Confidence Intervals and the Stability of Standard Errors

Bradley Efron

Stanford University
From a Proportional Hazards Analysis

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\theta}$</th>
<th>$\hat{\text{se}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>Age</td>
<td>-0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Date</td>
<td>-1.35</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Standard 0.95 interval  $\hat{\theta} \pm 1.96 \cdot \hat{\text{se}}$
A Tacit Assumption of the Standard Intervals

- That the standard error is constant over the interval ("stability") or at least changes much more slowly than $\theta$

- **This talk** Begin with measures of stability

- **Eventual goal** Better confidence intervals

- **Poisson example** $y \sim \text{Poi}(\mu)$, observe $y = 16$, $\hat{se} = 4$
Poisson model $y \sim \text{Poi}(\mu)$: observe $y=16$, $\text{se}=4$; 95% standard interval is $[8.2, 23.9]$; endpoint sterrrs are 2.9 and 4.9.
One-Parameter Exponential Families

- \( f_\mu(y) = e^{\eta y} f_0(y) / c(\eta) \) [Poisson: \( \eta = \log(\mu) \)]
- \( y \) is “sufficient statistic”
- \( \mu = E\{y\} \) “expectation parameter”
- \( \hat{\mu} = y \)
- \( \eta = \eta(\mu) \) “natural parameter”
- \( se(\mu) = \) standard deviation of \( y \)
- Fundamental fact

\[
\frac{d\; se_\mu(\hat{\mu})}{d\mu} = \frac{\text{SKEW}_\mu\{y\}}{2}
\]
“Big-A” Acceleration

- $\theta = t(\mu) = \theta_\mu$ parameter of interest [e.g., $\theta = \log(\mu)$]
- MLE $\hat{\theta} = t(\hat{\mu})$

$$A(\hat{\theta}) = \left. \frac{d \text{se}_\mu(\hat{\theta})}{d\theta_\mu} \right|_{\hat{\mu}}$$

(how fast se is changing as $\hat{\theta}$ changes)

- Fundamental fact says $A(\hat{\mu}) = \text{SKEW}_{\hat{\mu}}(y)/2$
- Poisson $A(\hat{\mu}) = \frac{1}{2 \sqrt{\hat{\mu}}} \quad (= 0.125 \text{ for } \hat{\mu} = 16)$
Poisson example: green points are actual sterrrs

se

8.2 16 23.9

A=.125
Poisson Parameter Examples

- $y \sim \text{Poi}(\mu)$, $y = \hat{\mu} = 16$:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Acceleration $A(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>.125 expectation</td>
</tr>
<tr>
<td>$\sqrt{\mu}$</td>
<td>.006 variance stabilized</td>
</tr>
<tr>
<td>$\log(\mu)$</td>
<td>−.146 natural</td>
</tr>
</tbody>
</table>

- Last two by bootstrap calculations...
A Multiparameter Example

- 22 students each took 5 tests
- Data matrix
  \[ \mathbf{x}_{22 \times 5} = (x_1, x_2, \ldots, x_{22})' \]
- **Multivariate normal model** \( x_i \overset{iid}{\sim} \mathcal{N}_5(\gamma, \Sigma) \)
- \( \theta = \text{tr}(\Sigma) \)
- MLE \( \hat{\theta} = \text{tr}(\hat{\Sigma}) = 976 \)
- \( \text{se}(\hat{\theta})? \) How stable?
The Student Score Data
(Mardia, Kent and Bibby 2003)

<table>
<thead>
<tr>
<th>student</th>
<th>mech</th>
<th>vecs</th>
<th>alg</th>
<th>analy</th>
<th>stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>51</td>
<td>43</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>69</td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>41</td>
<td>61</td>
<td>49</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>59</td>
<td>70</td>
<td>68</td>
<td>62</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>42</td>
<td>50</td>
<td>47</td>
<td>29</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>21</td>
<td>46</td>
<td>49</td>
<td>53</td>
<td>59</td>
<td>37</td>
</tr>
<tr>
<td>( n = 22 )</td>
<td>63</td>
<td>63</td>
<td>65</td>
<td>70</td>
<td>63</td>
</tr>
</tbody>
</table>
**Parametric Bootstrap Replications**

- $\mathcal{N}_5(\gamma, \Sigma) \overset{iid}{\longrightarrow} (x_1, x_2, \ldots, x_{22}) \overset{\text{mle}}{\longrightarrow} (\hat{\gamma}, \hat{\Sigma}) \longrightarrow \hat{\theta} = \text{tr}(\hat{\Sigma}) = 976$

