Bayesian Inference and the Parametric Bootstrap

Bradley Efron

Stanford University

Importance Sampling for Bayes Posterior Distribution

- Newton and Raftery (1994 *JRSS-B*)

  "Nonparametric Bootstrap: good choice"

  (actually used a smoothed nonparametric bootstrap)

- Today   Parametric Bootstrap

  – Good computational properties when applicable

  – Connection between Bayes and frequentist inference

Student Score Data
(Mardia, Kent and Bibby)

- $n = 22$ students' scores on two tests: *mechanics*, *vectors*

- Data $\qquad \boldsymbol{y} = (y_1, y_2, \ldots, y_{22})$ with $y_i = (\text{mec}_i, \text{vec}_i)$

- *Parameter of interest* $\qquad \theta = \text{correlation (mec, vec)}$

- *Sample correlation coefficient* $\qquad \hat{\theta} = 0.498 \pm$??

Scores of 22 students on two tests 'mechanics' and 'vectors';
Sample Correlation Coefficient is .498 +−??

mechanics score
Sampled from 88 students, Mardia, Kent and Bibby
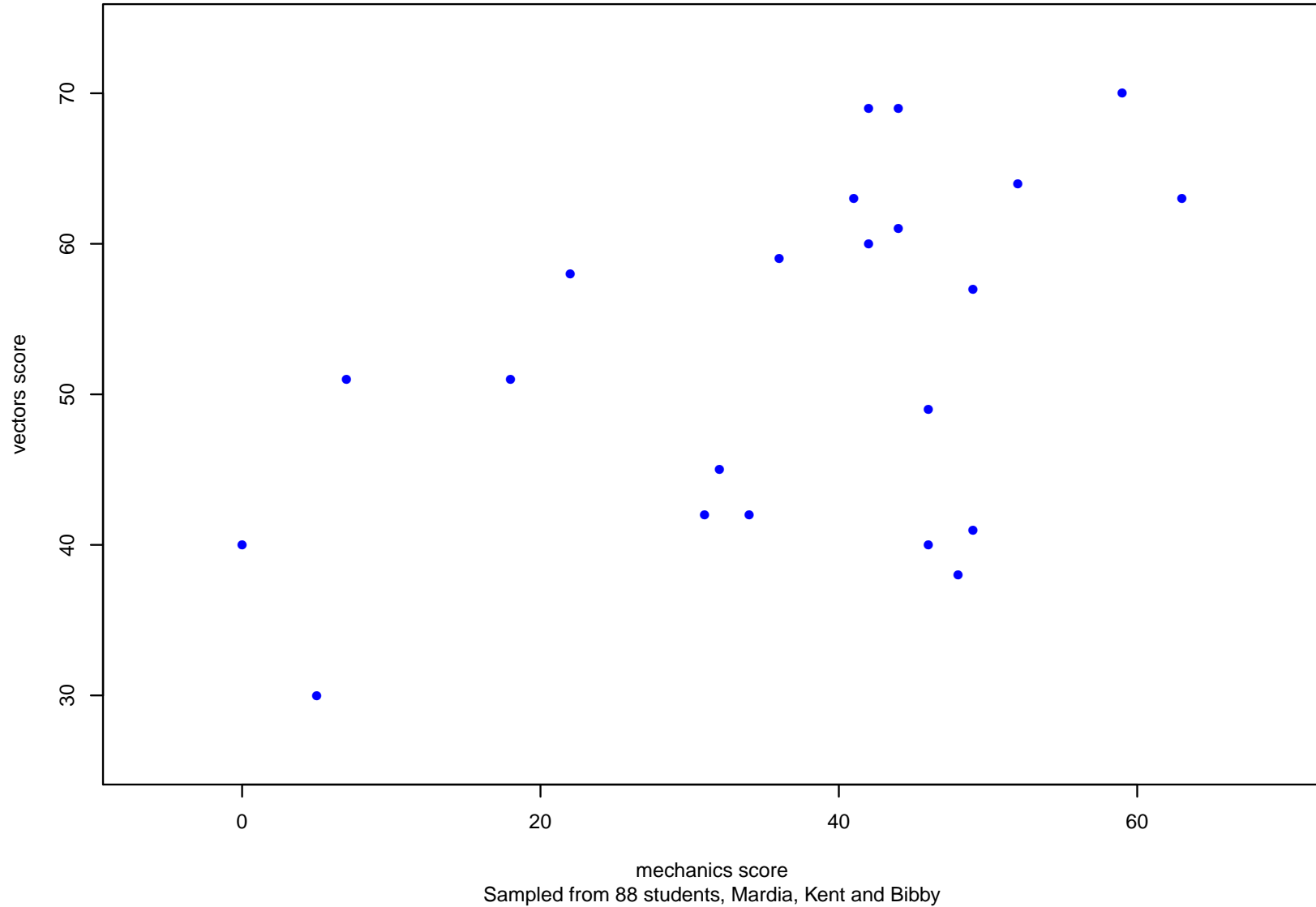
R. A. Fisher (1915)

- Density $f_\theta(\hat{\theta})$

$$= \frac{n-2}{\pi}(1-\theta^2)^{(n-1)/2}\left(1-\hat{\theta}^2\right)^{(n-4)/2}\int_0^\infty \left[\cosh(w) - \theta\hat{\theta}\right]^{-(n-1)} dw$$

- *Bivariate normal* $y_i \stackrel{\text{ind}}{\sim} \mathcal{N}_2(\mu, \Sigma)$ $i = 1, 2, \ldots, n$

