A 250-Year Argument

Belief, Behavior, and the Bootstrap

Bradley Efron

Stanford University

Bayes Rule (1763)

- 250th anniversary in 2013
- Always influential, usually controversial

$$\boxed{\pi(\theta|\boldsymbol{x}) = \pi(\theta)f(\boldsymbol{x}|\theta)/c(\boldsymbol{x})} \quad \begin{cases} \boldsymbol{x} = \text{ observed data} \\ \pi(\theta) = \text{ prior belief for parameter } \theta \\ f(\boldsymbol{x}|\theta) = \text{ likelihood} \\ c(\boldsymbol{x}) = \text{ scaling constant} \\ \pi(\theta|\boldsymbol{x}) = \text{ posterior belief} \end{cases}$$

- How to choose $\pi(\theta)$?

The Physicist's Twins

- Sonogram "twin boys"
- Prob. *identical*, not *fraternal*?
- Doctor "population probability *identical* = 1/3"
- Bayes Rule $$\frac{\pi(\text{ident}\,|\,\text{sono})}{\pi(\text{frat}\,|\,\text{sono})} = \left[\frac{\pi(\text{ident})}{\pi(\text{frat})}\right]\left[\frac{f(\text{sono}\,|\,\text{ident})}{f(\text{sono}\,|\,\text{frat})}\right]$$
- Prior Odds $= \dfrac{1/3}{2/3} = \dfrac{1}{2}$
- Likelihood Ratio $= \dfrac{1}{1/2} = 2$
- Posterior Odds $= \left[\dfrac{1}{2}\right] \cdot [2] = 1$
- $\pi(\text{ident}\,|\,\text{sono}) = 50\%$

The Complete Population of Twins

		Sexes		
		Same	*Different*	
Twins	*Identical*	**1/3**	**0**	1/3
	Fraternal	**1/3**	**1/3**	2/3
		sonogram		

Doctor

The Doctor's Beliefs

- Bayes Prior $\pi(\theta)$: Infinite list of relevant past cases (millions of prior twins)
- What if only 3 cases, one *identical* and two *fraternal*?
- **Harold Jeffreys** (1930s): theory of uninformative priors
- Dominates current Bayesian practice
- *Statistician can "always be a Bayesian"*

