Confidence Densities, Uninformative Priors, and the Bootstrap

Bradley Efron

Stanford University

The Holy Grail

- *Multiparameter family* $f_\mu(\boldsymbol{x})$
- *Real-valued parameter of interest* $\theta = t(\mu)$
- The Grail: posterior density for θ given $\boldsymbol{x}$ when prior $\pi(\mu)$ is completely unknown
- Searchers $\longrightarrow$ uninformative priors, matching priors, fiducial methods, confidence densities
- Bootstrap calculations

A Simple Example

- $X \sim \text{Poisson}(\theta)$
- Observe $x = 10$
- $\theta = ??$

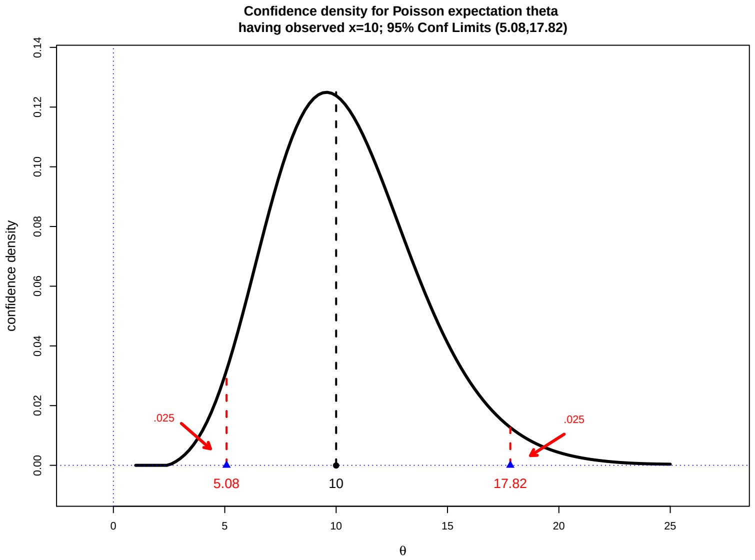
- *Neyman construction* α level confidence limit $\hat{\theta}[\alpha]$:

$$\Pr\left\{X \geq 10 \mid \theta = \hat{\theta}[\alpha]\right\} = \alpha$$

("$\geq$" splits prob atom at $x = 10$)

- **Confidence density** $\text{cd}(\theta)$: $\boxed{\int_0^{\hat{\theta}[\alpha]} \text{cd}(\theta)\, d\theta = \alpha}$ for all α

- Same as fiducial distribution for Poisson example

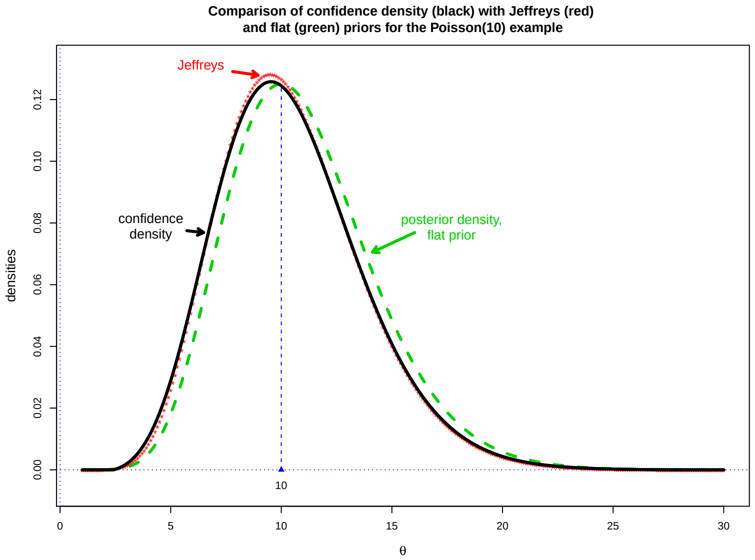

Confidence density for Poisson expectation theta
having observed x=10; 95% Conf Limits (5.08,17.82)
confidence density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
.025
.025
5.08
10
17.82
0
5
10
15
20
25
θ

