

Frequentist Accuracy of Bayesian Estimates

Bradley Efron

Stanford University

Bayesian Inference

- Parameter: $\mu \in \Omega$
- Observed data: x
- Prior: $g(\mu)$
- Probability distributions: $\{f_\mu(x), \mu \in \Omega\}$
- Parameter of interest: $\theta = t(\mu)$

$$E\{\theta|x\} = \int_{\Omega} t(\mu) f_\mu(x) g(\mu) d\mu \Big/ \int_{\Omega} f_\mu(x) g(\mu) d\mu$$

- What if we don't know g ?

Jeffreysonian Bayes Inference

“Uninformative Priors”

- Jeffreys: $g(\mu) = |\mathcal{I}(\mu)|^{1/2}$ where $\mathcal{I}(\mu) = \text{cov} \{ \nabla_{\mu} \log f_{\mu}(x) \}$
(the Fisher information matrix)
- Can still use Bayes theorem but how accurate are the estimates?
- Today: frequentist variability of $E \{ t(\mu) | x \}$

General Accuracy Formula

- μ and $x \in \mathcal{R}^p$
- $x \sim (\mu, V_\mu)$
- $\alpha_x(\mu) = \nabla_x \log f_\mu(x) = \left(\dots, \frac{\partial \log f_\mu(x)}{\partial x_i}, \dots \right)^T$

Lemma

$E = E\{t(\mu)|x\}$ has gradient $\nabla_x E = \text{cov}\{t(\mu), \alpha_x(\mu)|x\}$.

Theorem

The delta-method standard deviation of E is

$$\text{sd}(E) = \left[\text{cov}\{t(\mu), \alpha_x(\mu)|x\}^T V_x \text{cov}\{t(\mu), \alpha_x(\mu)|x\} \right]^{1/2}.$$

Implementation

- **Posterior Sample** $\mu_1^*, \mu_2^*, \dots, \mu_B^*$ (MCMC)
- $\hat{\theta}_i^* = t(\mu_i^*) = t_i^*$ and $\alpha_i^* = \alpha_x(\mu_i^*)$

$$\widehat{\text{cov}} = \sum_{i=1}^B (\alpha_i^* - \bar{\alpha})(t_i^* - \bar{t}) / B$$

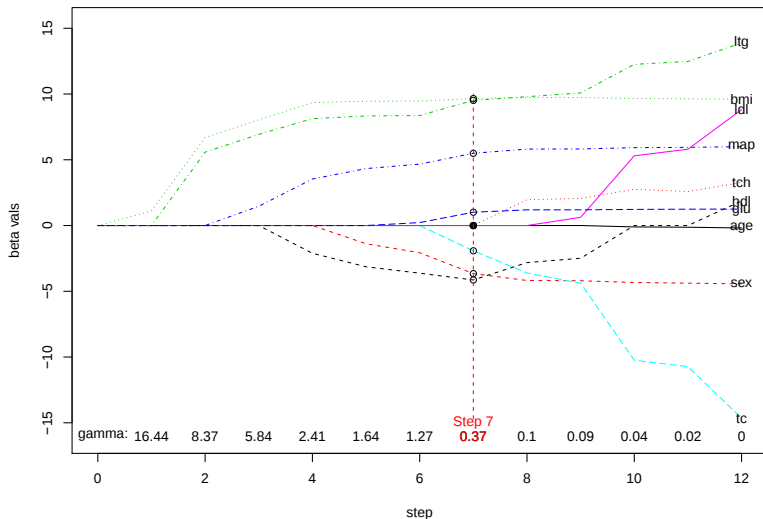
- $\widehat{\text{sd}}(E) = \left[\widehat{\text{cov}}^T V_x \widehat{\text{cov}} \right]^{1/2}$

Diabetes Data

Efron et al. (2004), "LARS"

- $n = 442$ subjects
- $p = 10$ predictors: age, sex, bmi, glu, . . .
- Response: y = disease progression at one year
- Model:
$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times p}{\mathbf{X}} \underset{p \times 1}{\beta} + \underset{n \times 1}{\mathbf{e}} \quad [\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, I)]$$

