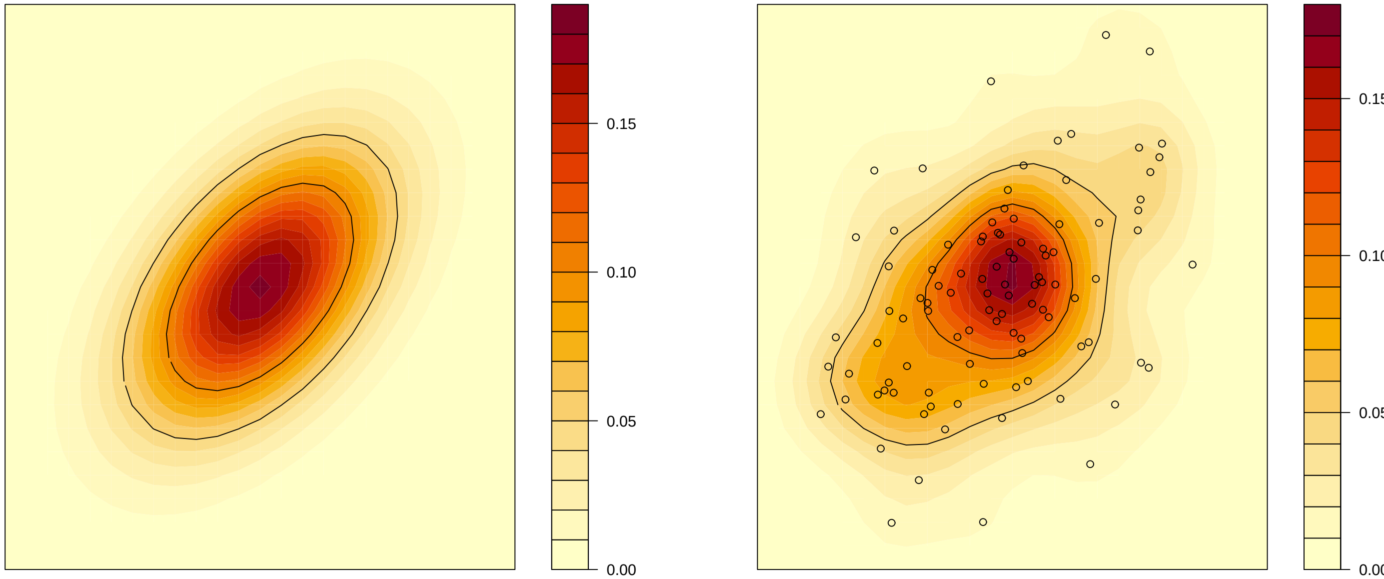
$$\text{s.t.}\quad \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\left\{h(\widehat{\theta};\boldsymbol{D}^{*(b)})\in C^*_\alpha\right\}=1-\alpha$$

$$\Downarrow$$

$$\mathbb{P}(h(\theta;\boldsymbol{D})\in C^*_\alpha)\approx 1-\alpha,$$

$$\Downarrow$$

$$h^{-1}(C^*_\alpha;\boldsymbol{D})\equiv\{\theta: h(\theta;\boldsymbol{D})\in C^*_\alpha\}$$



Contours of C_α [left] and of C^*_α [right], where

$$\mathbb{P}\left(h(\theta;\boldsymbol{D}) \in C_\alpha\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(h(\widehat{\theta};\boldsymbol{D}^{*(b)}) \in C^*_\alpha\right) = 1-\alpha,$$

for different levels of α.

Example 7.7 (pp. 129–131)

$$h(\theta; \boldsymbol{D}) = \widehat{\theta} - \theta \quad \Rightarrow \quad h(\widehat{\theta}; \boldsymbol{D}^{*(b)}) = \underset{\substack{\uparrow \\ \text{Think!}}}{\widehat{\theta}^{*(b)}} - \widehat{\theta}$$

$$\Downarrow$$

$$c_\alpha^{*\text{lo}}, c_\alpha^{*\text{hi}} = \text{empirical quantiles of } \{\overbrace{\widehat{\theta}^{*(b)} - \widehat{\theta}}^{z^{*(b)}}\}_{b=1}^{B},$$

$$\left[\text{e.g., } z^{*(1)} < \ldots < \underset{\substack{\uparrow \\ c_{0.05}^{*\text{lo}}}}{z^{*(25)}} < \quad \ldots \quad \underset{\substack{\uparrow \\ c_{0.05}^{*\text{hi}}}}{z^{*(976)}} < \ldots < z^{*(1000)}\right]$$

$$\Downarrow$$

$$\{\theta : c_\alpha^{*\text{lo}} < \widehat{\theta} - \theta < c_\alpha^{*\text{hi}}\} \Leftrightarrow (\widehat{\theta} - c_\alpha^{*\text{hi}},\ \widehat{\theta} - c_\alpha^{*\text{lo}})$$