Jeffreys Prior

- For family $\{f_\mu(\boldsymbol{x})\}$ Jeffreys "uninformative" prior is

$$\pi(\mu) = |I(\mu)|^{1/2} \qquad [I(\mu) = \text{Fisher info matrix}]$$

- *Poisson case:* $\pi(\theta) = 1\Big/\sqrt{\theta}$
- Almost a matching prior (posterior quantiles match $\hat{\theta}[\alpha]$)



Comparison of confidence density (black) with Jeffreys (red)
and flat (green) priors for the Poisson(10) example
Jeffreys
confidence
density
posterior density,
flat prior
densities
0.00
0.02
0.04
0.06
0.08
0.10
0.12
10
0
5
10
15
20
25
30
θ

An Elementary Mistake?

- CD amounts to believing

$$\Pr\left\{\hat{\theta}[0.90] \le \theta \le \hat{\theta}[0.91]\right\} = 0.01 \quad \text{etc.}$$

- *Frequentist justification:*
  - cd(θ) is just a convenient way to describe confidence limits $\hat{\theta}[\alpha]$
  - Many good frequentist properties

    (Xie and Singh, 2013; Schweder and Hjort, 2002)

Bayesian Justification

- Suppose $\pi(\theta)$ were a perfect matching prior:

$$\int_0^{\hat{\theta}[\alpha]} \pi(\theta \mid x)\, d\theta = \alpha \qquad \text{for all } \alpha$$

- Since also $\displaystyle\int_0^{\hat{\theta}[\alpha]} \mathrm{cd}(\theta)\, d\theta = \alpha \qquad \text{for all } \alpha$:

$$\pi(\theta \mid x) = \mathrm{cd}(\theta)$$

- "cd(θ) is posterior density starting from a perfectly uninformative prior"

Fisher's Fiducial Argument

- $x_1, \ldots, x_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$
- $\hat{\theta} = \bar{x}, \quad \widehat{\text{se}}^2 = \frac{\sum (x_i - \bar{x})^2}{n(n-1)}$
- Then $\frac{\hat{\theta} - \theta}{\widehat{\text{se}}} \sim T_{n-1}$ (Student's t):

$$\boxed{\theta = \hat{\theta} - \widehat{\text{se}} \cdot T_{n-1}}$$

- *Fisher:* sufficient stats $(\hat{\theta}, \widehat{\text{se}})$ exhaust information in $\boldsymbol{x}$, after which T_{n-1} is irreducible component of remaining randomness
- Fiducial limits same as usual t limits

Nuisance Parameters

- Parametric family $f_\mu(\boldsymbol{x}), \quad \mu \in \mathcal{R}^p$
- Real-valued parameter of interest $\theta = t(\mu)$
- How to get rid of the $p - 1$ nuisance parameters?
- *Bayes* Integrate them out: $\pi(\theta \mid \boldsymbol{x}) = \int_{t(\mu)=\theta} \pi(\mu \mid \boldsymbol{x})\, d\mu$
- $\pi(\mu)$ uninformative?

Example: Student Score Data

(Mardia, Kent and Bibby 2003)

- $n = 22$ students each scored on 5 tests: $\boldsymbol{x} = \begin{pmatrix} -x_1- \\ -x_2- \\ \vdots \\ -x_n- \end{pmatrix}$

- Multivariate normal model:

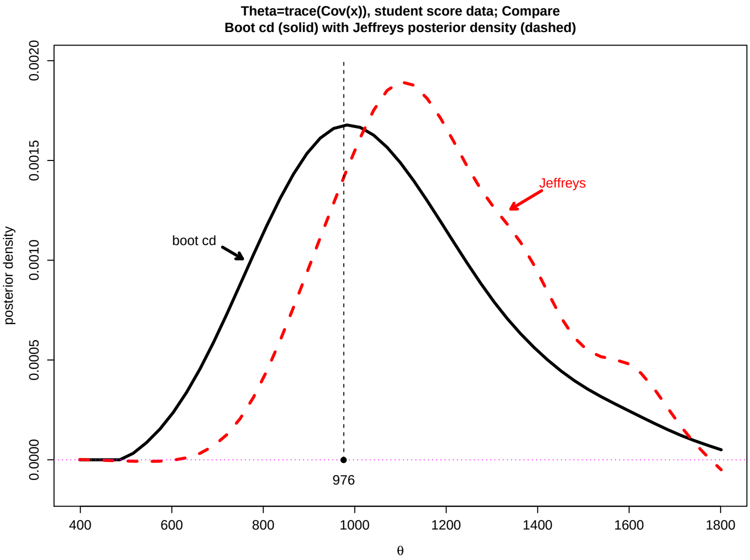
$$x_j \overset{\text{iid}}{\sim} \mathcal{N}_5(\mu, \Sigma) \qquad [j = 1, \ldots, n]$$

