Frequentist Accuracy of Bayesian Estimates

Bradley Efron

Stanford University

RSS Journal Webinar

Objective Bayesian Inference

- Probability family $\mathcal{F} = \{f_\mu(x),\ \mu \in \Omega\}$
- Parameter of interest: $\theta = t(\mu)$
- Prior $\pi(\mu)$

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

- Objective Bayes $\pi(\mu)$ " uninformative" e.g., Jeffreys:

$$\pi(\mu) = |I(\mu)|^{1/2}$$

- How accurate is $\hat{\theta}$?

General Accuracy Formula

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

Lemma

$\hat{\theta} = E\{t(\mu) \mid x\}$ has gradient $\nabla_x \hat{\theta} = \operatorname{cov}\{t(\mu), \alpha_x(\mu) \mid x\} = \widehat{\operatorname{cov}}$

Theorem

The delta-method standard deviation of $\hat{\theta}$ is

$$\boxed{\operatorname{sd}\left(\hat{\theta}\right) = \left[\widehat{\operatorname{cov}}^\top V_x \widehat{\operatorname{cov}}\right]^{1/2}.}$$

Implementation

- Posterior Sample $\mu_1^*, \mu_2^*, \ldots, \mu_B^*$ *(MCMC)*
- $t(\mu_i^*) = t_i^*$ and $\alpha_i^* = \alpha_x(\mu_i^*)$
- $\hat{\theta} = \sum_{i=1}^{B} t_i^* / B$

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

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

Diabetes Data

Efron et al. (2004), Park and Casella (2008)

- $n = 442$ subjects
- $p = 10$ predictors: age, sex, bmi, glu,...
- *Response* $y =$ disease progression at one year
- Model: $\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}} \underset{p\times 1}{\alpha} + \underset{n\times 1}{\boldsymbol{e}} \qquad [\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, I)]$
- Prior $\pi(\alpha) = e^{-\lambda\|\alpha\|_1} \qquad (\lambda = 0.37)$

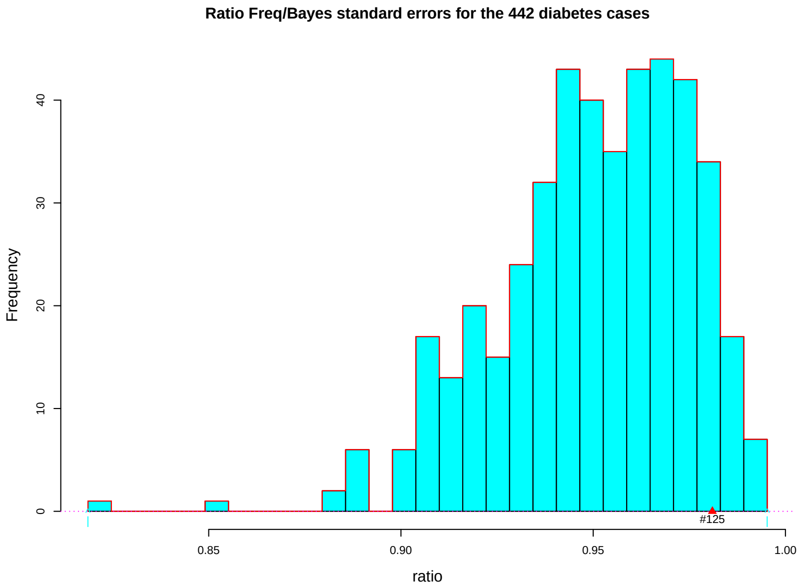
Estimating $\theta_j = x_j^\top \alpha$

The predicted value for patient j

- *MCMC* posterior sample $\{\alpha_1^*, \alpha_2^*, \ldots, \alpha_B^*,\ B = 10{,}000\}$
- $t_{ji}^* = x_j^\top \alpha^*$ gives $\hat{\theta}_j = \sum_{i=1}^{B} t_{ji}^*/B$
- $\hat{\theta}_{125} = 0.248$ $\qquad \mathrm{sd}_{\mathrm{Bayes}} = 0.072 \qquad \mathrm{sd}_{\mathrm{freq}} = 0.071$

$$\boxed{\frac{\mathrm{sd}_{\mathrm{freq}}}{\mathrm{sd}_{\mathrm{Bayes}}} = \left(\frac{x^\top \hat{\Sigma} \boldsymbol{G} \hat{\Sigma} x}{x^\top \hat{\Sigma} x}\right)^{1/2}}$$