- 95% confidence limits (from z transform)

$$\boxed{\theta \in (0.084, 0.754)}$$

Fisher density for correlation coefficient if corr=.498
And histogram for 10000 parametric bootstrap replications

correlation
Red triangles are Fisher 95% confidence limits
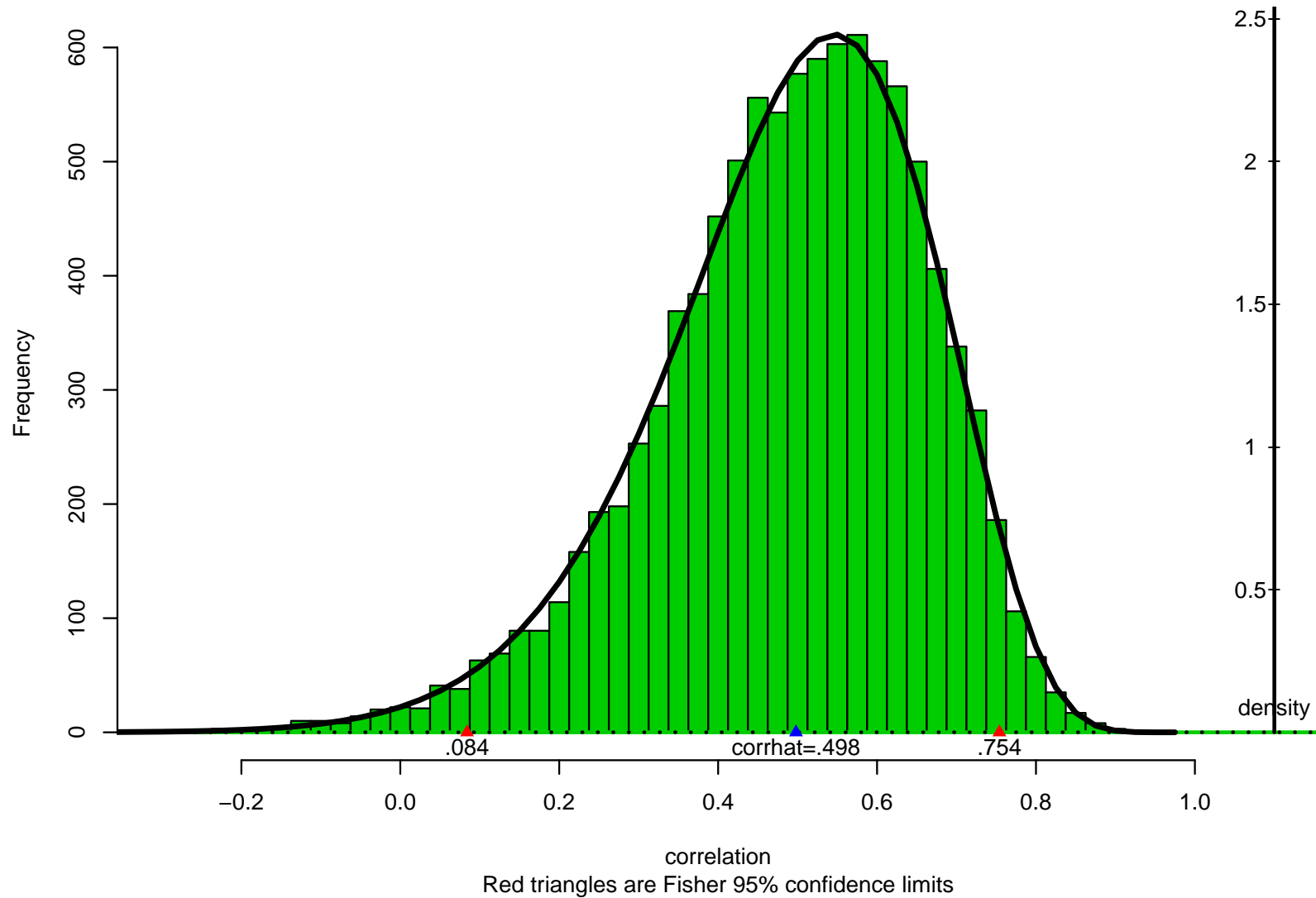
Parametric Bootstrap Computations

- *Assume* $y_i \overset{\text{ind}}{\sim} \mathcal{N}_2(\mu, \Sigma)$ • *Estimate* $\hat{\mu}$ and $\hat{\Sigma}$ by MLE

- Bootstrap sample $y_i^* \overset{\text{ind}}{\sim} \mathcal{N}_2\left(\hat{\mu}, \hat{\Sigma}\right)$ for $i = 1, 2, \ldots, 22$

- Bootstrap replication $\hat{\theta}^* = $ corr coeff for $\boldsymbol{y}^* = (y_1^*, y_2^*, \ldots, y_{22}^*)$

- Bootstrap standard deviation

$$\text{sd}\left(\hat{\theta}\right) = 0.168 \qquad (B = 10,000)$$

- *Percentile 95% confidence limits* $\boxed{(0.109,\ 0.761)}$

Better Bootstrap Confidence Limits (BCa)

- Idea Reweighting the B bootstrap replications improves coverage

- Weights involve z_0 (bias correction) and a (acceleration)

- $$W_{\text{BCa}}(\theta) = \frac{\varphi(Z_\theta/(1 + aZ_\theta) - z_0)}{(1 + aZ_\theta)^2 \varphi(Z_\theta + z_0)} \quad \text{where } Z_\theta = \Phi^{-1}G(\theta) - z_0$$

  $\uparrow$

  bootstrap cdf

- Reweighting the $B = 10,000$ bootstraps gave weighted percentiles

  $\boxed{(0.075,\ 0.748)}$ Fisher: $(0.084,\ 0.754)$

Bayes Posterior Intervals

- *Prior* $\pi(\theta)$ • Posterior expectation for $t(\theta)$:

$$E\left\{t(\theta)|\hat{\theta}\right\} = \int t(\theta)\pi(\theta)f_\theta\left(\hat{\theta}\right)d\theta \Bigg/ \int \pi(\theta)f_\theta\left(\hat{\theta}\right)d\theta$$

- Conversion factor   Ratio of likelihood to bootstrap density:

$$\boxed{R(\theta) = f_\theta\left(\hat{\theta}\right)\Big/ f_{\hat{\theta}}(\theta)} \qquad \left(\text{``}\theta\text{''} = \hat{\theta}^*\right)$$

- Bootstrap integrals

$$E\left\{t(\theta)|\hat{\theta}\right\} = \frac{\int t(\theta)\pi(\theta)R(\theta)f_{\hat{\theta}}(\theta)\,d\theta}{\int \pi(\theta)R(\theta)f_{\hat{\theta}}(\theta)\,d\theta}$$

Bootstrap Estimation of Bayes Expectation $E\{t(\theta)|\hat{\theta}\}$

- Parametric bootstrap replications $f_{\hat{\theta}}(\cdot) \rightarrow \theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_B$

- $t_i = t(\theta_i), \quad \pi_i = \pi(\theta_i), \quad R_i = R(\theta_i)$:

$$\hat{E}\left\{t(\theta)|\hat{\theta}\right\} = \sum_{i=1}^{B} t_i \pi_i R_i \bigg/ \sum_{i=1}^{B} \pi_i R_i$$

- *Reweighting*   Weight $\pi_i R_i$ on θ_i

- Importance sampling estimate:

$$\hat{E}\left\{t(\theta)|\hat{\theta}\right\} \rightarrow E\left\{t(\theta)|\hat{\theta}\right\} \qquad \text{as } B \rightarrow \infty$$

Jeffreys Priors

- Harold Jeffreys (1930s):

  Theory of "uninformative" prior distributions

  ("invariant", "objective", "reference", ...)