Diabetes data. Lasso: $\min\{\text{RSS}/2 + \gamma \cdot L1(\beta)\}$
 C_p minimized at Step 7, with $\gamma = .37$



Bayesian Lasso

Park and Casella (2008)

- Model: $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\beta, I)$ “ μ ” = β and “ $\mathbf{x}$ ” = $\mathbf{y}$
- Prior: $g(\beta) = e^{-\gamma L_1(\beta)}$ [$\gamma = 0.37$]
- Then posterior mode at Lasso $\hat{\beta}_\gamma$
- Subject 125: $\theta_{125} = \mathbf{x}_{125}^T \beta$
- How accurate are Bayes posterior inferences for θ_{125} ?

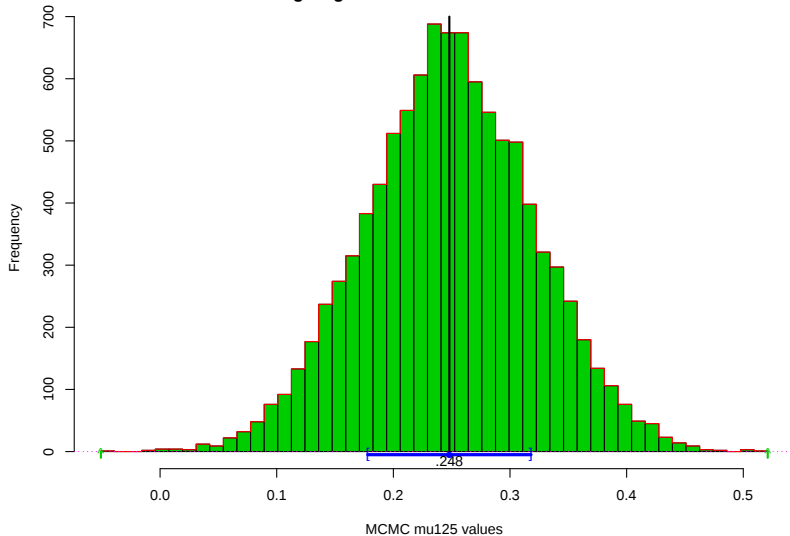
Bayesian Analysis

- MCMC: posterior sample $\{\beta_i^* \text{ for } i = 1, 2, \dots, 10,000\}$
- Gives $\{\theta_{125,i}^* = \mathbf{x}_{125}^T \beta_i^*, i = 1, 2, \dots, 10,000\}$

$$\theta_{125,i}^* \sim 0.248 \pm 0.072$$

- General accuracy formula: frequentist sd 0.071 for $E = 0.248$
- Other subjects freq sd < Bayes sd

Figure 2. 10000 MCMC values for Theta =Subject 125 estimate;
mean=.248, stdev=.0721; Frequentist Sd for $E\{\theta|data\}$ is .0708,
giving Coefficient of Variation 29%



Posterior CDF of θ_{125}

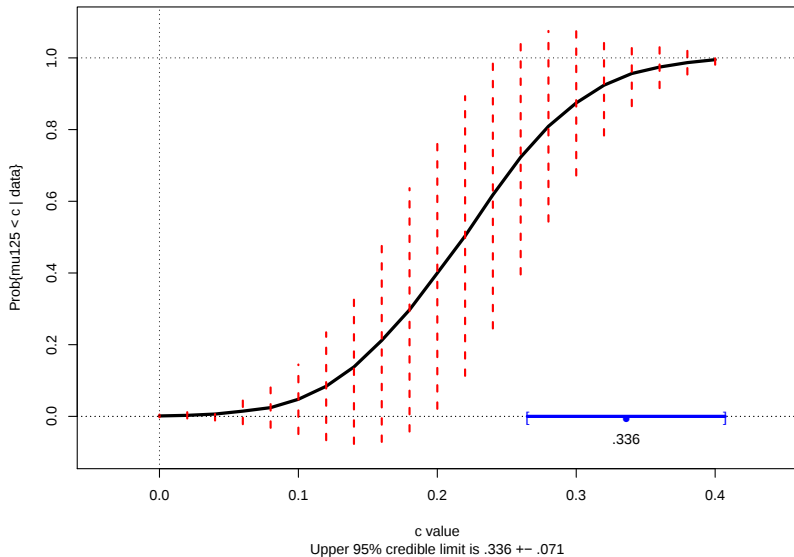
- Apply GAC to $s_i^* = I\{t_i^* < c\}$ so