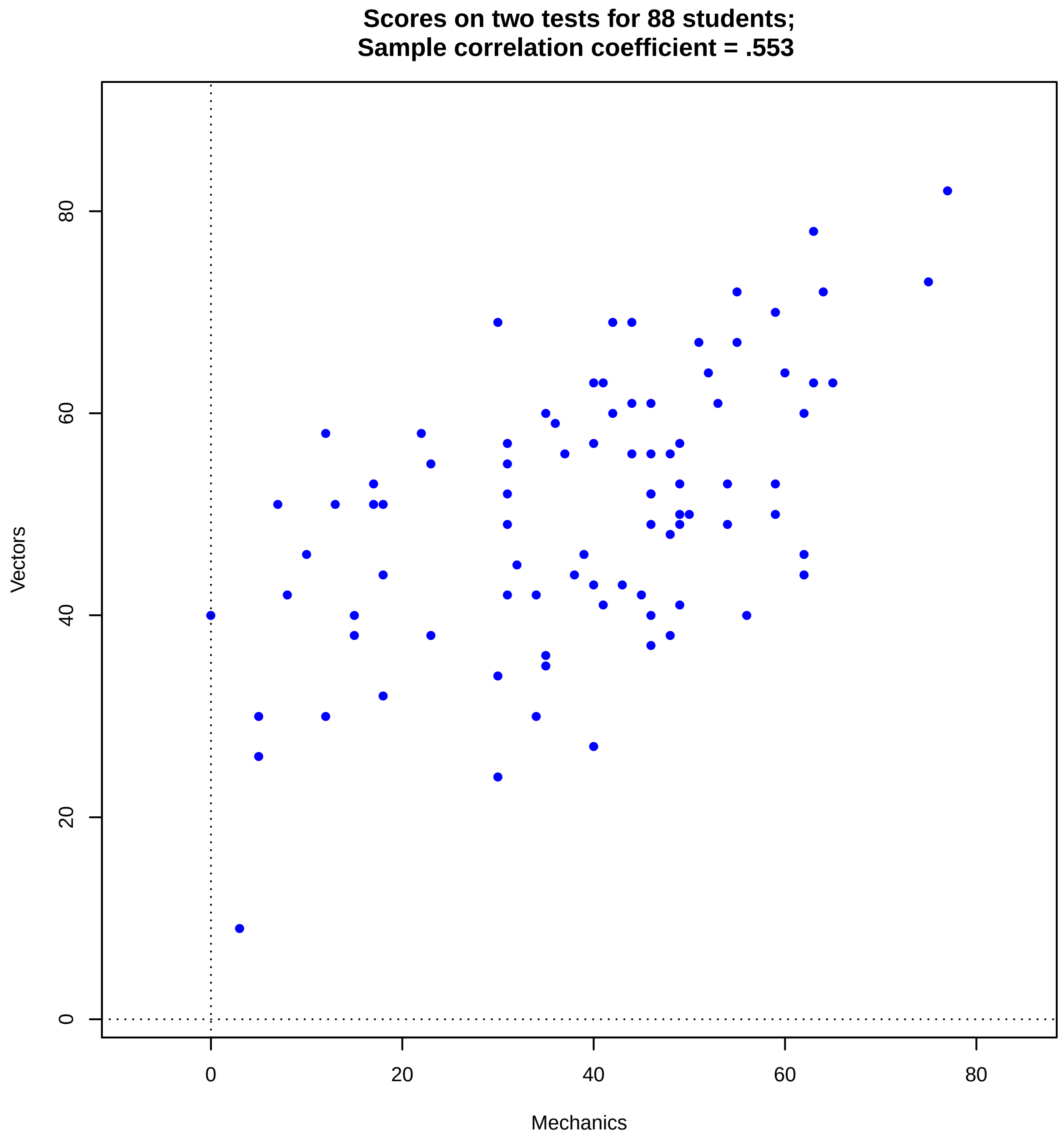
Jeffreys and the Twins

- θ = population proportion of *identical* twins
  [Doctor: "$\theta = 1/3$"]
- Uninformative hyperprior $\pi(\theta) \propto \theta^{-1}(1-\theta)^{-1}$
  for θ in (0,1)
- *Doctor should believe*: $\pi(\theta|\text{his experience}) = 2 \cdot (1-\theta)$
  ["Beta(1,2)"]
- Then $\boxed{\pi(\text{ident}\,|\,\text{sono}, \text{Doc's experience}) = 50\%}$
- Same "50%"?

Frequentist Inference ("Behavior")

- No prior beliefs "$\pi(\theta)$"
- Specific estimation rule $\hat{\theta} = s(\boldsymbol{x})$

  [test, prediction, model selection, etc.]
- Judge $s(\boldsymbol{x})$ by its behavior in repeated use
  - Bayes   infinite past experience
  - Frequentist   infinite future usage


Scores on two tests for 88 students;
Sample correlation coefficient = .553
Vectors
0
20
40
60
80
Mechanics

Correlation Example (Mardia, Kent, and Bobby)

- Data $n = 88$ students:

  Scores on two exams "mechanics" and "vectors"

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_{88}) \quad \text{with} \quad x_i = (x_{i1}, x_{i2})$$

- *Parameter of interest*: correlation coefficient $\rho = \text{cor}(\text{mec}, \text{vec})$
- *Estimate* $\hat{\rho} = s(\boldsymbol{x})$
- Sample correlation coefficient $= 0.553 \pm ?$
- Bivariate normal assumption $x_i \overset{\text{ind}}{\sim} \mathcal{N}_2(\mu, \Sigma)$

  for $i = 1, 2, \ldots, 88$

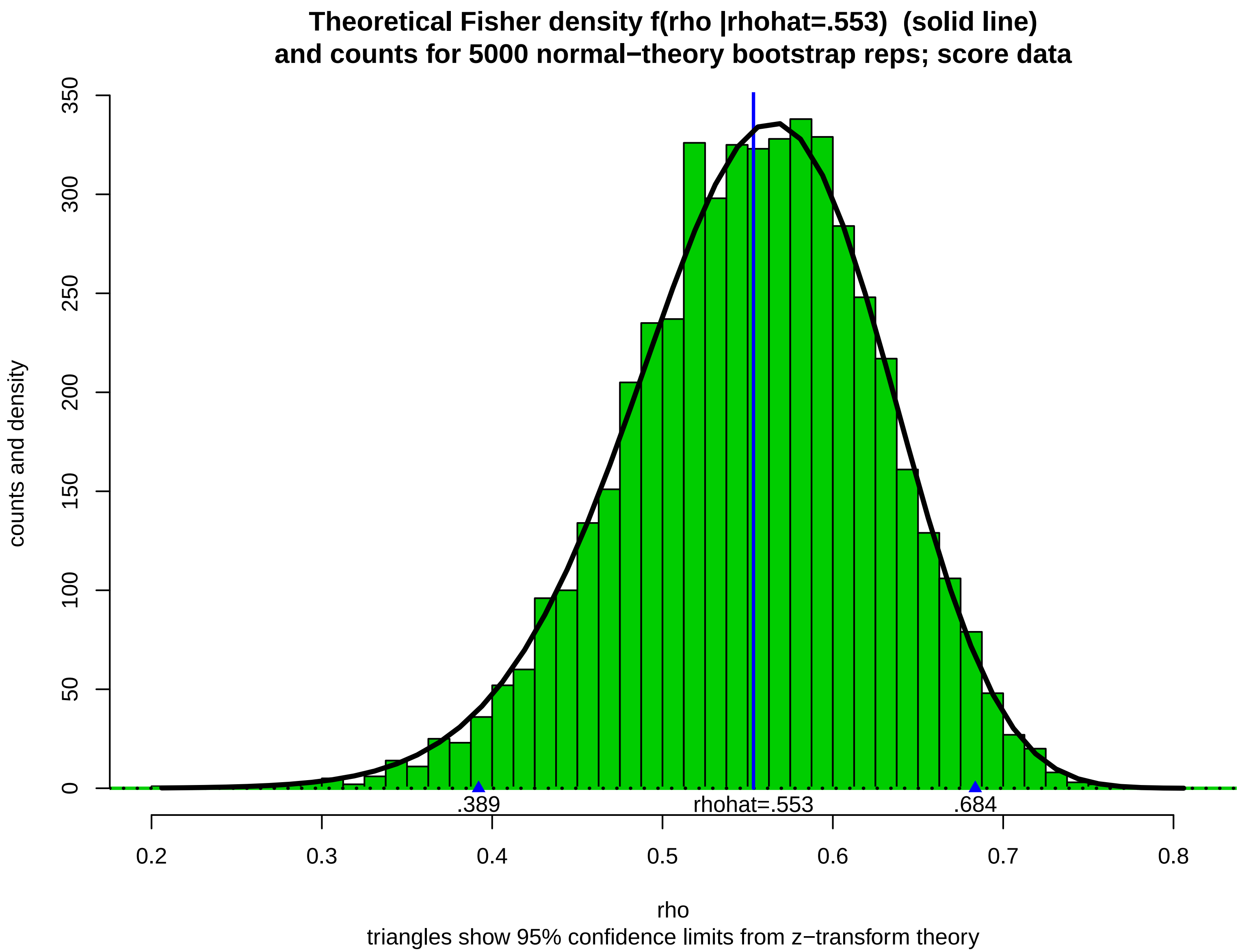
Fisher: Normal Theory for $\hat{\rho}$ (1915)