- For density family $f_\theta(y)$, $\theta \in \mathcal{R}^p$:

$$\boxed{\pi(\theta) = c|I(\theta)|^{1/2}} \qquad I(\theta) = \text{Fisher info matrix}$$

- *Normal correlation* $\quad \pi(\theta) = 1/(1 - \theta^2)$

**Importance sampling posterior density for correlation (black)
using Jeffreys prior pi(theta)=1/(1−theta^2);
BCa weighted density (green); raw bootstrap density (red)**

correlation
triangles show 95% posterior limits
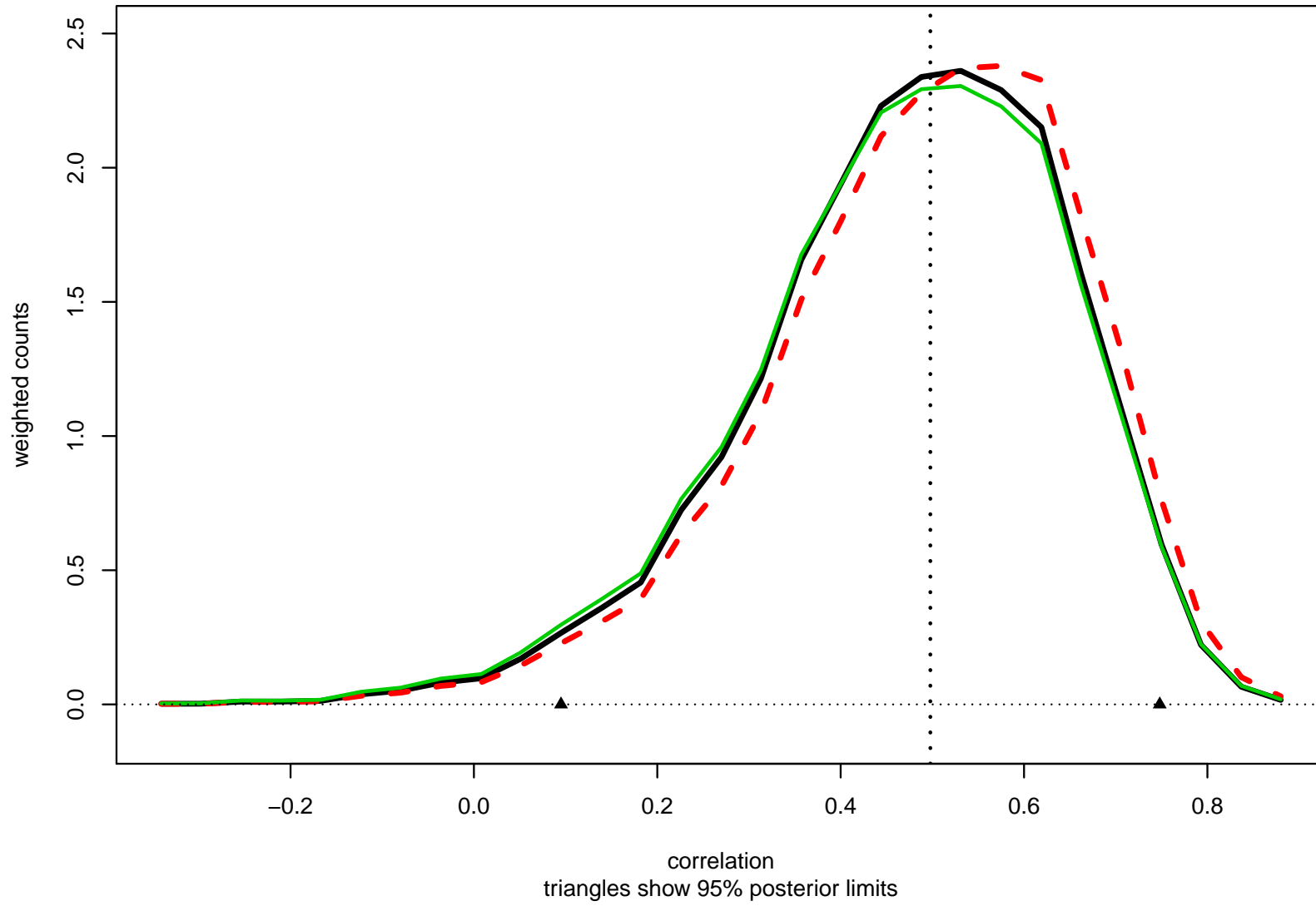
Multiparameter Exponential Families $\mathcal{R}^P$

- p-dimensional sufficient statistic $\hat{\beta}$

- p-dimensional parameter vector $\beta = E\left\{\hat{\beta}\right\}$
  (*Poisson*: $e^{-\lambda}\lambda^x/x!$ has $\hat{\beta} = x,\ \beta = \lambda$)

- Hoeffding $\boxed{f_\beta\left(\hat{\beta}\right) = f_{\hat{\beta}}\left(\hat{\beta}\right)e^{-D(\hat{\beta},\beta)/2}}$

- *Deviance* $\quad D(\beta_1,\beta_2) = 2E_{\beta_1}\log\left\{f_{\beta_1}\left(\hat{\beta}\right)\Big/f_{\beta_2}\left(\hat{\beta}\right)\right\}$

Conversion Factor in Exponential Families

- Conversion factor $\quad R(\beta) = f_\beta\left(\hat{\beta}\right)\big/f_{\hat{\beta}}(\beta) \quad$ (with $\hat{\beta}$ fixed)

$$R(\beta) = \nu(\beta)e^{\Delta(\beta)}$$

  where $\Delta(\beta) = \left[D\left(\beta,\hat{\beta}\right) - D\left(\hat{\beta},\beta\right)\right]\big/2$

- $\nu(\beta) = f_{\hat{\beta}}\left(\hat{\beta}\right)\big/f_\beta(\beta) \doteq 1/\text{Jeffreys prior} \quad$ (Laplace)