$$E\{s|\text{data}\} = \Pr\{\theta_{125} \leq c|\text{data}\}$$

$$c = 0.3 : \hat{E} = 0.762 \pm 0.304$$

↗ ↗
Bayes GAC sd
estimate

Figure 3. MCMC posterior cdf of μ_{125} , Diabetes data,
Prior $\exp\{-.37 \cdot L_1(\beta)\}$; verts are $\pm$ One Frequentist Standard Dev



Exponential Families

- $f_{\alpha}(\hat{\beta}) = e^{\alpha^T \hat{\beta} - \psi(\alpha)} f_0(\hat{\beta})$ “ x ” = $\hat{\beta}$ and “ μ ” = $\beta = E_{\alpha} \{\hat{\beta}\}$
- **GAC** $\alpha = \alpha_x(\mu)$
- For $\hat{E} = E \{t(\beta) | \hat{\beta}\}$

$$\widehat{\text{sd}} = \left[\text{cov}(t, \alpha | \hat{\beta})^T V_{\hat{\alpha}} \text{cov}(t, \alpha | \hat{\beta}) \right]^{1/2}$$

with $V_{\hat{\alpha}} = \text{cov}_{\alpha=\hat{\alpha}} \{\hat{\beta}\} = \ddot{\psi}(\hat{\alpha})$.

Bayesian Estimation Using the Parametric Bootstrap

Efron (2012)

- Parametric bootstrap: $f_{\hat{\alpha}}(\cdot) \rightarrow \{\hat{\beta}_1^*, \hat{\beta}_2^*, \dots, \hat{\beta}_B^*\}$
- Want to calculate: $\hat{E} = E\{t(\beta)|\hat{\beta}\}$ for prior $\pi(\beta)$
- Importance sampling:

$$\hat{E} \doteq \sum_1^B t_i \pi_i R_i \Bigg/ \sum_1^B \pi_i R_i$$

$t_i = t(\hat{\beta}_i^*)$, $\pi_i = \pi(\hat{\beta}_i^*)$, and $R_i =$ “conversion factor”

$$R_i = f_{\hat{\alpha}_i^*}(\hat{\beta}) / f_{\hat{\alpha}}(\hat{\beta}_i^*)$$

(Easy in exponential families.)

Bootstrap for GAC

- $p_i = \pi_i R_i / \sum_1^B \pi_j R_j$ is weight on i th Bootstrap Replication
- $\hat{E} = \sum_{i=1}^B p_i t_i^*$
- $\widehat{\text{cov}} = \sum_{i=1}^B p_i \hat{\alpha}_i^* (t_i^* - \hat{E})$ estimates $\text{cov}(\alpha, t | \hat{\beta})$

$$\widehat{\text{sd}} = \left[\widehat{\text{cov}}^T V_{\hat{\alpha}} \widehat{\text{cov}} \right]^{1/2}$$

Prostate Cancer Study

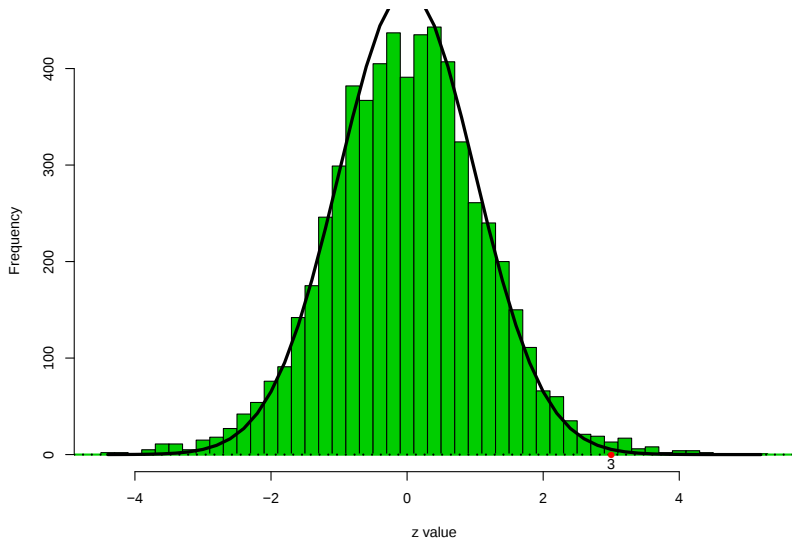
Singh et al. (2002)