- **Boots** $\mathcal{N}_5(\hat{\gamma}, \hat{\Sigma}) \longrightarrow (x_1^*, x_2^*, x_{22}^*) \longrightarrow (\hat{\gamma}^*, \hat{\Sigma}^*) \longrightarrow \hat{\theta}^* = \text{tr}(\hat{\Sigma}^*)$

  $\longrightarrow \{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\}$

- **Bootstrap standard error**

  \[
  \hat{\text{se}} = \left[ \sum_{i=1}^{B} \frac{(\hat{\theta}_i^* - \hat{\theta}^*)^2}{B} \right]^{1/2}
  \]

- $B = 2000$ gave $\hat{\text{se}} = 211.9$ for $\hat{\theta} = \text{tr}(\hat{\Sigma})$
2000 normal theory bootreps, trace(Cov), student score data; boot se=211.9

$A = .248 \pm .026$

Confidence Intervals & Stability of Standard Errors
**Multiparameter Exponential Families**

- $f_{\mu}(y) = e^{\eta'y} f_0(y)/c(\eta)$
- $y$ is $p$-dimensional sufficient vector
- $\mu = E_\mu(y)$ “expectation parameter”
- $\hat{\mu} = y$
- $\eta = \eta(\mu)$ “natural parameter”
- $V_\mu = \text{cov}_\mu\{y\}$
- **Student score matrix** $x$ $y$ has $p = 20$:
  - column averages, col² averages, col product averages
Scalar Parameter of Interest

- $\theta = t(\mu)$
- MLE $\hat{\theta} = t(y) = t(\hat{\mu})$
- Student score $\theta = \text{tr}(\Sigma)$, $\hat{\theta} = 976$
- Gradient $p$-vector $\dot{t} = \begin{pmatrix} \frac{\partial t}{\partial \mu_j} \\ \vdots \\ \vdots \end{pmatrix}_{\mu=y}$

$$t(y) \doteq t(\hat{\mu}) + \dot{t}'(y - \hat{\mu})$$

- Approx standard error $\hat{\text{se}}(\hat{\theta}) \doteq (\dot{t}' \hat{V}_\mu \dot{t})^{1/2} = \overline{\text{se}}$
**Parametric Bootstrap Replications**

- **y = MLE \( \hat{\mu} \)**
- **\( f_{\hat{\mu}} \overset{iid}{\longrightarrow} \{y_1^*, y_2^*, \ldots, y_B^*\} \)

\[
\hat{\theta}_i^* = t(y_i^*) \quad i = 1, 2, \ldots, B
\]

- **\( D_i = \hat{\theta}_i^* - \hat{\theta}^* \) and \( d_i = t'(y_i^* - \hat{\mu}) \)**

\[
\hat{se} = \left( \sum_{i=1}^{B} \frac{D_i^2}{B} \right)^{1/2}
\]

(211.9)

\[
\overline{se} = \left( \sum_{i=1}^{B} \frac{d_i^2}{B} \right)^{1/2}
\]

(215.7)
MLE $y$, gradient vector (red), and bootstrap replications $y^*$ (dots); Blue curve: $t(\mu) = t(y) = \theta$
Stein’s Least Favorable Family

- **Idea**  One-parameter subfamily of \( f_\mu(y) \)

\[
\eta_\lambda = \hat{\eta} + \dot{t}\lambda \quad (\hat{\eta} = \text{MLE of } \eta)
\]

- Just as hard to estimate \( \theta_\lambda = t(\mu_\lambda) \) as \( \theta = t(\mu) \)

- Now can use one-parameter formula \( \tilde{A} = \frac{d\ \text{se}_\lambda}{d\theta_\lambda} \bigg|_{\lambda=0} \)

**Theorem**

\[
\tilde{A} = \frac{1}{2} \sum_{i=1}^{B} \frac{D_i^2 d_i / B}{\hat{\text{se}} \ \text{se}^2} = 0.248
\]
**Bagging Formula**

(Efron 2014 *JASA*)

- $\hat{\theta}_i^* = t(y_i^*) = t_i^*$
- **Bagged estimate**
  
  \[ s = \frac{1}{B} \sum_{i=1}^{B} t_i^* : \]

  \[
  \overline{se}(s) = (\hat{\text{cov}}' \hat{\text{V}} \hat{\text{cov}})^{1/2}
  \]

  where $\hat{\text{cov}} = \text{cov}(y_i^*, t_i^*)$, $\hat{\text{V}} = \text{cov}(y_i^*)$