- Jeffreys prior $\quad \pi(\beta)R(\beta) \doteq e^{\Delta(\beta)}$

Example: Multivariate Normal

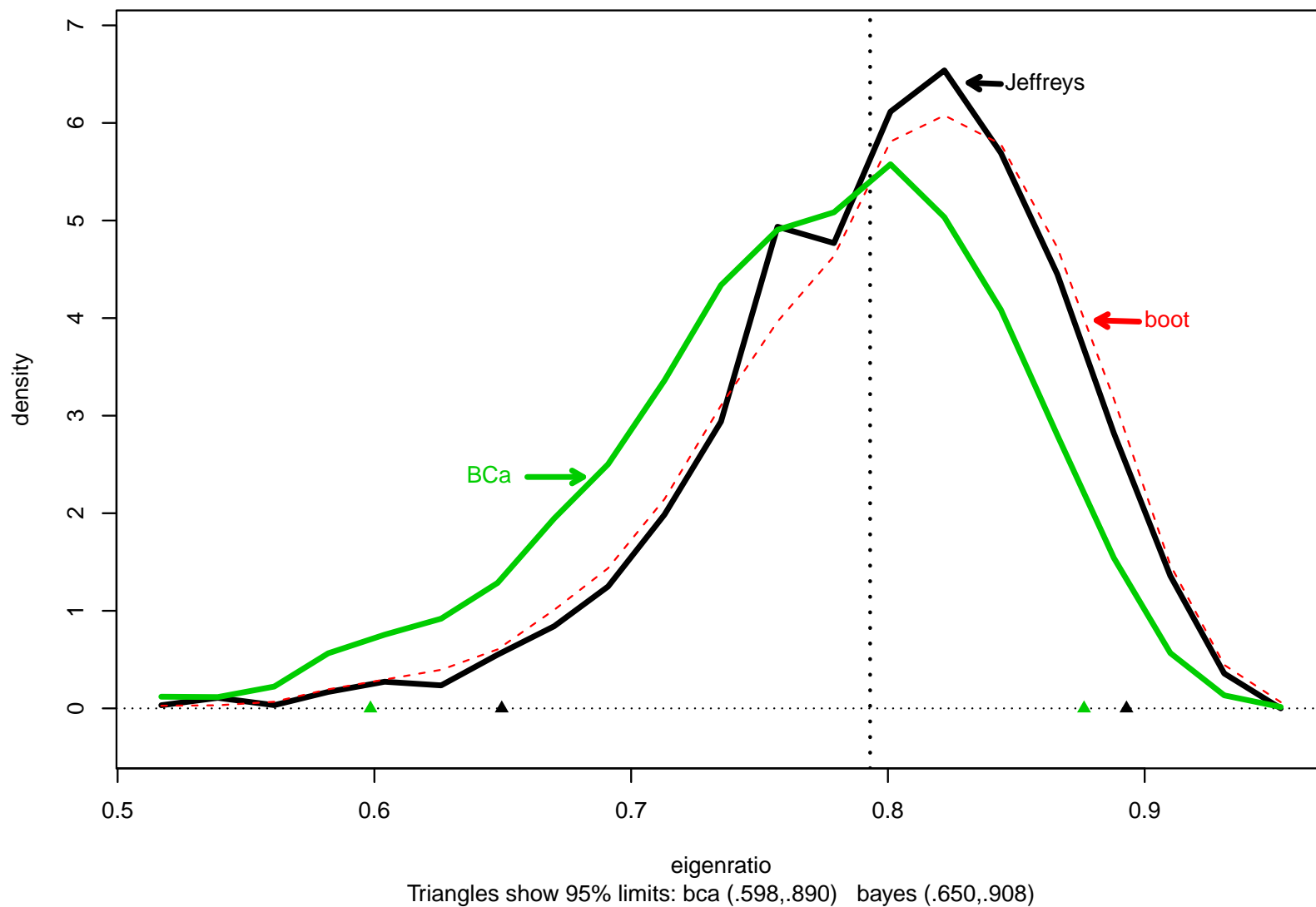- $y_i \stackrel{\text{ind}}{\sim} \mathcal{N}_d(\mu, \Sigma)$ for $i = 1, 2, \ldots, n$

- $\Delta =$

$$n \left\{ (\mu - \hat{\mu})' \frac{\hat{\Sigma}^{-1} - \Sigma^{-1}}{2} (\mu - \hat{\mu}) + \frac{1}{2} \text{tr}\left(\Sigma\hat{\Sigma}^{-1} - \hat{\Sigma}\Sigma^{-1} \right) + \log \frac{|\hat{\Sigma}|}{|\Sigma|} \right\}$$
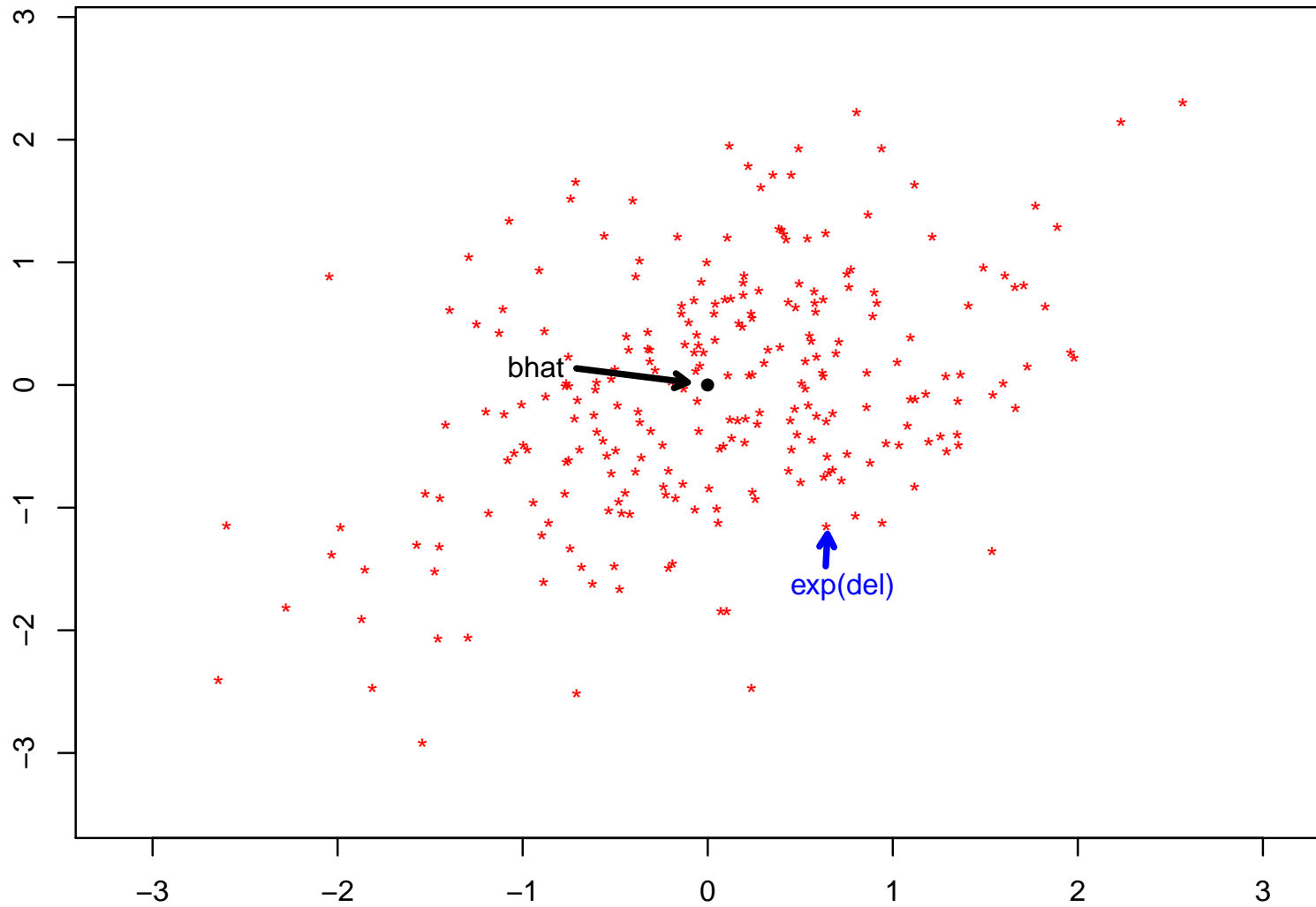
- "eigenratio" (student score data)