- $\boldsymbol{G} = \boldsymbol{X}^\top \boldsymbol{X}$ and $\hat{\Sigma}$ = empirical covariance matrix of $\{\alpha_i^*\}$
- Eigen range $\left(\hat{\Sigma}^{1/2} \boldsymbol{G} \hat{\Sigma}^{1/2}\right)^{1/2} = (1.007, 0.313)$



Ratio Freq/Bayes standard errors for the 442 diabetes cases
Frequency
0
10
20
30
40
#125
0.85
0.90
0.95
1.00
ratio

Posterior CDF of θ_{125}

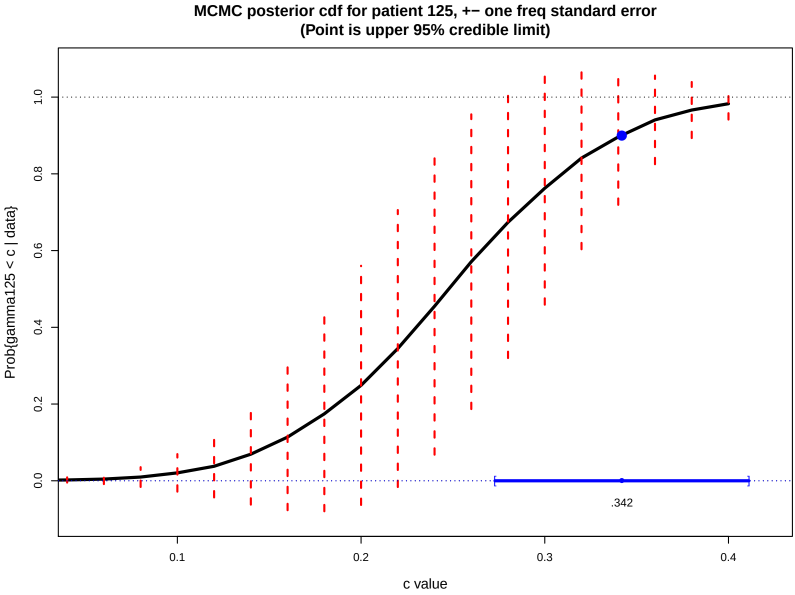
- Apply accuracy formula to $s_i^* = I\{t_i^* < c\}$
- $E\{s \mid \text{data}\} = \Pr\{\theta_{125} \le c \mid \text{data}\} = \widehat{\text{cdf}}(c)$

$$\widehat{\text{cdf}}(0.3) = 0.762 \pm 0.304$$

$\nearrow$ Bayes estimate

$\nearrow$ frequentist sd

- Upper 95% credible limit $= 0.34 \pm 0.07$


MCMC posterior cdf for patient 125, +− one freq standard error
(Point is upper 95% credible limit)
Prob{gamma125 < c | data}
0.0
0.2
0.4
0.6
0.8
1.0
.342
0.1
0.2
0.3
0.4
c value

