Interval Estimation: Confidence Densities, Uninformative Priors, and the Bootstrap

Bradley Efron

Stanford University

The Holy Grail

- *Multiparameter family* $f_\mu(\boldsymbol{x})$: *parameter of interest* $\theta = t(\mu)$
- **Bayes** posterior density of θ given $\boldsymbol{x}$
- 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

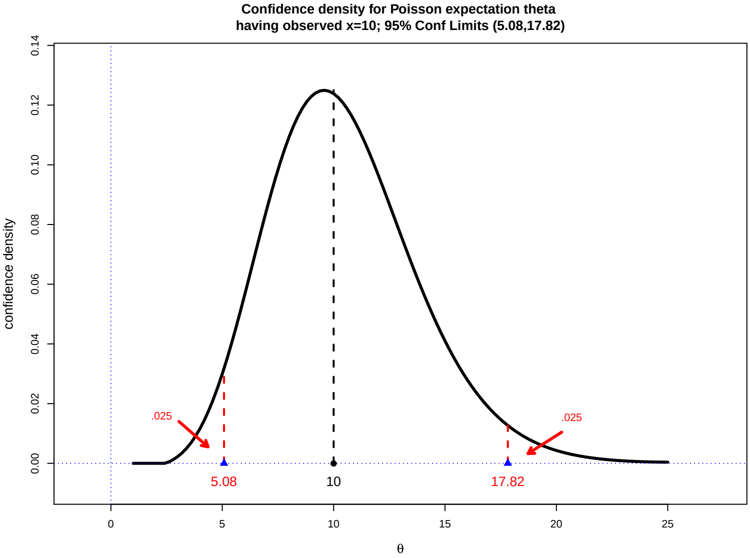
Confidence Distributions and Confidence Densities

- $\theta_{\boldsymbol{x}}[\alpha]$ upper level α confidence limit for θ given $\boldsymbol{x}$
- Confidence distribution $\mathrm{Cd}(\theta)$

  $\Pr\{\theta \in (\theta_{\boldsymbol{x}}[0.90], \theta_{\boldsymbol{x}}[0.91])\} = 0.01$ etc.
- *Confidence density* $\mathrm{cd}(\theta) = \dfrac{d}{d\theta}\mathrm{Cd}(\theta)$

$$\int_{\theta_{\boldsymbol{x}}[\alpha_1]}^{\theta_{\boldsymbol{x}}[\alpha_2]} \mathrm{cd}(\theta)\, d\theta = \alpha_2 - \alpha_1$$

- “Fiducial distribution”


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

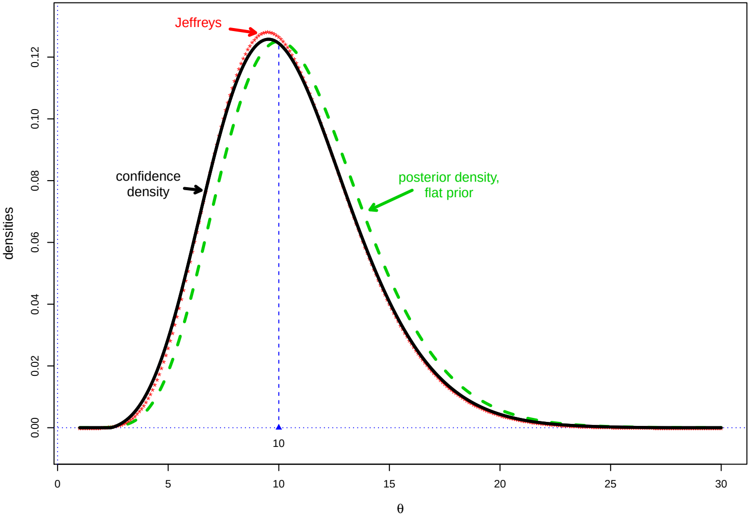
Jeffreys Prior

- 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



An Elementary Mistake?