- *Density* $f_\rho(\hat{\rho})$
- $\mathrm{sd}_\rho(\hat{\rho}) \doteq (1-\rho^2)\Big/\sqrt{n-3}$ [= 0.075 for $\rho = 0.553$]
- Transformation $\hat{\tau} = \frac{1}{2}\log\frac{1+\hat{\rho}}{1-\hat{\rho}}$ • $\hat{\tau} \dot{\sim} \mathcal{N}\left(\tau, \frac{1}{n-3}\right)$
- 95% confidence interval for ρ: transform $\hat{\tau} \pm 1.96\frac{1}{\sqrt{n-3}}$

  back to ρ scale:

  $$\boxed{\rho \in [0.389, 0.684]}$$

  - Bayes single probability distribution
  - Frequentist many distributions


Theoretical Fisher density f(rho |rhohat=.553) (solid line)
and counts for 5000 normal−theory bootstrap reps; score data
counts and density
0
50
100
150
200
250
300
350
.389
rhohat=.553
.684
0.2
0.3
0.4
0.5
0.6
0.7
0.8
rho
triangles show 95% confidence limits from z−transform theory

Parametric Bootstrap Calculations

- Assume $x_i \overset{\text{ind}}{\sim} \mathcal{N}_2(\mu, \Sigma)$ • Estimate $\hat{\mu}$ and $\hat{\Sigma}$ by MLE
  - *Bootstrap sample* $x_i^* \overset{\text{ind}}{\sim} \mathcal{N}_2\left(\hat{\mu}, \hat{\Sigma}\right)$
    for $i = 1, 2, \ldots, \; n = 88$
  - *Bootstrap replication* $\hat{\rho}^*$ = cor coef for $\boldsymbol{x}^* = (x_1^*, \ldots, x_{88}^*)$
- $B = 5000$ replications gave $\widehat{\mathrm{sd}} = 0.074$
- Bootstrap 95% confidence interval $[0.393, 0.682]$ ("BCa")

Bayes-Frequentist Conversion Factor

- Bayes $E\left\{t(\rho)|\hat{\rho}\right\} = \dfrac{\int t(\rho)\pi(\rho)f_\rho(\hat{\rho})\,d\rho}{\int \pi(\rho)f_\rho(\hat{\rho})\,d\rho} = \dfrac{\int t(\rho)\pi(\rho)R(\rho)f_{\hat{\rho}}(\rho)\,d\rho}{\int \pi(\rho)f_{\hat{\rho}}(\rho)\,d\rho}$

  where $R(\rho) = \dfrac{f_\rho(\hat{\rho})}{f_{\hat{\rho}}(\rho)}$ is "conversion factor."

- *Simulation* Parametric bootstrap $f_{\hat{\rho}}(\cdot) \longrightarrow B$ bootreps $\rho_1, \rho_2, \ldots, \rho_B$ $\left[\rho_i = \hat{\rho}_i^*\right]$

- $$\hat{E}\left\{t(\rho)|\hat{\rho}\right\} = \sum_{i=1}^{B} t_i\pi_i R_i \Big/ \sum \pi_i R_i$$

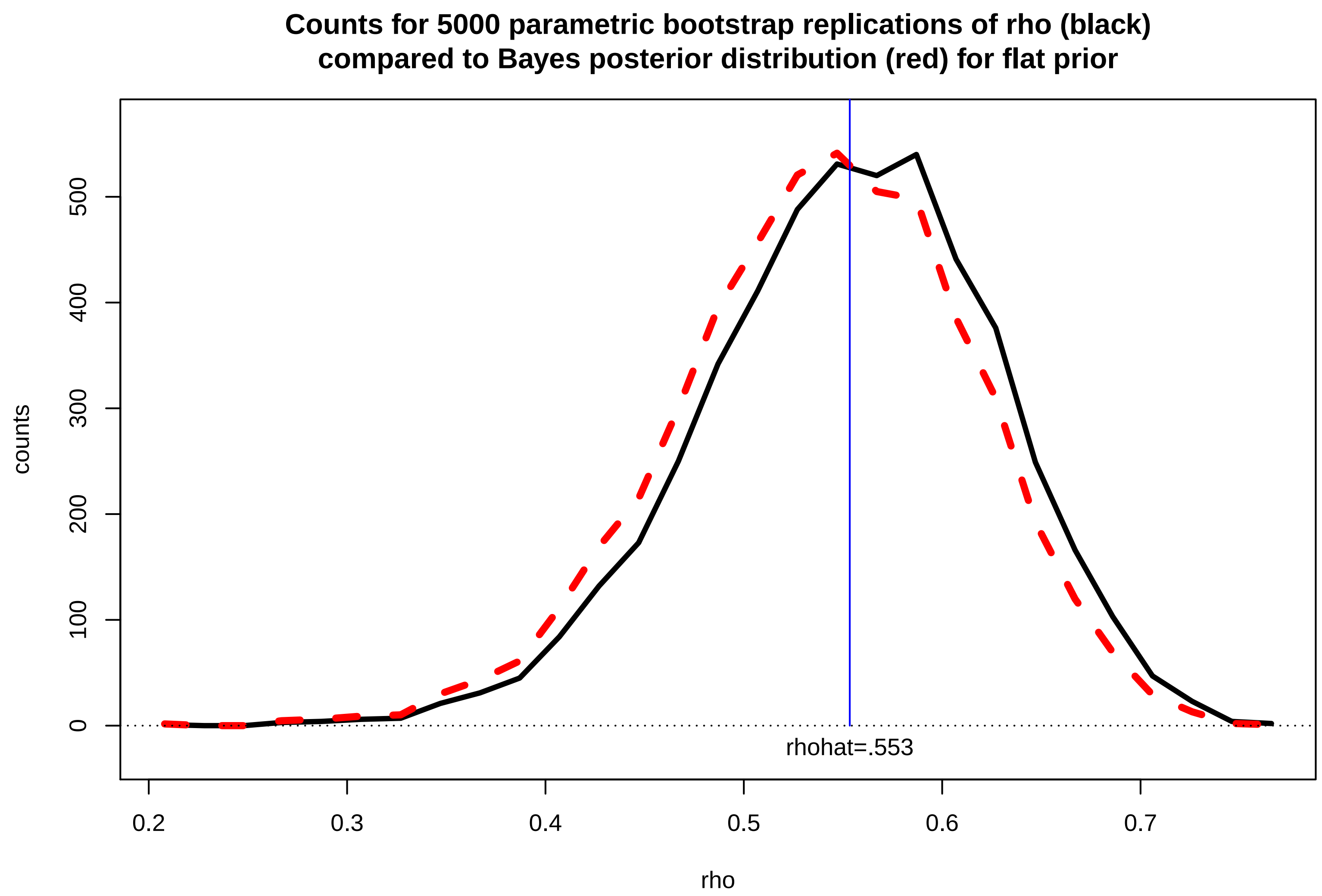
  where $t_i = t(\rho_i)$, $\pi_i = \pi(\rho_i)$, $R_i = R(\rho_i)$.