- $\theta = \text{trace}(\Sigma)$
- $\hat{\theta} = 976$

- *No exact confidence limits*

The Student Score Data

student	mech	vecs	alg	analy	stat
1	7	51	43	17	22
2	44	69	53	53	53
3	49	41	61	49	64
4	59	70	68	62	56
5	34	42	50	47	29
⋮	⋮	⋮	⋮	⋮	⋮
21	46	49	53	59	37
22	63	63	65	70	63



Theta=trace(Cov(x)), student score data; Compare
Boot cd (solid) with Jeffreys posterior density (dashed)
posterior density
0.0000
0.0005
0.0010
0.0015
0.0020
boot cd
Jeffreys
976
400
600
800
1000
1200
1400
1600
1800
θ

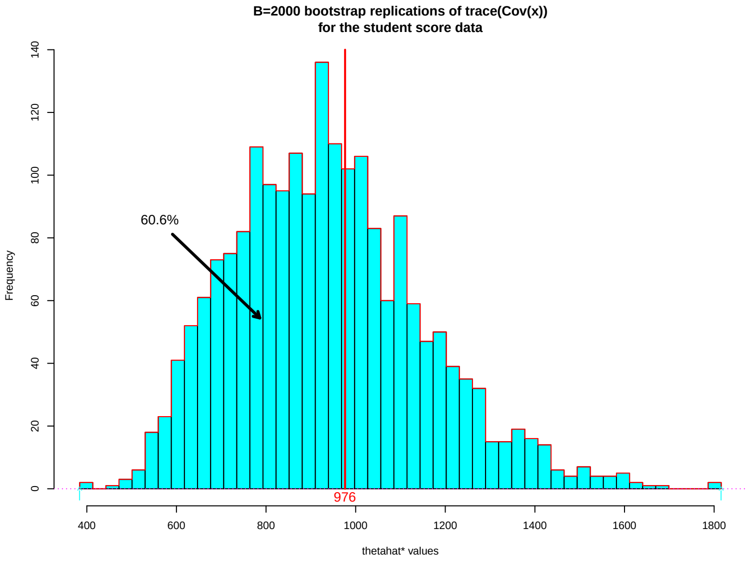
The Parametric Bootstrap

- **Parametric family** $f_\mu(\boldsymbol{x}) \to \boldsymbol{x} \to \text{MLE } \hat{\mu} \to \hat{\theta} = t(\hat{\mu})$
- *Resampling* from $f_{\hat{\mu}}(\cdot)$:

$$f_{\hat{\mu}} \longrightarrow \boldsymbol{x}^* \longrightarrow \hat{\mu}^* \longrightarrow \hat{\theta}^* = t(\hat{\mu}^*)$$

- *Bootstrap CDF:* Resampling B times ($B \sim 2000$)

$$\boxed{\widehat{G}(t) = \#\left\{\hat{\theta}^{*i} \leq t\right\} \Big/ B}$$



B=2000 bootstrap replications of trace(Cov(x))
for the student score data
60.6%
976
Frequency
thetahat* values
0
20
40
60
80
100
120
140
400
600
800
1000
1200
1400
1600
1800