- $\theta = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ 　　 $\bullet$ $\hat{\theta} = \dfrac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2} = 0.793 \pm$??

- Bootstrap 　 $y_i^* \stackrel{\text{ind}}{\sim} \mathcal{N}_d\left(\hat{\mu}, \hat{\Sigma} \right)$ 　 $i = 1, 2, \ldots, n \rightarrow \theta_1, \theta_2, \ldots, \theta_B$

Density estimates for student score eigenratio (B=10,000);
Boot (red),Jeffreys (5–dim prior) Black, and BCa Green (z0=−.182, a=0)

eigenratio

Triangles show 95% limits: bca (.598,.890)   bayes (.650,.908)

Reweighting the parametric bootstrap points
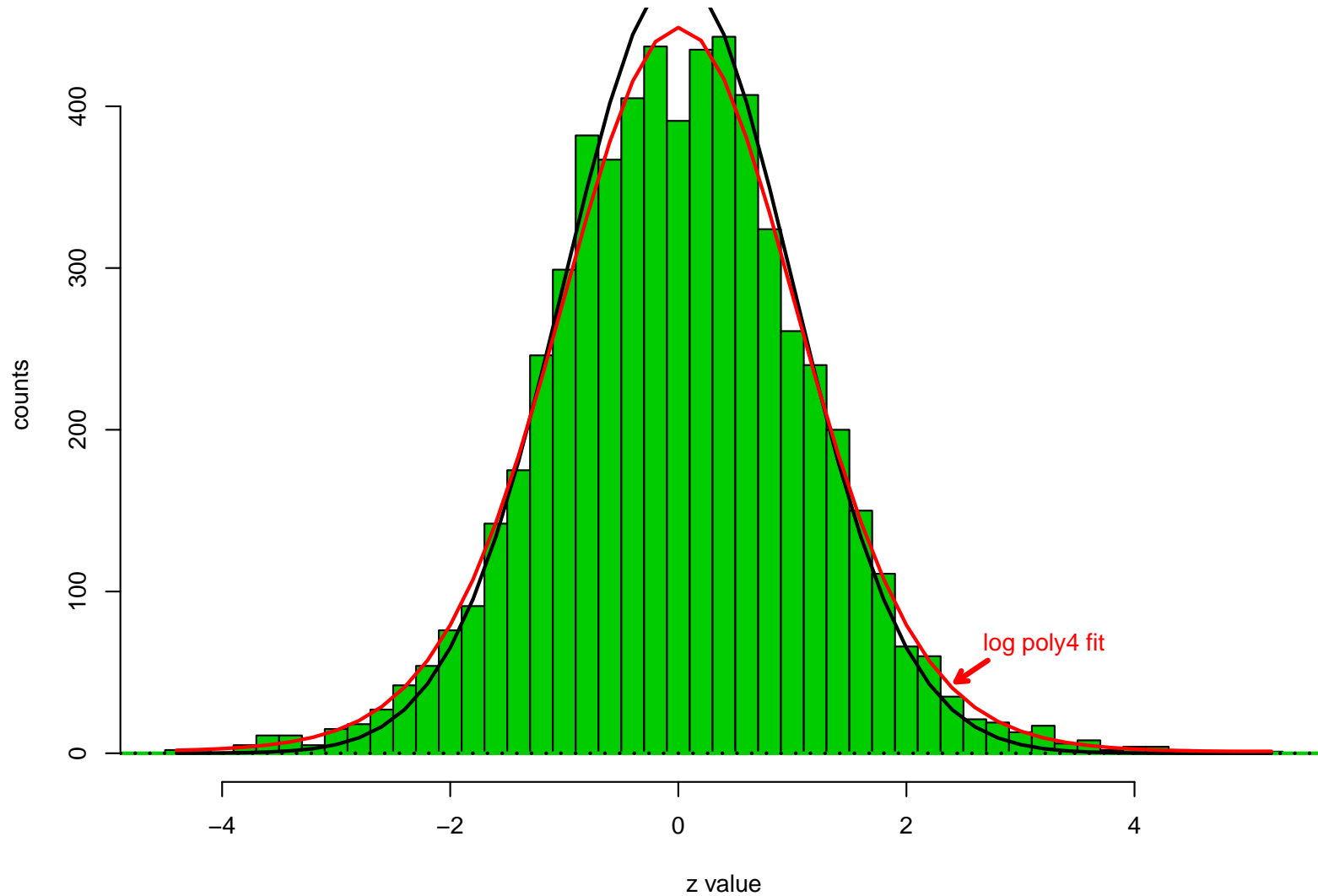
Prostate Cancer Study
(Singh et al, 2002)