- Newton–Raftery *JRSS-B* (1994)

Bayes Posterior Density for Flat Prior: $\pi(\rho) = 1$

- Bin the $B = 5000$ bootstrap replications, as for histogram
- y_k = count in kth bin = $\sum_1^B I_k(\rho_i)$

  [$I_k(\cdot)$ indicator for ρ in kth bin]
- Bayes counts $$\tilde{y}_k = \sum_1^B R(\rho_i) I_k(\rho_i) \Big/ \bar{R}$$
- Plot y_k and $\tilde{y}_k$ versus bin centers to see Bayes-frequentist differences (i.e., usual bootstrap histogram compared to Bayes posterior dist.)


Counts for 5000 parametric bootstrap replications of rho (black)
compared to Bayes posterior distribution (red) for flat prior
counts
0
100
200
300
400
500
0.2
0.3
0.4
0.5
0.6
0.7
rho
rhohat=.553

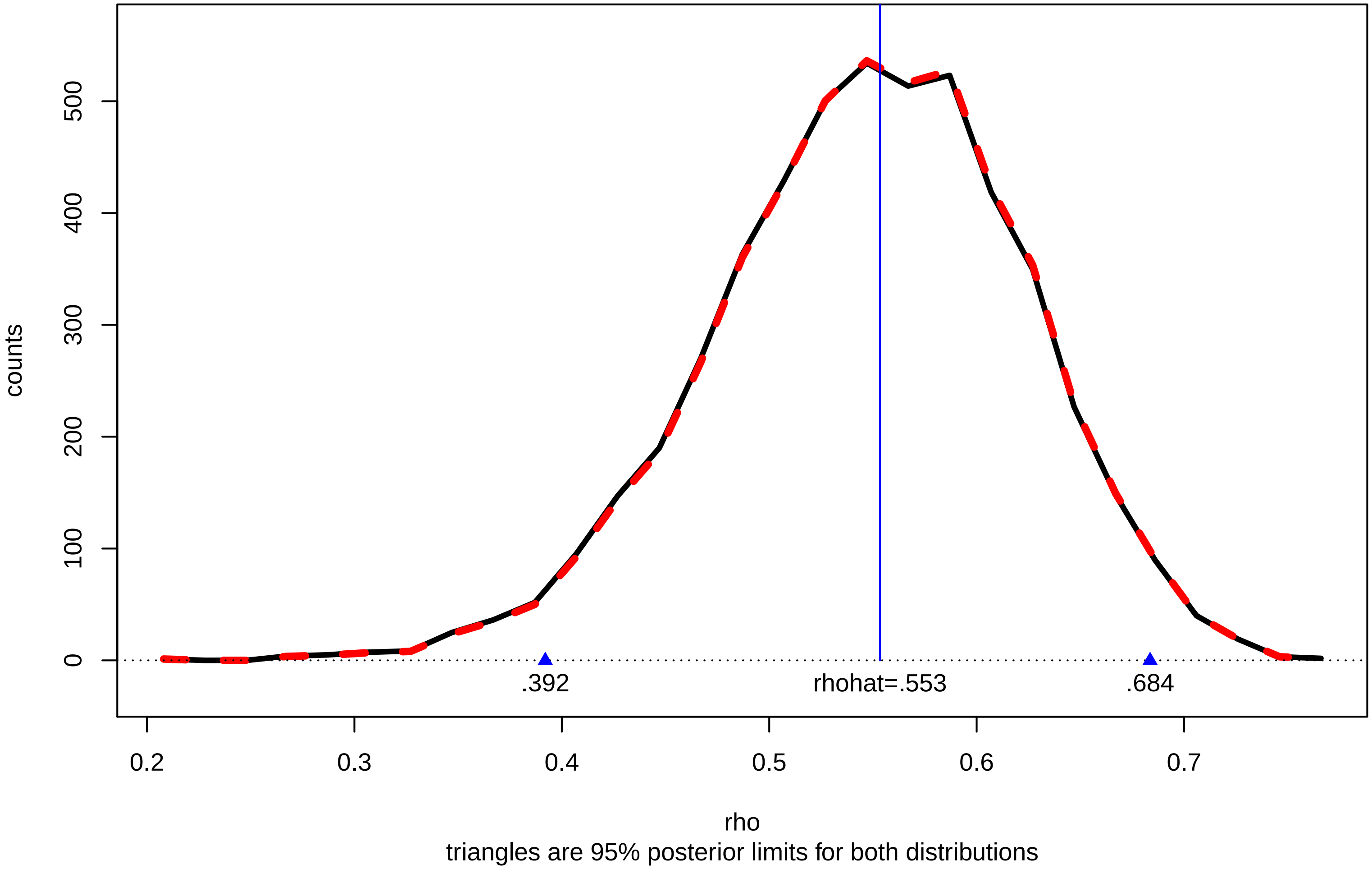
Jeffreys Prior for ρ

