Frequentist Accuracy of Bayesian Estimates

Bradley Efron

Stanford University
Bayesian Inference

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

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

- What if we don’t know $g$?
Jeffreysonian Bayes Inference

“Uninformative Priors”

- Jeffreys: $\pi(\mu) = |I(\mu)|^{1/2}$ where $I(\mu) = \text{cov}\left\{\nabla_\mu \log f_\mu(x)\right\}$ (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

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

Lemma

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

Theorem

The delta-method standard deviation of $E$ is

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

- Posterior Sample $\mu^*_1, \mu^*_2, \ldots, \mu^*_B$ (MCMC)
- $\hat{\theta}^*_i = t(\mu^*_i) = t_i^*$ and $\alpha^*_i = \alpha_x(\mu^*_i)$

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

$$\hat{sd}(\hat{E}) = \left[ \hat{\text{cov}}^T V_x \hat{\text{cov}} \right]^{1/2}$$
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: $y_{n \times 1} = X_{n \times p} \alpha_{p \times 1} + e_{n \times 1}$ \[ e \sim \mathcal{N}_n(0, I) \]
Diabetes data. Lasso: \( \text{min}\{\text{RSS}/2 + \gamma \text{L1}(\beta)\} \)

Cp minimized at Step 7, with \( \gamma = .37 \)
Bayesian Lasso

Park and Casella (2008)

- Model: \( y \sim \mathcal{N}_n(X \alpha, I) \) \( \mu = \alpha \) and \( x = y \)
- Prior: \( \pi(\alpha) = e^{-\gamma L_1(\alpha)} \) \[ \gamma = 0.37 \]
- Then posterior mode at Lasso \( \hat{\alpha}_\gamma \)
- Subject 125: \( \theta_{125} = x_{125}^T \alpha \)
- How accurate are Bayes posterior inferences for \( \theta_{125} \)?
Bayesian Analysis

- MCMC: posterior sample \( \{\alpha_i^* \text{ for } i = 1, 2, \ldots, 10,000\} \)
- Gives \( \{\theta_{125,i}^* = x_{125}^T \alpha_i^*, \text{ } i = 1, 2, \ldots, 10,000\} \)

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

- General accuracy formula: frequentist sd 0.071 for \( \hat{E} = 0.248 \)
- 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 $\theta_{125}$

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 $\mu_{125}$, Diabetes data, Prior $\exp\{-0.37*L_1(\alpha)\}$; verts are $\pm$ One Frequentist Standard Dev

Upper 95% credible limit is $0.336 \pm 0.071$

$\text{Prob}\{\mu_{125} < c \mid \text{data}\}$
**Exponential Families**

- \( f_\alpha(\hat{\beta}) = e^{\alpha^T \hat{\beta} - \psi(\alpha)} f_0(\hat{\beta}) \) 
  “\( x = \hat{\beta} \) and “\( \mu = \alpha \)”

- **General Accuracy Formula** \( \alpha = \alpha_x(\mu) \)

- For \( \hat{E} = E\{t(\alpha)|\hat{\beta}\} \)

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

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: \( f_\alpha(\cdot) \rightarrow \{ \hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_B^* \} \)
- Want to calculate: \( \hat{E} = E \left\{ t(\alpha) \mid \hat{\beta} \right\} \) for prior \( \pi(\alpha) \)
- Importance sampling:

\[
\hat{E} = \frac{\sum_{1}^{B} t_i \pi_i R_i}{\sum_{1}^{B} \pi_i R_i}
\]

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

\[
R_i = \frac{f_{\hat{\alpha}_i^*}(\hat{\beta})}{f_{\hat{\alpha}}(\hat{\beta}_i^*)}
\]

(Easy in exponential families.)
Bootstrap for GAC

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

\[
\hat{\text{sd}} = \left[ \hat{\text{cov}}^T V_{\hat{\alpha}} \hat{\text{cov}} \right]^{1/2}
\]
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: $y_j^{\text{ind}} \sim \text{Poi}(\mu_j)$