- *Microarray study*

  102 men: 52 prostate cancer, 50 healthy controls

- 6033 genes z_i test statistic for H_{0i}: "no difference"

$$\boxed{H_{0i} : z_i \sim \mathcal{N}(0,1)}$$

- *Goal*   Identify genes involved in prostate cancer

Prostate cancer z−values for 6033 genes; 52 patients vs 50 healthy controls. N(0,1) black; LogPoly4 red

log poly4 fit

counts

z value

Poisson Regression Models

- *Histogram* $\quad y_j = \#\{z_i \in \text{bin}_j\}$

- $x_j = \text{midpoint bin}_j, \; j = 1, 2, \ldots, J$

- Poisson model $\quad y_j \overset{\text{ind}}{\sim} \text{Poi}(\mu_j)$
  with $\log(\mu_j) = \text{poly}(x_j, \text{degree} = m)$

- Exponential family $\quad$ degree $p = m + 1$ $\qquad$ • Let $\eta_j = \log(\mu_j)$

$$\Delta = (\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})' \, (\boldsymbol{\mu} + \hat{\boldsymbol{\mu}}) - 2 \left(\sum \mu_j - \sum \hat{\mu}_j \right)$$
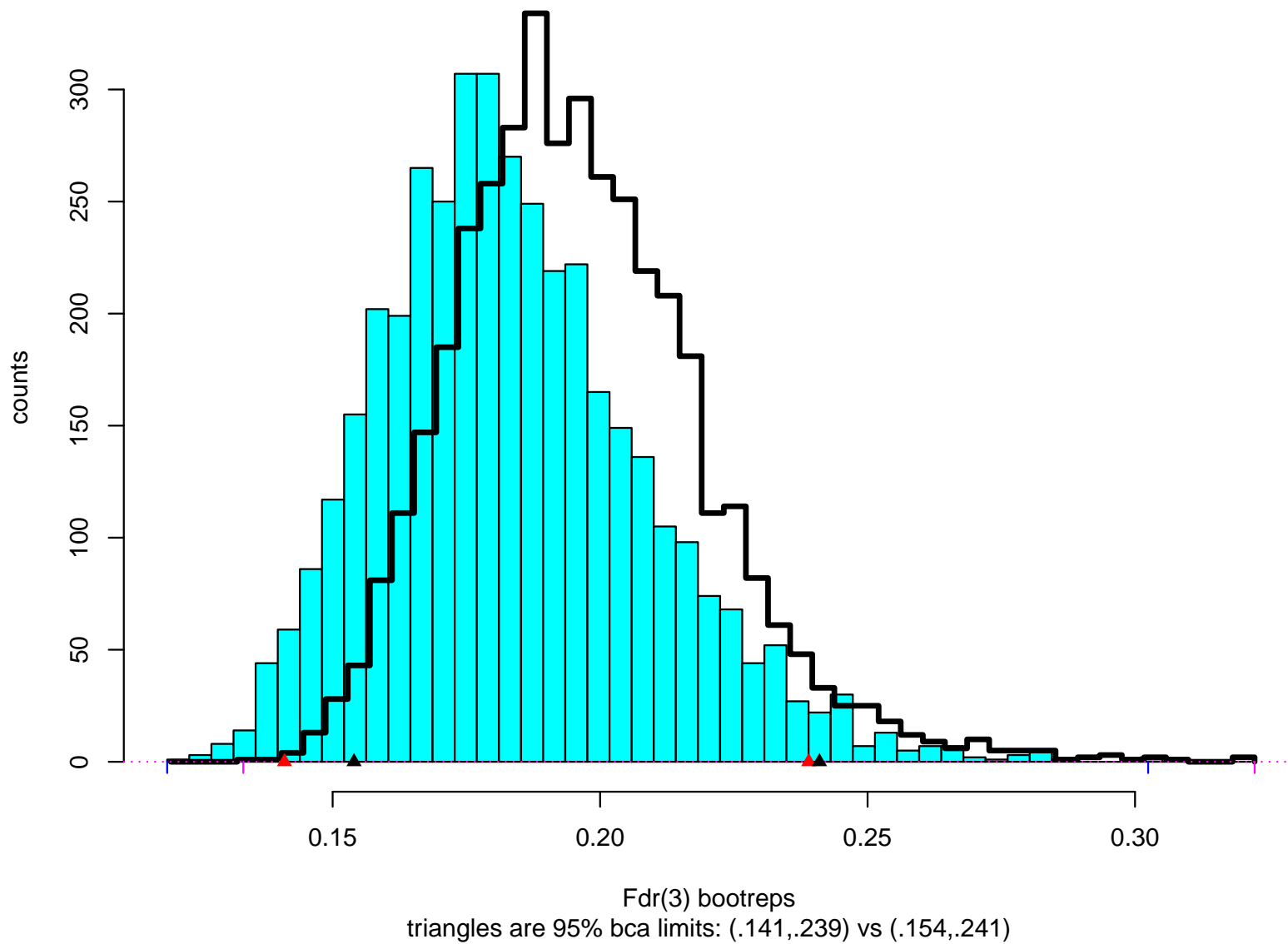
Parametric False Discovery Rate Estimates

- $\texttt{glm(y\sim poly(x,4),Poisson)} \longrightarrow \left\{\hat{\mu}_j\right\}$ ("log poly 4")

- $\theta = \boxed{\widehat{\text{Fdr}}(3) = [1 - \Phi(3)]\Big/\left[1 - \hat{F}(3)\right]} \approx \text{Pr}\{\text{null}|z \geq 3\}$

  ($\hat{F}$ is smoothed CDF estimate)

- Model 4: $\quad \widehat{\text{Fdr}}(3) = 0.192\pm$??