- $\pi(\rho) = 1/(1-\rho^2)$ "uninformative for ρ" [makes posterior intervals have nearly correct frequentist coverage]
- BCa method frequentist weights w_i that reweight raw bootstrap counts to have nearly correct coverages:

$$\hat{E}_{\text{BCa}}\{t\} = \sum t_i w_i / w_i$$

- In this case Bayes $\doteq$ frequentist (not always so!)

Corrected BCa bootstrap distribution (black) and Jeffreys Bayes posterior density (red) for rho|rhohat



Comparison of Intervals

	.025	**.975**	
Fisher	.389	.684	(confidence intervals)
Jeffreys	.392	.684	(credible intervals)
BCa	.393	.682	(both?)

Prostate Cancer Study (Singh et al, 2002)

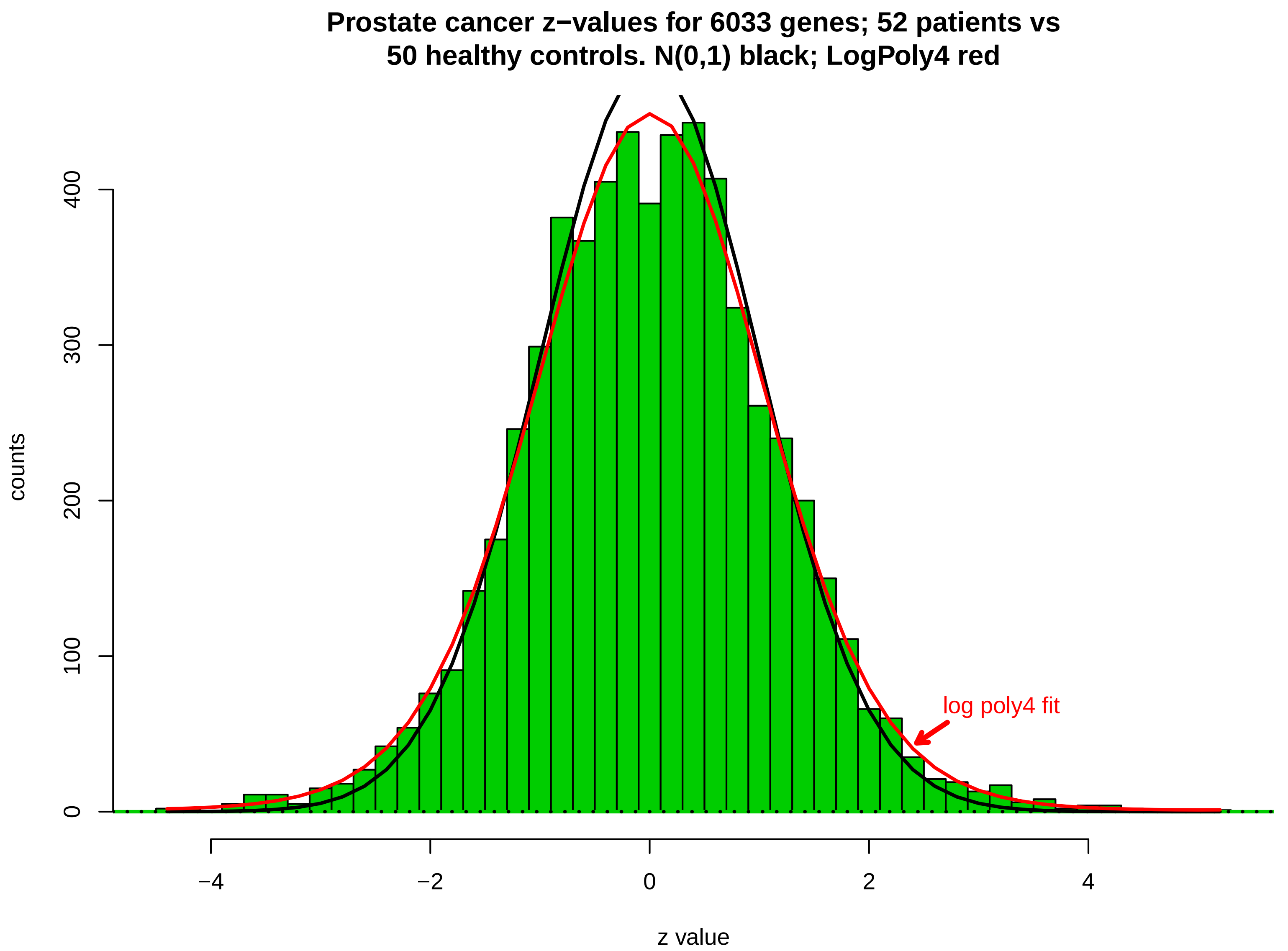
- *Microarray study*

  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
counts
0
100
200
300
400
log poly4 fit
−4
−2
0
2
4
z value