Bootstrap Confidence Limits

- *BCa confidence limits*

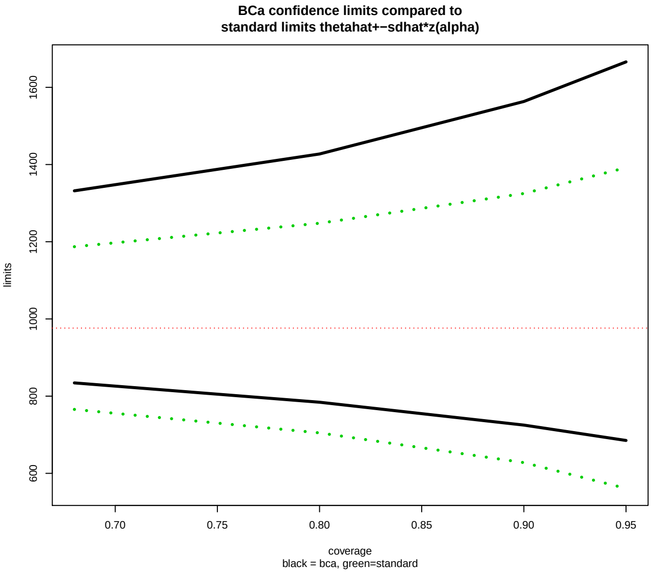
$$\boxed{\hat{\theta}[\alpha] = \widehat{G}^{-1}\Phi\left(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}\right)}$$

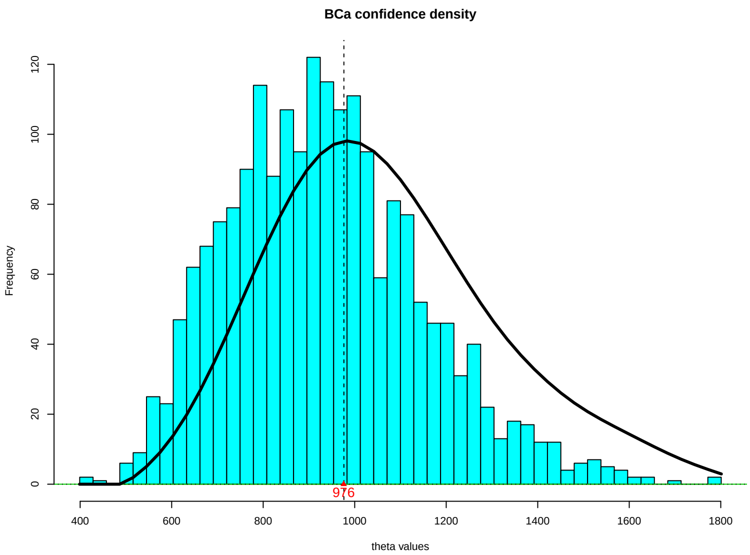
$z^{(\alpha)} = \Phi^{-1}(\alpha)$; z_0 and a calculated from resamples

- *2nd order accuracy:* actual α = claimed $\alpha + O(1/n)$ (Hall, 1988)

- **Confidence density** weight $w\left(\hat{\theta}_i^*\right)$ on $\hat{\theta}_i^*$:

$$w(\theta) = \frac{\phi\left[z_\theta/(1 + az_\theta) - z_0\right]}{(1 + az_\theta)^2\phi(z_\theta + z_0)} \qquad \left[z_\theta = \Phi^{-1}\widehat{G}(\theta) - z_0\right]$$


BCa confidence limits compared to
standard limits thetahat+−sdhat*z(alpha)
limits
600
800
1000
1200
1400
1600
0.70
0.75
0.80
0.85
0.90
0.95
coverage
black = bca, green=standard


BCa confidence density
Frequency
0
20
40
60
80
100
120
976
400
600
800
1000
1200
1400
1600
1800
theta values

The BCa Constants z_0 and a

- *Bias-correction estimate* $\hat{z}_0 = \Phi^{-1}\widehat{G}\left(\hat{\theta}\right)$
- *Acceleration* $\hat{a} = {}^1\!/_3 \left.\frac{dsd\{\hat{\theta}\}}{dE\{\hat{\theta}\}}\right|_{\hat{\mu}}$ in "least favorable family"
- p-parameter exp family w/ minimal sufficient statistic $\hat{\beta} = s(\boldsymbol{x})$
- LFF: $\beta = \hat{\beta} + \lambda \widehat{V}\dot{t}$ where $\widehat{V} = \text{cov}\left(\hat{\beta}\right)$ and $\dot{t} = \nabla_\mu t(\mu)\big|_{\hat{\mu}}$
- $a = {}^1\!/_6$ skewness of log likelihood in LFF