- Parametric bootstrap $\quad y_j^* \stackrel{\text{ind}}{\sim} \text{Poi}\left(\hat{\mu}_j\right) \to \left\{\hat{\mu}_j^*\right\} \to \widehat{\text{Fdr}}(3)^*$

**Fdr(3) estimation, Model 4; prostate data from 4000 parboots;
Jeffreys Bayes (black), raw boot (red), bca (green)**

posterior density

Fdr(3)*
triangles show 95% Jeffreys/BCa limits (.154, .241)
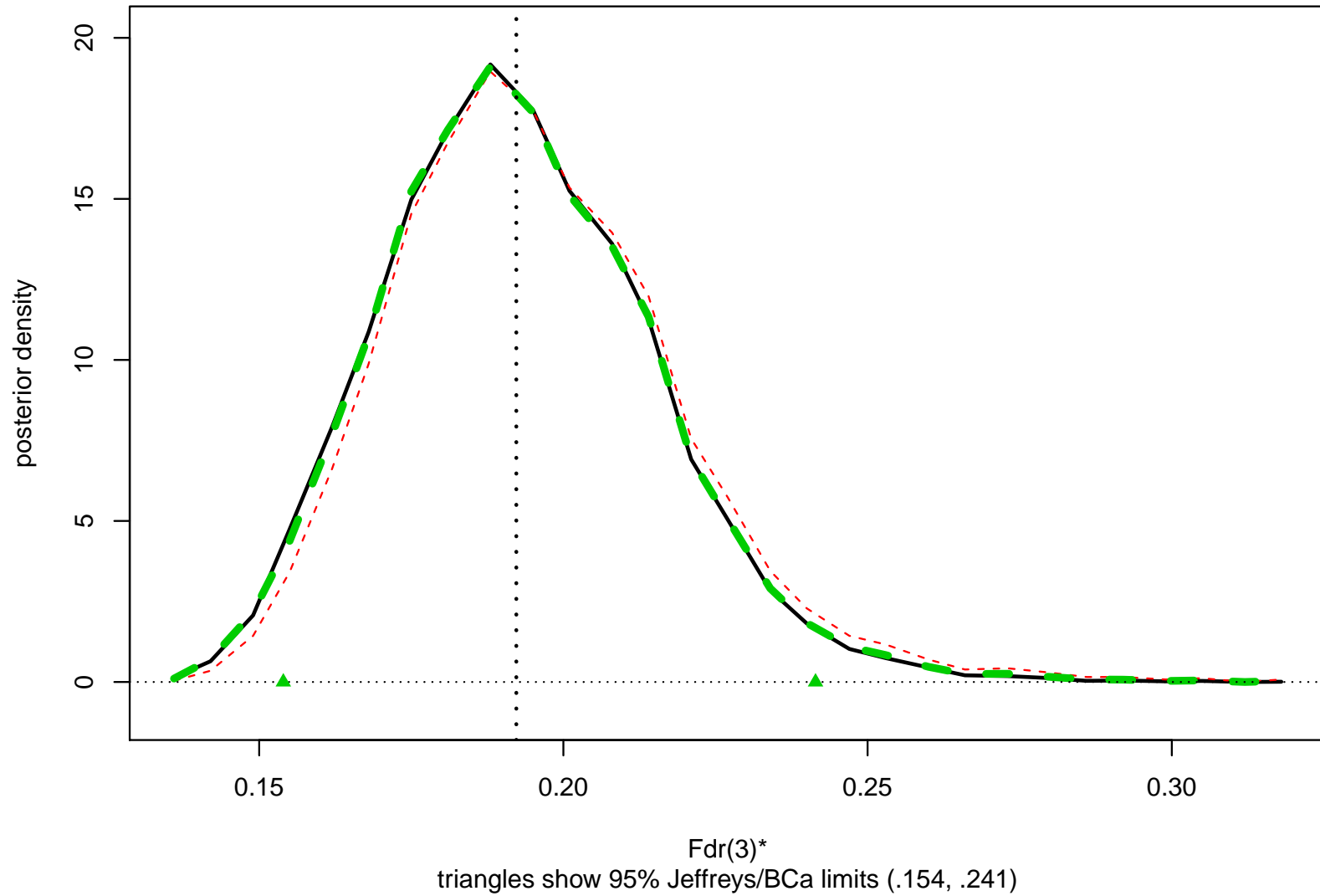
Compare Model 4 (line) with Model 8 (solid);
4000 parametric bootstrap replications of Fdr(3)

Fdr(3) bootreps
triangles are 95% bca limits: (.141,.239) vs (.154,.241)

Model Selection: AIC = Deviance + 2 · df

Model	AIC	boot counts	Bayes counts[†]	
M2	142.6	0	0.0	
M3	143.1	0	0.0	
M4	**73.3**	1266	1458	(36%)
M5	74.3	415	462	(12%)
M6	75.8	215	197	(5%)
M7	77.8	54	67	(2%)
M8	75.6	2050	1816	(45%)

[†] Model 8, Jeffreys prior

Internal Accuracy of Bayes Estimates

- B for computing $\hat{E}\left\{t(\beta)|\hat{\beta}\right\} = \left.\sum_1^B t_i\pi_i R_i\middle/ \sum_1^B \pi_i R_i\right.$?

- Define $r_i = \pi_i R_i$ and $s_i = t_i\pi_i R_i$ • $\hat{c}_{ss} = \sum_1^B (s_i - \bar{s})^2/B$, etc.

$$\mathrm{CV}\left\{\hat{E}\right\}^2 = \frac{1}{B}\left(\frac{\hat{c}_{ss}}{\bar{s}^2} - 2\frac{\hat{c}_{sr}}{\bar{s}\bar{r}} + \frac{\hat{c}_{rr}}{\bar{r}^2}\right)$$

- *Example* $t = \mathrm{Fdr}(3)$

   4000 parametric bootreps, Jeffreys Bayes

- Model 4 $\hat{E} = 0.193$ $\widehat{\mathrm{CV}} = \mathbf{0.0019}$

- Model 8 $\hat{E} = 0.179$ $\widehat{\mathrm{CV}} = \mathbf{0.0025}$

Sampling Variability of Bayes Estimates

- *How much* would $\hat{E}\left\{t(\beta)|\hat{\beta}\right\}$ vary for new $\hat{\beta}$'s?
  (i.e., frequentist properties of Bayes estimates)

- Need to bootstrap the parametric bootstrap calculations
  for $\hat{E}\left\{t(\beta)|\hat{\beta}\right\}$

- Shortcut   "Bootstrap after bootstrap"