False Discovery Rate: $\mathrm{Fdr}(z_0) = \Pr\{H_{0i} | z_i \geq z_0\}$

- Let $S_0(z_0) = \Pr\{\mathcal{N}(0.1) \geq z_0\}$ and $S(z) = \Pr\{z \geq z_0\}$
- $\mathrm{Fdr}(z_0) \doteq S_0(z_0)/S(z_0)$
- Empirical Bayes $\boxed{\widehat{\mathrm{Fdr}}(z_0) = S_0(z_0) \Big/ \hat{S}_{\text{nonpar}}(z_0)}$ where

  $\hat{S}_{\text{nonpar}}(z_0) = \#\{z_i \geq z_0\}/n$ (Benjamini–Hochberg 1995)
- *Reject* H_{0i} if $\widehat{\mathrm{Fdr}}(z_i)$ small (≤ 0.1 or 0.2)

Poisson Model

- Let $y_1, y_2, \ldots, y_K$ be heights of histogram bars (i.e., bin counts), $\mu_1, \mu_2, \ldots, \mu_K$ their unknown expectations.
- Poisson model $\quad y_k \overset{\text{ind}}{\sim} \text{Poi}(\mu_k)$
- $\log(\mu)$ polynomial function of z
- GLM gives parametric cdf estimate $\hat{S}_{\text{par}}(z_0)$

$$\boxed{\widehat{\text{Fdr}}_{\text{par}}(z_0) = S_0(z_0) \Big/ \hat{S}_{\text{par}}(z_0) = \hat{\theta}}$$

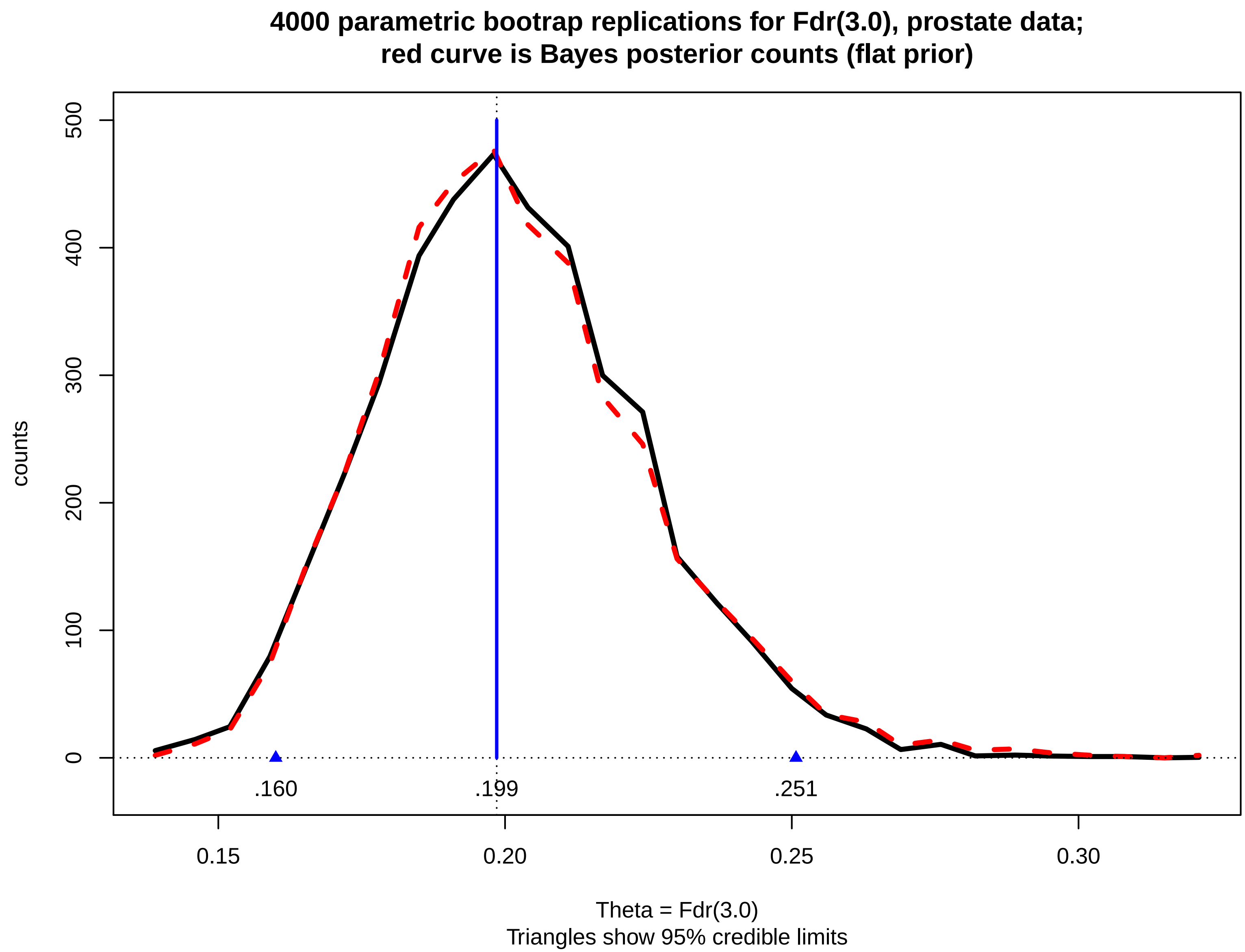
- *Example* $\quad$ polynomial degree 4: $\hat{\theta} = \widehat{\text{Fdr}}_{\text{par}}(3.0) = 0.199 \pm ?$