- Microarray study:
102 men — 52 prostate cancer, 50 healthy controls
- 6033 genes
- z_i test statistic for H_{0i} : “no difference”

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

- Goal: identify genes involved in prostate cancer

Figure 4. Prostate study: 6033 z-values and matching $N(0,1)$ density



Poisson GLM

- Histogram: 49 bins, c_j midpoint of bin j
- $y_j = \#\{z_i \text{ in bin } j\}$
- Poisson GLM: $y_j \stackrel{\text{ind}}{\sim} \text{Poi}(\mu_j)$

$$\log(\mu) = \text{poly}(\mathbf{c}, \text{degree}=8)$$

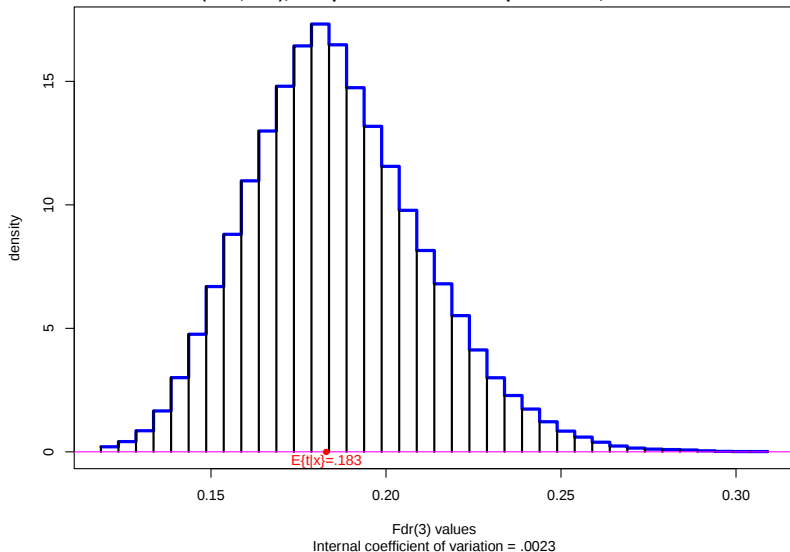
[MLE: `glm(y ~ poly(c,8), Poisson)`]

Bayesian Estimation for the Poisson Model

- Model: $\mathbf{y} \sim \text{Poi}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = e^{\mathbf{X}\boldsymbol{\beta}} \quad [\mathbf{X} = \text{poly}(\mathbf{c}, 8)]$
- Prior: Jeffreys prior for $\boldsymbol{\beta}$
- cdf: $\boldsymbol{\beta} \rightarrow \boldsymbol{\mu} \rightarrow \text{cdf: } F_{\boldsymbol{\beta}}(z) = \sum_{c_i \leq z} \mu_i \bigg/ \sum \mu_i$
- Fdr parameter: $t(\boldsymbol{\beta}) = \frac{1 - \Phi(3)}{1 - F(3)} = \text{Fdr}(3)$

$$[\text{MLE: } \text{Fdr}_{\hat{\boldsymbol{\beta}}}(3) = 0.183]$$

Figure 5. Prostate study: posterior density for Fdr(3) based on 4000 parametric bootstraps from Poisson poly(8) gave posterior (m,sd) =(.183,.025); Frequentist Sd for .183 equalled .026; CV=14%