Example 7.7 (pp. 129–131)

If

$$q_\alpha^{*\text{lo}},\ q_\alpha^{*\text{hi}} = \text{empirical quantiles of } \{\widehat{\theta}^{*(b)}\}_{b=1}^{B},$$

$$\text{rather than those of } \{\widehat{\theta}^{*(b)} - \widehat{\theta}\}_{b=1}^{B},$$

then

$$c_\alpha^{*\text{lo}} = q_\alpha^{*\text{lo}} - \widehat{\theta} \quad \text{and} \quad c_\alpha^{*\text{hi}} = q_\alpha^{*\text{hi}} - \widehat{\theta}, \quad \text{[Why?]}$$

so

$$(\widehat{\theta} - c_\alpha^{*\text{hi}},\ \widehat{\theta} - c_\alpha^{*\text{lo}}) \Leftrightarrow (2\widehat{\theta} - q_\alpha^{*\text{hi}},\ 2\widehat{\theta} - q_\alpha^{*\text{lo}}).$$

Remark Widely quoted form for a bootstrap confidence interval, but many people don't really understand where the "extra factor of 2" comes from! Beware of which quantiles are being used!!

An Important “Special” Case

$$h(\theta;\boldsymbol{D}) = h(\boldsymbol{D}) \quad [\text{e.g., } h(\boldsymbol{D}) = \widehat{\theta}]$$

“merely” a function of data alone

$$\Downarrow$$

$$\left\{h(\boldsymbol{D}^{*(1)}),\ldots,h(\boldsymbol{D}^{*(B)})\right\} \quad [\text{e.g., } \{\widehat{\theta}^{*(1)},\ldots,\widehat{\theta}^{*(B)}\}]$$

allows us to estimate the sampling distribution of $h(\boldsymbol{D})$ [e.g.,$\widehat{\theta}$]

<u>Example</u>

$$\mathbb{V}\text{ar}(\widehat{\theta}) \approx \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\theta}^{*(b)} - \overline{\theta^{*}})^2 \quad \text{where} \quad \overline{\theta^{*}} \equiv \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{*(b)}$$

Exercise Again, censored Poisson problem (where $X_i \sim \text{Poisson}(\theta v_i)$ but we observe only whether $X_i > 0$ or $X_i = 0$). How big is $\mathbb{V}\text{ar}(\widehat{\theta})$?

Test of Significance

- a related—but seemingly different—problem, which also requires explicit consideration of uncertainty

- want to decide if a particular statement about θ, e.g.,

$$H_0 : \quad \theta = \theta_0 \quad [\text{the null hypothesis}],$$

  is supported by data

- intuitively, check if $\widehat{\theta}$ is "close enough" to θ_0

- judgment of "close enough" has to do with uncertainty

Formal Framework

- rely on a test statistic, $t(\theta; \boldsymbol{D})$ [e.g., $t(\theta; \boldsymbol{D}) = |\widehat{\theta} - \theta|$]

- reject H_0 if $t(\theta_0; \boldsymbol{D}) \geq c$

- choose decision threshold c s.t. the significance level† is controlled at a pre-specified level, say α, i.e., solve

$$\mathbb{P}[t(\theta_0; \boldsymbol{D}) \geq c_\alpha] = \alpha$$

- conventionally, take $\alpha = 0.05, 0.01$

†**Definition** The probability of a test incorrectly rejecting H_0 when H_0 is actually true is called its significance level.

Test Statistic vs Pivotal Quantity

Like $h(\theta; \boldsymbol{D})$, the test statistic $t(\theta; \boldsymbol{D})$ must also be a quantity

(i) which depends only on the parameter of interest θ and data $\boldsymbol{D}$, and

(ii) whose distribution is completely known.