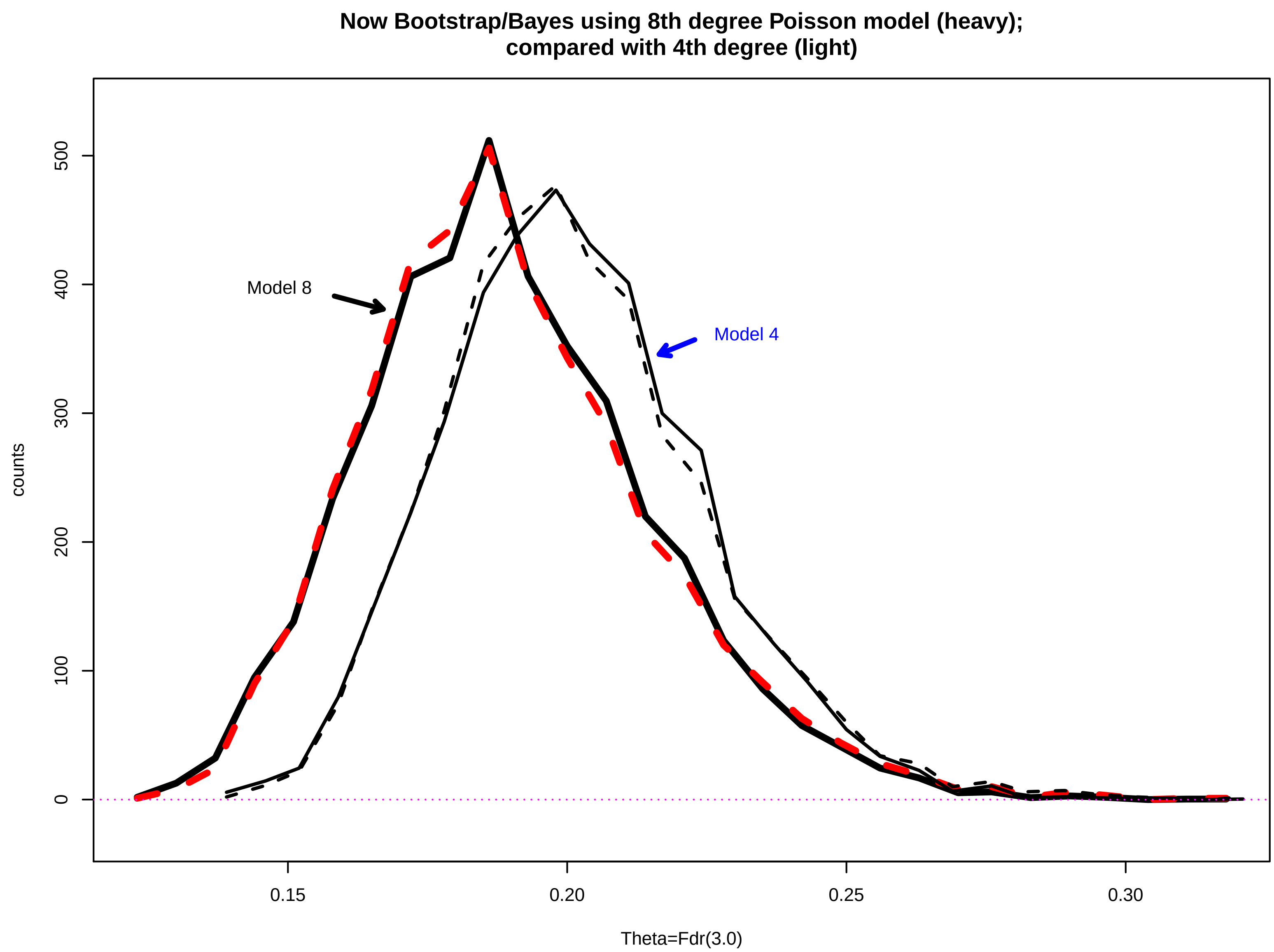
Parametric Bootstrap Replications

- Use estimates $\hat{\mu}_k$ from Poisson GLM
- Bootstrap sample $y_k^* \overset{\text{ind}}{\sim} \text{Poi}\,(\hat{\mu}_k)$ gives $\hat{S}^*_{\text{par}}(z_0)$ and