- CD amounts to believing "wrong" interpretation

  of confidence limits
- *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:

$$\int_{\theta_{\boldsymbol{x}}[\alpha_1]}^{\theta_{\boldsymbol{x}}[\alpha_2]} \pi(\theta \mid x)\, d\theta = \alpha_2 - \alpha_1 \qquad \text{for all } \alpha_1, \alpha_2$$

- Since also $\displaystyle\int_{\theta_{\boldsymbol{x}}[\alpha_1]}^{\theta_{\boldsymbol{x}}[\alpha_2]} \mathrm{cd}(\theta)\, d\theta = \alpha_2 - \alpha_1 \qquad \text{for all } \alpha_1, \alpha_2$:

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

- "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:

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

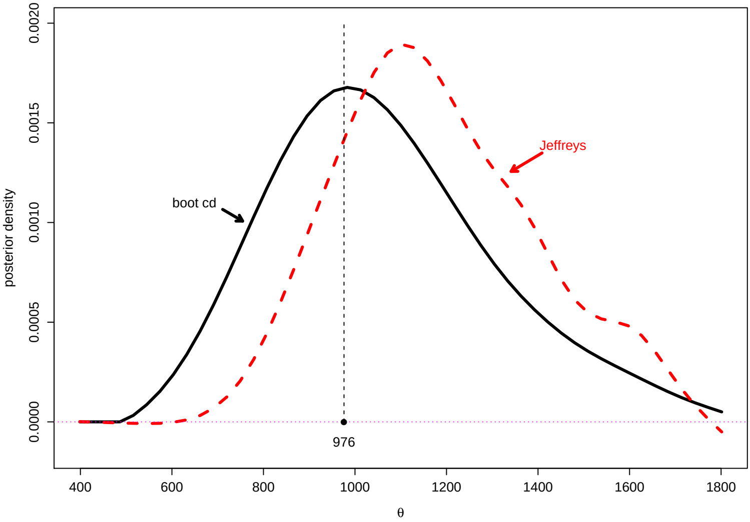
- $\theta = \text{trace}(\Sigma)$
- $\hat{\theta} = 976$

- *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)



The Parametric Bootstrap

- **Parametric family** $f_\mu(\boldsymbol{x}) \rightarrow \boldsymbol{x} \rightarrow \text{MLE } \hat{\mu} \rightarrow \hat{\theta} = t(\hat{\mu})$
- *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 ($B \sim 2000$)

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

Bootstrap Confidence Limits

- *BCa confidence limits*

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

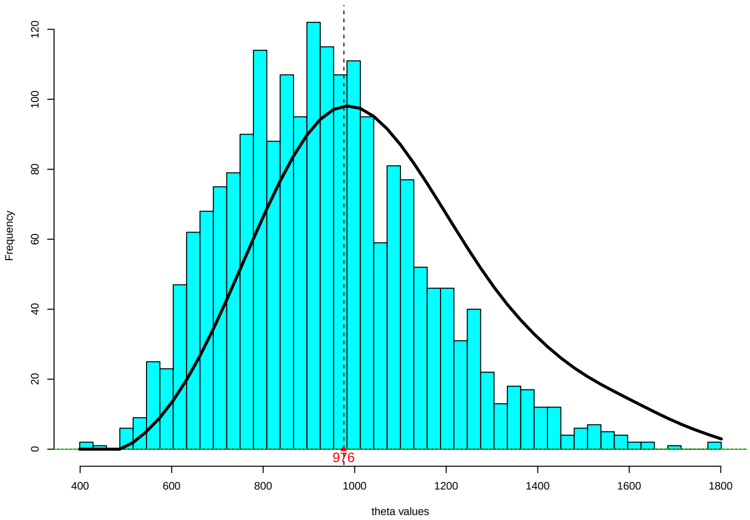
$z^{(\alpha)} = \Phi^{-1}(\alpha)$; z_0 and a calculated from resamples

- *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 density



Program bcaj for BCa Confidence Density

- Model $x_j \stackrel{\text{iid}}{\sim} \mathcal{N}_5(\mu, \Sigma)$ 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* $\mathcal{N}_5\left(\hat{\mu}, \hat{\Sigma}\right) \rightarrow \boldsymbol{x}^{*i} \rightarrow \left(\hat{\beta}^{*i}, \hat{\theta}^{*i}\right)$ for $i = 1, \dots, B = 2000$
- Histogram of $\left\{\hat{\theta}^{*i}\right\}$ gives boot cdf $\widehat{G}$
- Program `bcaj` gives $\hat{\theta}[\alpha]$, $\hat{z}_0$, $\hat{a}$, $\widehat{\text{cd}}$, and also jackknife estimates of internal accuracy (i.e., error from stopping at B)

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)

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

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.