Exponential Families

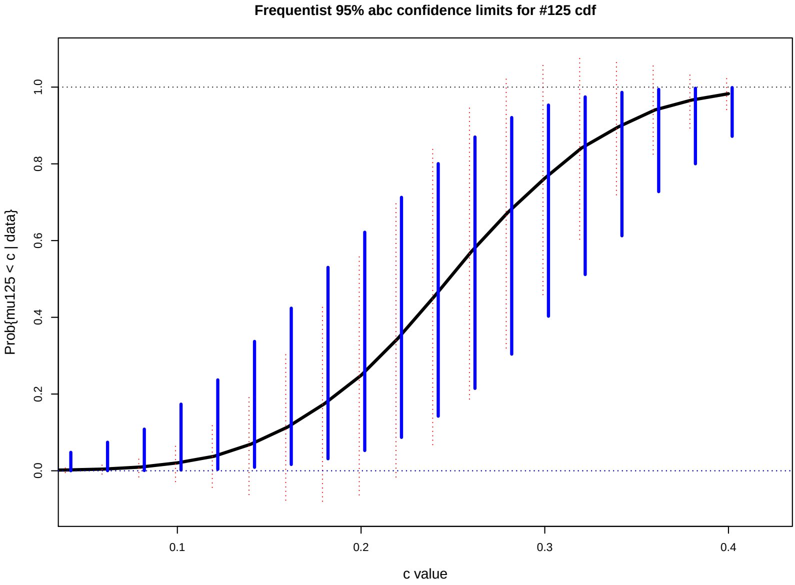
- $f_\alpha\left(\hat{\beta}\right) = e^{\alpha^\top \hat{\beta} - \psi(\alpha)} f_0\left(\hat{\beta}\right)$
- natural parameter α
- sufficient statistic $\hat{\beta}$

  (Now $\alpha =$"μ", $\hat{\beta} =$"x")
- "$\alpha_x(\mu)$"$= \alpha$

$$\boxed{\widehat{\text{sd}}\left(\hat{\theta}\right) = \text{cov}\left(t, \alpha \,\middle|\, \hat{\beta}\right)^\top V_{\hat{\alpha}} \, \text{cov}\left(t, \alpha \,\middle|\, \hat{\beta}\right)}$$

- $V_{\hat{\alpha}} = \ddot{\psi}(\hat{\alpha}) = \text{cov}_{\alpha=\hat{\alpha}}\left\{\hat{\beta}\right\}$

- *MCMC* $\{\alpha_1^*, \alpha_2^*, \ldots, \alpha_B^*\}$
- $\hat{g}(\alpha \mid \beta = \hat{\beta})$ prob $\frac{1}{B}$ on α_i^*
- *Reweighting* $\hat{g}(\alpha \mid \beta = b)$: prob $w_i(b) = ce^{(b-\hat{\beta})^\top \alpha_i}$ on α_i^* (exponential family in b)
- Obtain bootstrap confidence intervals for $t(\alpha)$ without further simulation (BCa, ABC)


Frequentist 95% abc confidence limits for #125 cdf
Prob{mu125 < c | data}
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
c value

Bootstrap Estimates of a Posterior Distribution

- $\mathcal{F} = \left\{ f_\mu(x),\ \mu \in \Omega \right\}$
- *MCMC* sample $\{\mu_1^*, \mu_2^*, \ldots, \mu_B^*\}$:

  Approximate $g(\mu \mid x)$ with uniform weight $\frac{1}{B}$ on each μ_i^*
- *Bootstrap* MLE $\hat{\mu} \longrightarrow$ bootsample $\hat{\mu}_i^* \sim f_{\hat{\mu}},\ i = 1, 2, \ldots, B$
- Approximate $g(\mu \mid x)$ with weight p_i on $\hat{\mu}_i^*$ where $p_i = \pi(\hat{\mu}_i^*) R_i$,

  R_i “conversion factor” (easy in exponential families)

Simulation Accuracy

- How big to take B?
- *Boot sampling* $\quad q_i = t_i^* \cdot p_i \quad$ and $\quad d_i = \frac{q_i}{\bar{q}} - \frac{p_i}{\bar{p}} \qquad \left[t_i^* = t\left(\hat{\mu}_i^*\right)\right]$
- Then $\hat{\theta} = \sum_1^B p_i t_i^*$ has simulation coefficient of variation

$$\boxed{\text{cv} \doteq \left(\sum_{i=1}^{B} d_i^2\right)^{1/2} \Big/ B}$$

(about 0.01 for $\hat{\theta}_{125}$ with $B = 10{,}000$)

How Uninformative is a Prior?