$$\boxed{\hat{\theta}^* = S_0(z_0) \Big/ \hat{S}^*_{\text{par}}(z_0)}$$

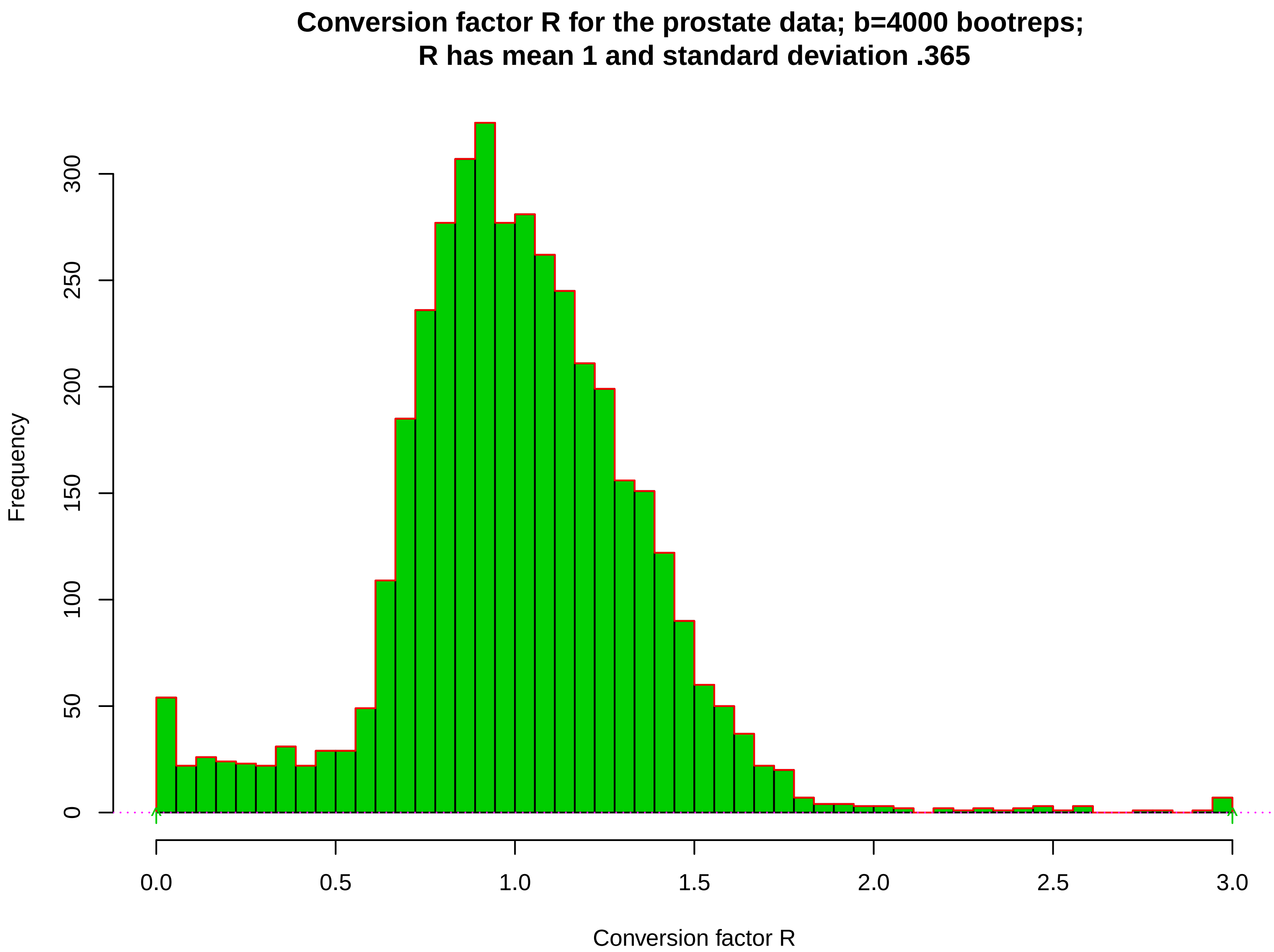
- $B = 4000$ bootstrap replications gave boot standard error 0.024 for $z_0 = 3$


4000 parametric bootrap replications for Fdr(3.0), prostate data;
red curve is Bayes posterior counts (flat prior)
counts
0
100
200
300
400
500
.160
.199
.251
0.15
0.20
0.25
0.30
Theta = Fdr(3.0)
Triangles show 95% credible limits


Now Bootstrap/Bayes using 8th degree Poisson model (heavy);
compared with 4th degree (light)
Model 8
Model 4
counts
0
100
200
300
400
500
0.15
0.20
0.25
0.30
Theta=Fdr(3.0)

Model Selection: AIC = Deviance + 2 · df

Model	deviance	AIC	boot counts	Bayes counts
M2	138.6	142.6	0	0.0
M3	137.1	143.1	0	0.0
M4	65.3	73.3*	1265	1294.8
M5	64.3	74.3	416	421.0
M6	63.8	75.8	215	214.6
M7	63.8	77.8	54	54.7
M8	59.6	75.6	2050	2033.4


Conversion factor R for the prostate data; b=4000 bootreps;
R has mean 1 and standard deviation .365
Frequency
0
50
100
150
200
250
300
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Conversion factor R

The Conversion Factor R

- Converts frequentist boot density function $f_{\hat{\beta}}\left(\hat{\beta}^*\right)$ to likelihood $f_{\beta}\left(\hat{\beta}\right)$
- $\hat{t} = \sum_1^B t_i R_i \Big/ \sum_1^B R_i$ and $\bar{t} = \sum_1^B t_i / B$, $\overline{\text{sd}}$ = boot st dev of t:

$$\frac{\hat{t} - \bar{t}}{\overline{\text{sd}}} = \overline{\text{cor}}(t, R) \cdot \overline{\text{CV}}(R)$$

- *Prostate example*

  $\overline{\text{CV}} = 0.365$ $\overline{\text{cor}} = -0.01$ so $\dfrac{\hat{t} - \bar{t}}{\overline{\text{sd}}} = -0.004$

- *Correlation example*

  $\overline{\text{CV}} = 0.172$ $\overline{\text{cor}} = -0.999$ so " " $= -0.17$

One-Parameter Exponential Families

- θ the expectation parameter
  (e.g., p in binomial, λ in Poisson)

- Repeated sampling

$$\hat{\theta} \sim \left(\theta, \sigma_\theta \Big/ \sqrt{n}, \gamma_\theta \Big/ \sqrt{n}\right) \quad \text{(mean, sd, skewness)}$$

- $Z \equiv \sqrt{n}\,\dfrac{\theta - \hat{\theta}}{\sigma_{\hat{\theta}}} \longrightarrow \mathcal{N}(0,1)$

$$\boxed{\log R(\theta) \doteq \frac{\gamma_{\hat{\theta}}}{2\sqrt{n}} Z \cdot (1 + Z^2/3)}$$

- Rate Bayes $\longrightarrow$ frequentist depends on skewness $\gamma_{\hat{\theta}}$

Some Conclusions

- Parametric bootstrap connects Bayes and frequentist analyses, especially Jeffreys-type uninformative Bayes.
  [bootstrap: only one probability distribution]
- Convenient way to compute $\pi(\theta|\boldsymbol{x})$.
  [competition to MCMC]
- Provides easy estimates of simulation error.
- "Behavior" important for Jeffreys-Bayes methods.