Property (i) is necessary so we can decide whether the hypothesis is supported by data.

Property (ii) is necessary for choosing the right decision threshold c_α.

Proposition

(i) If a test, which rejects $H_0 : \theta = \theta_0$ when $t(\theta_0; \boldsymbol{D}) \geq c_\alpha$, has a significance level of α, then $\{\theta : t(\theta; \boldsymbol{D}) < c_\alpha\}$ will contain the true parameter with probability $(1-\alpha) \times 100\%$.

(ii) Conversely, if $\{\theta : h(\theta; \boldsymbol{D}) \in C_\alpha\}$ contains the true parameter with probability $(1-\alpha) \times 100\%$, then a test, which rejects $H_0 : \theta = \theta_0$ when $h(\theta_0; \boldsymbol{D}) \notin C_\alpha$, will have a significance level of α.

Exercise Prove the proposition above.

Examples

t Test Reject $H_0 : \mu = \mu_0$ if

$$T(\mu_0; \boldsymbol{D}) \equiv \frac{|\bar{X} - \mu_0|}{S/\sqrt{n}} \geq c_\alpha.$$

Choose c_α based on a t-distribution. [Example 8.1 (p. 135)]

Likelihood Ratio Test (LRT) Reject $H_0 : \theta = \theta_0$ if

$$2 \log \Lambda(\theta_0; \boldsymbol{D}) \equiv 2[\ell(\widehat{\theta}_{mle}, \widehat{\psi}_{mle}) - \ell(\theta_0, \widehat{\psi}_{mle|\theta_0})] \geq c_\alpha.$$

Choose c_α based on a χ^2-distribution. [Example 8.2 (p. 136)]

LRT Very Flexible

$$H_0 : \theta \in \Omega_0$$

[other than simply $\Omega_0 = \{\theta_0\}$]

Example

$$\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2, \quad \Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2\}$$

LRT

[assuming no nuisance parameter here]

$$\Lambda(\Omega_0; \boldsymbol{D}) = \frac{\sup\limits_{\theta \in \Omega_0^c} L(\theta)}{\sup\limits_{\theta \in \Omega_0} L(\theta)} \quad \Rightarrow \quad 2[\ell(\widehat{\theta}_{mle|\Omega_0^c}) - \ell(\widehat{\theta}_{mle|\Omega_0})] \sim \chi^2_{(df)}$$

$$df = \dim(\Omega_0^c) - \dim(\Omega_0)$$

Exercise 8.2 (p. 137)

Given $\{X_1, ..., X_n \overset{iid}{\sim} \text{Poisson}(\lambda_1)\} \perp\!\!\!\perp \{Y_1, ..., Y_m \overset{iid}{\sim} \text{Poisson}(\lambda_2)\}$, want to test $H_0 : \lambda_1 = \lambda_2$. Using the LRT, we get

$$\Lambda = \frac{\prod_{i=1}^{n} \frac{e^{-\widehat{\lambda}_1}(\widehat{\lambda}_1)^{X_i}}{X_i!} \prod_{j=1}^{m} \frac{e^{-\widehat{\lambda}_2}(\widehat{\lambda}_2)^{Y_j}}{Y_j!}}{\prod_{i=1}^{n} \frac{e^{-\widehat{\lambda}}(\widehat{\lambda})^{X_i}}{X_i!} \prod_{j=1}^{m} \frac{e^{-\widehat{\lambda}}(\widehat{\lambda})^{Y_j}}{Y_j!}},$$

$\Big\}_{\Omega_0^c=\{\lambda_1\neq\lambda_2\};\ \dim(\Omega_0^c)=2}$

$\Big\}_{\Omega_0=\{\lambda_1=\lambda_2\};\ \dim(\Omega_0)=1}$

where

$$\widehat{\lambda}_1 = \bar{X}, \quad \widehat{\lambda}_2 = \bar{Y}, \quad \widehat{\lambda} = (n\bar{X} + m\bar{Y})/(n+m).$$

Compute $2 \log \Lambda = ...$ and compare with $\chi^2_{(1)}$ threshold.

Think How to do this with the bootstrap if you didn't know about the (approximate) distribution of $2 \log \Lambda$?