Model Selection Calculations

- Full model: $\mathbf{y} \sim \text{Poi}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = \mathbf{e}^{\mathbf{X}\boldsymbol{\beta}} \quad \left[\boldsymbol{\beta} \in \mathcal{R}^9, \boldsymbol{\mu} \in \mathcal{R}^{49} \right]$
- Submodel $\mathcal{M}_m$: $\{\boldsymbol{\mu} : \text{only } 1\text{st } m + 1 \text{ coordinates of } \boldsymbol{\beta} \neq 0\}$
- Bayesian model selection: prior probabilities on and within each $\mathcal{M}_m$
- Poor man's model estimates: partition $\mathcal{R}^{49}$ into “preference regions” $\mathcal{R}_m = \{\boldsymbol{\mu} \text{ closest to } \mathcal{M}_m\}$
- Calculate posterior probabilities $\Pr\{\mathcal{R}_m|\mathbf{y}\}$

Posterior Model Probabilities

- Distance: minimum AIC from μ to point in $\mathcal{R}_m$
- Preferred model: is one with smallest distance

$\mathcal{R}_m$	4	5	6	7	8
posterior prob	.36	.12	.05	.02	.45
sd	.32	.16	.08	.03	.40

Empirical Bayes Accuracy

- z_k = observed value for gene k
- θ_k = “true effect size”

$$z_k \sim \mathcal{N}(\theta_k, 1)$$

- Unknown prior $g(\cdot) \rightarrow \theta_1, \theta_2, \dots, \theta_N$ [$N = 6033$]
- From observations $z_1, z_2, \dots, z_N$ we wish to estimate

$$E_z = E\{t(\theta)|z\} = \int t(\theta)f_\theta(z)g(\theta) d\theta \bigg/ \int f_\theta(z)g(\theta) d\theta$$

[if $t(\theta) = \delta_0(\theta)$ then $E_z = \Pr\{\theta = 0|z\}$].

Parametric Families of Priors

- p -parameter exponential family:

$$\log g_{\alpha}(\theta) = \underset{1 \times p}{q(\theta)}^T \underset{p \times 1}{\alpha} + \text{constant}$$

- α the unknown parameter
- Marginal: $f_{\alpha}(z) = \int f_{\theta}(z) g_{\alpha}(\theta) d\theta$
- e.g., $q(\theta) = (\delta_0(\theta), \theta, \theta^2, \theta^3, \theta^4, \theta^5)^T$ [“spike and slab”]
- MLE: $\alpha \rightarrow g_{\alpha} \rightarrow f_{\alpha} \rightarrow (z_1, z_2, \dots, z_N) \xrightarrow{\text{marginal MLE}} \hat{\alpha}$
- $\hat{E}_z = \int t(\theta) f_{\theta}(z) g_{\hat{\alpha}}(\theta) d\theta / \int f_{\theta}(z) g_{\hat{\alpha}}(\theta) d\theta \pm ??$

Delta-Method Standard Deviation of $\hat{E}_z$

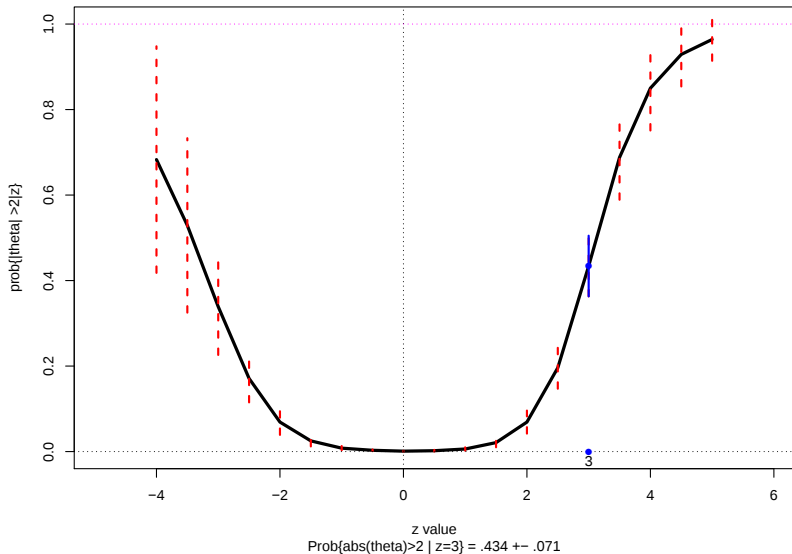
- $\text{sd}(\hat{E}_z) = \left\{ \left(\frac{dE_z}{d\alpha} \right)^T \mathcal{I}_\alpha^{-1} \left(\frac{dE_z}{d\alpha} \right) \right\}^{1/2}$
- Fisher information matrix:
 - ▶ $\bar{q}_\alpha = \int q(\theta) g_\alpha(\theta) d\theta$
 - ▶ $h_\alpha(z) = \int f_\theta(z) g_\alpha(\theta) [q(\theta) - \bar{q}] d\theta$