- But bootstrap standard error is

  \[
  \left( \frac{1}{B} \sum t_i^{*2} - \left( \frac{1}{B} \sum t_i^* \right)^2 \right)^{1/2}
  \]
**BCa Confidence Limits**

*(Efron 1987 JASA)*

- **Idea**  Improve convergence rate of standard intervals

$$\theta_{\text{BCa}}(\alpha) = \hat{G}^{-1} \left\{ \Phi \left[ \hat{z}_0 + \frac{\hat{z}_0 + z(\alpha)}{1 - \hat{a}(\hat{z}_0 + z(\alpha))} \right] \right\}$$

- $\hat{G}(t) = \frac{\#\{\hat{\theta}_i^* \leq t\}}{B}$

- $z(\alpha) = \Phi^{-1}(\alpha)$

- $\hat{z}_0 = \Phi^{-1} \left\{ \hat{G}(\hat{\theta}) \right\}$ is the “bias corrector”

- $\hat{a}$ is “little-a” acceleration
**Three Corrections to the Standard Intervals**

- Standard intervals $\hat{\theta} \sim \mathcal{N}(\theta, \text{se}^2)$
  1. Correct for non-normality ($\hat{G}$)
  2. Correct for bias ($\hat{z}_0$)
  3. Correct for acceleration ($\hat{a}$)

- Asymptotic error rate $O(1/n)$ instead of $O(1/\sqrt{n})$
Student score data trace{Cov}: Nonnormality, Downward Bias, and positive acceleration (A=.248)

61% bootreps <= 976

long right tail

Bradley Efron
Stanford University

CONFIDENCE INTERVALS & STABILITY OF STANDARD ERRORS
"LITTLE-A" ACCELERATION

- Least favorable family: \( \eta_{\lambda} = \hat{\eta} + \dot{\lambda} \) (\( \hat{\eta} \) fixed)
- One-parameter exponential family

\[
\text{“y”} = d = \dot{t}'(y^* - \hat{\mu}), \quad \text{“\( \eta \)”} = \lambda \quad (\hat{\lambda} = 0)
\]

\[
\hat{a} = \frac{1}{3} \frac{1}{2} \text{SKEW}_{\lambda}(d) \bigg|_{\lambda=0}
\]

- \( 1/3 \) of big-A acceleration in LFF

  (correcting for non-normality removes \( 2/3 \) of acceleration effect)
Program BCAJ

- Automates calculation of BCa intervals
- Enter $\hat{\theta}$ and $B$ bootreps \( \{y_i^*, \hat{\theta}_i^*\} \)
- Calculates boot cdf $\tilde{G}(\cdot)$
- $\hat{z}_0 = \Phi^{-1}\{\tilde{G}(\hat{\theta})\}$
- Gradient $\dot{t}$ from local linear regression

\[
\hat{\theta}_i^* = \hat{\theta} + \dot{t}'(y_i^* - \hat{\mu}) \quad (y_i^* \text{ near } \hat{\mu})
\]
Estimating “little a”

- Because $d$ is sufficient statistic in LFF

$$\hat{a} = \frac{1}{3} \frac{\text{SKEW}(d)}{2} = \frac{1}{6 \hat{\text{se}}^3} \sum_{i=1}^{B} d_i^3 / B$$

$$\hat{A} = \frac{1}{2 \hat{\text{se}} \hat{\text{se}}^2} \sum_{i=1}^{B} D_i^2 d_i / B, \quad D_i = \hat{\theta}_i^* - \hat{\theta}_*, \quad d_i = \dot{t}' (y_i^* - \hat{\mu})$$

- Alternatively

$$\hat{a}_0 = \Phi^{-1} \left\{ \frac{\# d_i \leq 0}{B} \right\}$$

- Why not

$$\frac{1}{6 \hat{\text{se}}^3} \sum_{i=1}^{B} D_i^3 / B$$
## BCAJ Output for Student Score Data

tr(cov), $B = 2000$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>bcalims</th>
<th>jacksd</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>685</td>
<td>(9.7)</td>
<td>561</td>
</tr>
<tr>
<td>.05</td>
<td>725</td>
<td>(9.9)</td>
<td>628</td>
</tr>
<tr>
<td>.16</td>
<td>834</td>
<td>(10.1)</td>
<td>766</td>
</tr>
<tr>
<td>.84</td>
<td>1332</td>
<td>(39.9)</td>
<td>1187</td>
</tr>
<tr>
<td>.95</td>
<td>1563</td>
<td>(51.8)</td>
<td>1325</td>
</tr>
<tr>
<td>.975</td>
<td>1666</td>
<td>(74.9)</td>
<td>1392</td>
</tr>
</tbody>
</table>