Bootstrap-after-Bootstrap Calculations

- $f_{\hat{\beta}}(\cdot) \longrightarrow \beta_1, \beta_2, \ldots, \beta_i, \ldots, \beta_B$,

  the original bootstrap reps under $\hat{\beta}$

- Want bootreps under some other value $\hat{\gamma}$

- Idea   Reweight the originals

- Weight $Q_{\hat{\gamma}}(\beta) = f_\beta(\hat{\gamma}) / f_\beta(\hat{\beta})$:

$$\hat{E}\{t|\hat{\gamma}\} = \sum_1^B t_i \pi_i R_i Q_{\hat{\gamma}}(\beta_i) \Bigg/ \sum_1^B \pi_i R_i Q_{\hat{\gamma}}(\beta_i)$$

Bootstrap-after-Bootstrap Standard Errors

- BAB

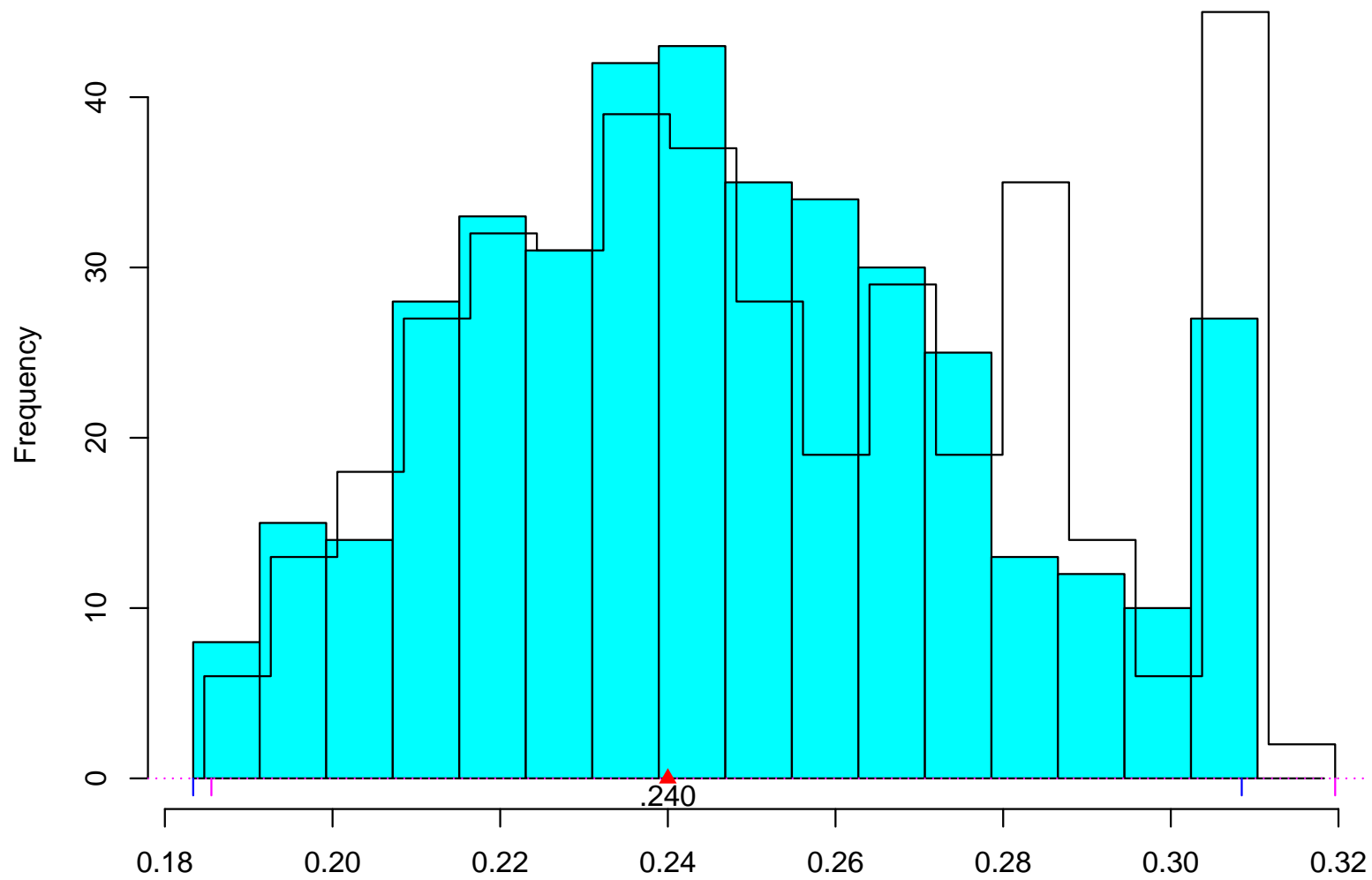 (1) $$f_{\hat{\beta}}(\cdot) \rightarrow \gamma_1, \gamma_2, \ldots, \gamma_k$$

  (2) $$\hat{E}_k = \hat{E}\{t|\hat{\gamma}_k\}$$

  (3) $$\text{se}\{\hat{E}\} = \left[\sum \left(\hat{E}_k - \hat{E}_.\right)^2 \Big/ K\right]^{1/2}$$

- *Example* Model 4

  97.5% credible limit for Fdr(3) = 0.241± ??

400 BAB replications of 97.5% credible upper limit for Fdr(3);
line histogram: same for BCa upper limit

.240

97.5% credible upper limit
BAB Standard Errors: .030 (Bayes)   .034 (BCa);   corr .98

Bayesian Inference

BAB SDs for Bayes Model Selection Proportions

| | Bayes | |
Model	post. prob.	BAB SD
M4	36%	± 20%
M5	12%	± 14%
M6	5%	± 8%
M7	2%	± 6%
M8	45%	± 27%

- *Model averaging?*

A Few Conclusions . . .

- Parametric bootstrap closely related to objective Bayes. (That's why it's a good importance sampling choice.)

- When it applies, parboot approach has both computational and interpretational advantages over MCMC/Gibbs.

- Objective Bayes analyses should be checked frequentistically.

References

- **Bootstrap** DiCiccio & Efron 1996 *Statist. Sci.* 189–228

- **Bayes and Bootstrap** Newton & Raftery 1994 *JRSS-B* 3–48 (see Discussion!)

- **Objective Priors** Berger & Bernardo 1989 *JASA* 200–207; Ghosh 2011 *Statist. Sci.* 187–202

- **Exponential Families** Efron 2010 *Large-Scale Inference*, Appendix 1

- **MCMC and Bootstrap** Efron 2011 *J. Biopharm. Statist.* 1052–1062