Program bcaj for BCa Confidence Density

- Model $x_j \overset{\text{iid}}{\sim} \mathcal{N}_5(\mu, \Sigma)$ is $p = 20$ parameter exp family:

$$\hat{\beta}(\boldsymbol{x}) = \left(\bar{x}_1, \dots, \bar{x}_5, \overline{x_1^2}, \dots, \overline{x_5^2}, \overline{x_1 x_2}, \dots, \overline{x_4 x_5}\right) \qquad [p = 20]$$

- *Bootstrap* $\quad \mathcal{N}_5\left(\hat{\mu}, \hat{\Sigma}\right) \to \boldsymbol{x}^{*i} \to \left(\hat{\beta}^{*i}, t^{*i}\right)$ for $i = 1, \dots, B = 2000$
- Histogram of $\{t^{*i}\}$ gives boot cdf $\widehat{G}$
- Program `bcaj` gives $\hat{\theta}[\alpha]$, $\hat{z}_0$, $\hat{a}$, $\widehat{\mathrm{cd}}$, and also jackknife estimates of internal accuracy

Output of Program bcaj for tr[Cov(x)]

Student Score Data

α	BCa-lims	(jacksd)	Standard
.025	685	(9.81)	561
.05	725	(11.53)	628
.1	784	(8.81)	705
.16	834	(9.73)	766
.5	1034	(10.27)	976
.84	1332	(35.26)	1187
.9	1427	(21.15)	1248
.95	1563	(51.87)	1325
.975	1666	(77.61)	1392

	θ	a	z_0
estimate	976	.083	.269
(jacksd)		(.011)	(.030)

Relation of Jeffreys to BCa

- Parametric bootstrap gives $\left\{\left(\hat{\beta}_i^*, t_i^*\right),\ i = 1, \ldots, B\right\}$
- *Bootstrap distribution* $\widehat{G}$: weight $^1/_B$ on each t_i^*
- *BCa confidence density:* weight w_i^{BCa} on t_i^*
- Jeffreys weight e^{Δ_i} on t_i^* where

$$\Delta_i = \left[\text{Dev}\left(\hat{\beta}_i^*, \hat{\beta}\right) - \text{Dev}\left(\hat{\beta}, \hat{\beta}_i^*\right)\right] / 2$$

Uninformative Priors and Matching Priors

- Change parameter of interest $\theta = \text{tr}(\Sigma)$ to $\theta = \mu' \Sigma^{-1} \mu$
- Weights w_i^{BCa} change but *not* Jeffreys weights

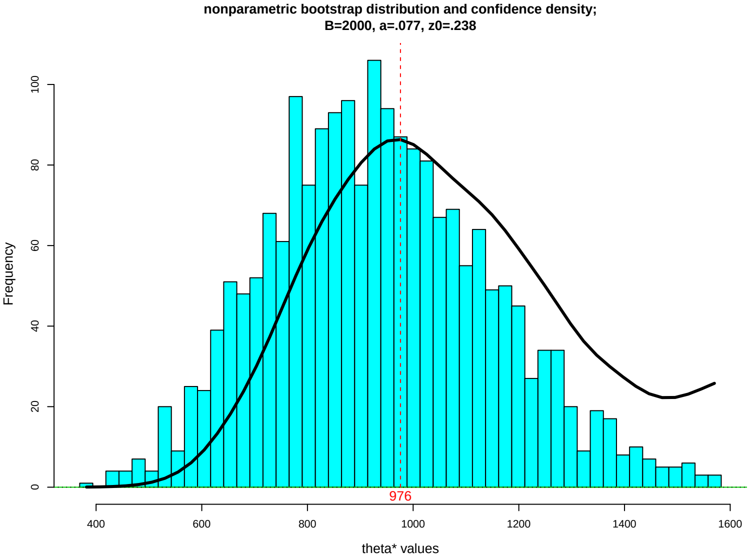
  (Δ_i only depends on β_i^*, not t_i^*)
- No one multidimensional prior $\pi(\mu)$ can be matching for all single parameters $\theta = t(\mu)$
- Bayes and frequentist methods diverge