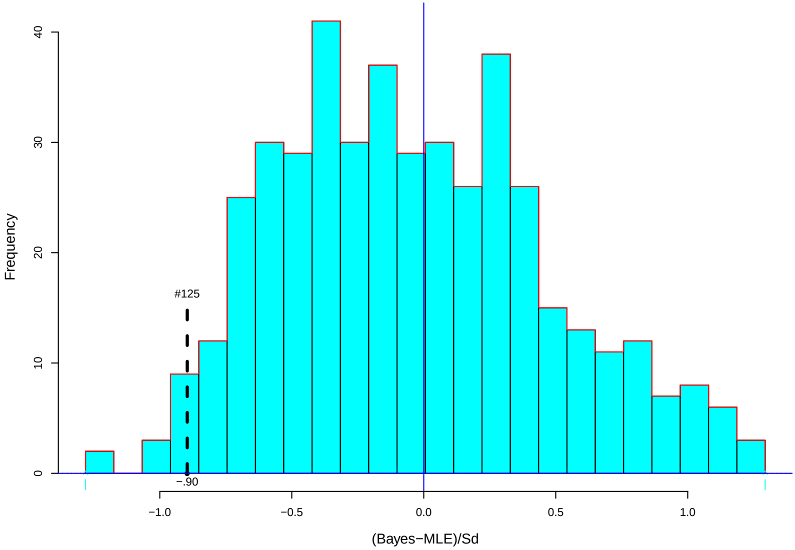
- Measure of prior's effect on estimate of $\theta = t(\mu)$ is

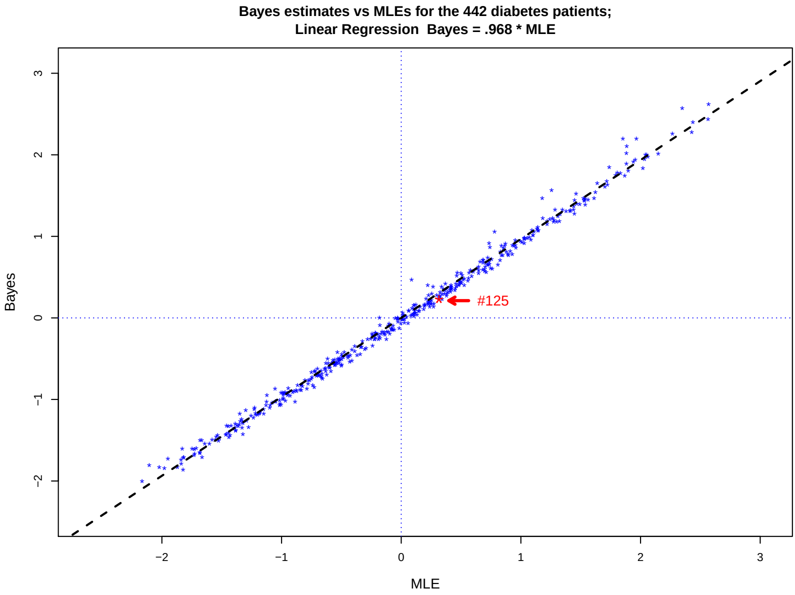
$$\delta = \frac{\overset{\text{Bayes}}{E\{\theta \mid x\}} - \overset{\text{MLE}}{t(\hat{\mu})}}{\widehat{\mathrm{sd}}_{\mathrm{MLE}}}$$

- For patient 125,

$$\delta_{125} = \frac{0.248 - 0.316}{0.076} = -0.90$$

Effect of Park/Cassella prior on patient estimates




Bayes estimates vs MLEs for the 442 diabetes patients;
Linear Regression Bayes = .968 * MLE
Bayes
MLE
#125
−2
−1
0
1
2
3
−2
−1
0
1
2
3

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