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

[MLE: \text{glm}(y \sim \text{poly}(c,8), \text{Poisson})]
Bayesian Estimation for the Poisson Model

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

[MLE: Fdr_\hat{\alpha}(3) = 0.183]
Figure 5. Prostate study: posterior density for Fdr(3) based on 4000 parametric bootstraps from Poisson poly(8) gave posterior \((m, sd) = (0.183, 0.025); \) Frequentist Sd for 0.183 equalled 0.026; CV=14% 

Internal coefficient of variation = 0.0023

\[ E(t|x) = 0.183 \]
Model Selection Calculations

- **Full model:** \( y \sim \text{Poi}(\mu), \quad \mu = e^{X\alpha} \quad [\alpha \in \mathbb{R}^9, \ \mu \in \mathbb{R}^{49}] \)

- **Submodel** \( \mathcal{M}_m \): \( \{ \mu : \text{only 1st } m + 1 \text{ coordinates of } \beta \neq 0 \} \)

- **Bayesian model selection:** prior probabilities on and within each \( \mathcal{M}_m \)

- **Poor man’s model estimates:** partition \( \mathbb{R}^{49} \) into “preference regions” \( \mathcal{R}_m = \{ \mu \text{ closest to } \mathcal{M}_m \} \)

- **Calculate posterior probabilities** \( \text{Pr}\{\mathcal{R}_m|y\} \)
Posterior Model Probabilities

- Distance: minimum AIC from $\mu$ to point in $R_m$
- Preferred model: is one with smallest distance

<table>
<thead>
<tr>
<th>$R_m$</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>posterior prob</td>
<td>.36</td>
<td>.12</td>
<td>.05</td>
<td>.02</td>
<td>.45</td>
</tr>
<tr>
<td>sd</td>
<td>.32</td>
<td>.16</td>
<td>.08</td>
<td>.03</td>
<td>.40</td>
</tr>
</tbody>
</table>
Empirical Bayes Accuracy

- $z_k =$ observed value for gene $k$
- $\theta_k =$ “true effect size”

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

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

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

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

- $p$-parameter exponential family:

\[
\log g_\alpha(\theta) = q(\theta)^T \alpha_{1 \times p} + \text{constant}
\]

- $\alpha$ 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: $\alpha \rightarrow g_\alpha \rightarrow f_\alpha \rightarrow (z_1, z_2, \ldots, z_N) \xrightarrow{\text{marginal MLE}} \hat{\alpha}$

- $\hat{E}_z = \int t(\theta) f_\theta(z) g_{\hat{\alpha}}(\theta) \ d\theta / \int f_\theta(z) g_{\hat{\alpha}}(\theta) \ d\theta \pm ??$
Delta-Method Standard Deviation of $\hat{E}_z$

- $\text{sd}(\hat{E}_z) = \left\{ \left( \frac{dE_z}{d\alpha} \right)^T I^{-1}_\alpha \left( \frac{dE_z}{d\alpha} \right) \right\}^{1/2}$

- Fisher information matrix:
  - $\bar{q}_\alpha = \int q(\theta) g_\alpha(\theta) \, d\theta$
  - $h_\alpha(z) = \int f_\theta(z) g_\alpha(\theta) [q(\theta) - \bar{q}] \, d\theta$

\[
I_\alpha = N \int \frac{h_\alpha(z) h_\alpha(z)^T}{f_\alpha(z)} \, dz
\]

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

\[
w(\theta) = \frac{t(\theta)f_\theta(z)g_\alpha(\theta)}{\int tfg_\alpha - f_\theta(z)g_\alpha(\theta)} \frac{1}{\int fg_\alpha}
\]
Figure 6. Estimated prob\{abs(\theta)>2 \mid z \text{ value}\} Prostate study; Bars show $\pm$ one freq stdev. Using g-model \{0, ns(5)\}

$\text{Prob}\{\text{abs}(\theta)>2 \mid z=3\} = .434 \pm .071$
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 $\text{prob}\{\theta=0 \mid z\}$ 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 $z = 3$:

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