### Estimated Parameters

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\hat{se}$</th>
<th>$\overline{se}$</th>
<th>$\hat{z}_0$</th>
<th>$\hat{a}$</th>
<th>$\hat{A}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>est</td>
<td>976</td>
<td>211.9</td>
<td>215.7</td>
<td>.269</td>
<td>.083</td>
</tr>
<tr>
<td>jsd</td>
<td>0</td>
<td>3.4</td>
<td>3.4</td>
<td>.028</td>
<td>.011</td>
</tr>
</tbody>
</table>
Two-sided bootstrap confidence intervals for \( tr(Cov) \) are shifted up from the Standard intervals.

Black = bca, Green = standard

Coverage limits

68% 80% 90% 95%

976 1666 685
The Bootstrap Sample Size $B$

- Was $B = 2000$ big enough? Too big?
- `bcaj`:
  - randomly divides the 2000 into 10 groups of 200
  - omits each group in turn, reruns `bca` on 1800
- “jacksd” is jackknife estimate of standard deviation
- Error from stopping at $B = 2000$ rather than $B = \infty$
**Nonparametric BCa is Easier! (bcanonj)**

- **Call:** `bcanonj(x, B, tfunc)`
- **x** is iid sample $(x_1, x_2, \ldots, x_n)$
- **B** is boot sample size
- **tfunc** is R function: $\hat{\theta} = tfunc(x)$
- Computes $B$ bootreps $\hat{\theta}^* = t(x^*)$, $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$, then full BCa output
**BCANONJ OUTPUT FOR STUDENT SCORE DATA**

\[ \text{tr(cov), } B = 2000 \]

<table>
<thead>
<tr>
<th>( bcalims )</th>
<th>( jacksd )</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>649 (10.23)</td>
<td>557</td>
</tr>
<tr>
<td>.05</td>
<td>704 (11.66)</td>
<td>625</td>
</tr>
<tr>
<td>.16</td>
<td>823 (8.62)</td>
<td>764</td>
</tr>
<tr>
<td>.84</td>
<td>1276 (10.54)</td>
<td>1189</td>
</tr>
<tr>
<td>.95</td>
<td>1418 (30.68)</td>
<td>1328</td>
</tr>
<tr>
<td>.975</td>
<td>1553 (16.43)</td>
<td>1396</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \hat{\theta} )</th>
<th>( \hat{\sigma} )</th>
<th>( \hat{z}_0 )</th>
<th>( \hat{a} )</th>
<th>( \hat{A} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{est}</td>
<td>976</td>
<td>213.87</td>
<td>.217</td>
<td>.077</td>
</tr>
<tr>
<td>\text{jsd}</td>
<td>0</td>
<td>3.08</td>
<td>.027</td>
<td>.000</td>
</tr>
</tbody>
</table>
Parametric bca conlims, B=2000
student score trace(Cov)

Now nonparametric limits,
B=2000 bootreps

black = bca, green=standard
Guatemala Abandonment Study

- 500 pediatric cancer subjects
- 47 abandoned by family

\[
\text{coxph}(\text{Aband} \sim \text{Distance} + \text{Age} + \text{Sex} + \text{Date})
\]

Date: \( \hat{\theta} = -1.35 \quad \hat{se} = 0.18 \quad (\hat{se}_{\text{boot}} = 0.165) \)

\( S = \text{Surv(Aband, days since entry)} \)

\( \text{bcanonj}(S, 2000, \text{coxph...}) \)
2000 bootreps of 'Date' coef in coxph for Abandonment study; seh=.165, z0=.128, a=.021

-2.0 -1.8 -1.6 -1.4 -1.2 -1.0 -0.8 0 20 40 60 80 100 120

\( \theta_{\text{hat}} = -1.35 \)

55%

Long left tail
bca confidence intervals for 'Date' coefficient

black = bca, green=standard

coverage limits

68% 80% 90% 95% −1.35

−1.02

−1.65

Bradley Efron

Confidence Intervals & Stability of Standard Errors

Stanford University