$$\mathcal{I}_\alpha = N \int \frac{h_\alpha(z) h_\alpha(z)^T}{f_\alpha(z)} dz$$

- $\frac{dE_z}{d\alpha} = E_z \int w(\theta) g_\alpha(\theta) [q(\theta) - \bar{q}] d\theta$ where

$$w(\theta) = t(\theta) f_\theta(z) g_\alpha(\theta) / \int t f g_\alpha - f_\theta(z) g_\alpha(\theta) / \int f g_\alpha$$

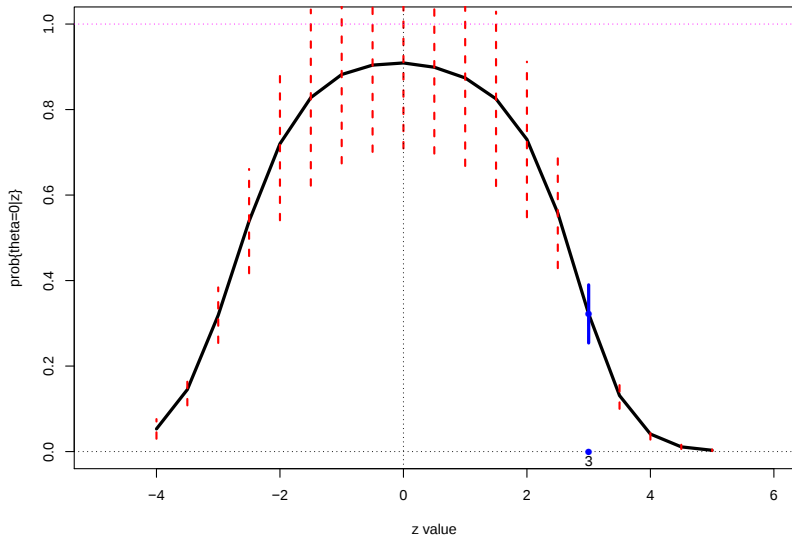
Figure 6. Estimated $\text{prob}\{\text{abs}(\theta) > 2 \mid z \text{ value}\}$ Prostate study;
Bars show $\pm$ one freq stdev. Using g-model $\{0, \text{ns}(5)\}$



Estimated False Discovery Rate

- Local false discovery rate: $\text{fdr}(z) = \Pr\{\theta = 0|z\} = E\{t(\theta)|z\}$
where $t(\theta) = \delta_0(\theta)$
- Next: applied to prostate study

Figure 7. Estimated $\text{prob}\{\theta=0 \mid z \text{ value}\}$ Prostate study;
Bars show $\pm$ one freq stdev. Using g-model $\{0, \text{ns}(5)\}$



$\text{Prob}\{\theta=0 \mid z=3\} = .322 \pm .068$, coeff of var = .21

For $z = 3$:

- $\widehat{\Pr}\{|\theta| > 2 | z = 3\} = 0.43 \pm 0.07$
- $\widehat{\text{fdr}}(3) = 0.32 \pm 0.07$
- “locfdr”: exfam modeling of z 's, not θ 's, gave $\widehat{\text{fdr}}(3) = 0.35 \pm 0.03$
but doesn't work for $|\theta| > 2$, etc.

References

- Park and Casella (2008) “The Bayesian Lasso” *JASA* 681–686
- Singh et al. (2002) “Gene expression correlates of clinical prostate cancer behavior” *Cancer Cell* 203–209
- Efron, Hastie, Johnstone and Tibshirani (2004) “Least angle regression” *Ann. Statist.* 407–499
- Efron (2012) “Bayesian inference and the parametric bootstrap” *Ann. Appl. Statist.* 1971–1997
- Efron (2010) *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* IMS Monographs, Cambridge University Press