Nonparametric Confidence Density

- Nonparametric boot sample $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)$ gives $\hat{\theta}^* = t(\boldsymbol{x}^*)$
- B boot samples give $\widehat{G}$, cdf of $\left\{\hat{\theta}^{*i}\right\}$
- $\hat{z}_0 = \Phi^{-1}\left[\widehat{G}\left(\hat{\theta}\right)\right]$
- $\hat{a} = {}^1/_6$ skewness (jackknife values)

$$\left[d_i \equiv \hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)} : \hat{a} = \left(\sum d_i^3/n\right)\Big/\left(\sum d_i^2/n\right)^{3/2}\right]$$

- BCa limits and cd same formulas as before


nonparametric bootstrap distribution and confidence density;
B=2000, a=.077, z0=.238
Frequency
0
20
40
60
80
100
976
400
600
800
1000
1200
1400
1600
theta* values

Priors and Likelihoods

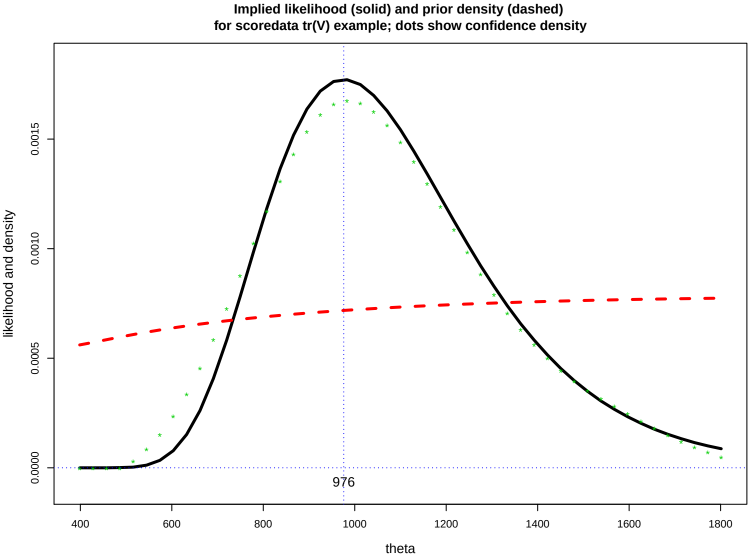
- *Bayes* $\quad \text{cd}_{\boldsymbol{x}}(\theta) = c\pi(\theta)L_{\boldsymbol{x}}(\theta)$? $\qquad [\pi(\theta)$ prior, $L_{\boldsymbol{x}}(\theta)$ likelihood]
- *Exponential families* (Efron, 1993)

$$\boxed{L_{\boldsymbol{x}}(\theta) = \text{cd}_{\boldsymbol{xx}}(\theta) \big/ \text{cd}_{\boldsymbol{x}}(\theta)}$$

$\left(\text{so } \pi(\theta) = {}^{[\text{cd}_{\boldsymbol{x}}(\theta)]^2}/_{\text{cd}_{\boldsymbol{xx}}(\theta)} \text{ is “implied prior”}\right)$

- *Combining independent cd's* $\quad L_{\boldsymbol{x}}(\theta)L_{\boldsymbol{y}}(\theta)$?

(Xie and Singh, 2013)



Implied likelihood (solid) and prior density (dashed)
for scoredata tr(V) example; dots show confidence density
likelihood and density
0.0000
0.0005
0.0010
0.0015
976
400
600
800
1000
1200
1400
1600
1800
theta

Some References

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statist. Sci.* 11: 189–228, with comments and a rejoinder by the authors.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* 80: 3–26, bootstrap-based confidence densities and likelihoods.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* 16: 927–985, with a discussion and a reply by the author.

Schweder, T. and Hjort, N. L. (2002). Confidence and likelihood. *Scand. J. Statist.* 29: 309–332, asymptotic theory of confidence-based likelihoods.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int. Stat. Rev.* 81: 3–39.