Frequentist Accuracy of Bayesian Estimates

Bradley Efron

Stanford University

Bayesian Inference

- Parameter: $\mu \in \Omega$

- Observed data: x

- Prior: $\pi(\mu)$

- Probability distributions: $\left\{f_\mu(x), \mu \in \Omega\right\}$

- Parameter of interest: $\theta = t(\mu)$

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

- What if we don't know g?

Jeffreysonian Bayes Inference

"Uninformative Priors"

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

  (the Fisher information matrix)

- Can still use Bayes theorem but how accurate are the estimates?

- Today: frequentist variability of $E\left\{t(\mu)|x\right\}$

General Accuracy Formula

- μ and $x \in \mathcal{R}^p$

- $x \sim (\mu, V_\mu)$

- $\alpha_x(\mu) = \nabla_x \log f_\mu(x) = \left(\ldots, \frac{\partial \log f_\mu(x)}{\partial x_i}, \ldots\right)^T$

Lemma

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

Theorem

The delta-method standard deviation of E is

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

Implementation

- Posterior Sample $\quad \mu_1^*, \mu_2^*, \ldots, \mu_B^* \quad$ (MCMC)

- $\hat{\theta}_i^* = t(\mu_i^*) = t_i^* \quad$ and $\quad \alpha_i^* = \alpha_x(\mu_i^*)$

$$\widehat{\text{cov}} = \sum_{i=1}^{B} \left(\alpha_i^* - \bar{\alpha} \right)\left(t_i^* - \bar{t} \right) \big/ B$$
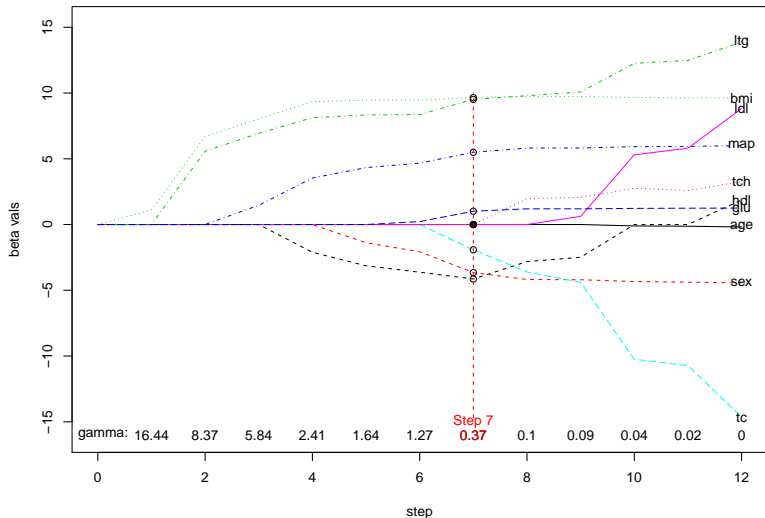
- $\widehat{\text{sd}}\left(\hat{E} \right) = \left[\widehat{\text{cov}}^T V_x \widehat{\text{cov}} \right]^{1/2}$

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: $\underset{n \times 1}{\boldsymbol{y}} = \underset{n \times p}{\boldsymbol{X}} \underset{p \times 1}{\alpha} + \underset{n \times 1}{\boldsymbol{e}} \qquad [\boldsymbol{e} \sim \mathcal{N}_n(\boldsymbol{0}, I)]$

Diabetes data. Lasso: min{RSS/2 + gamma*L1(beta)}
Cp minimized at Step 7, with gamma = .37

Bayesian Lasso

Park and Casella (2008)

- Model: $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\alpha, I)$ "μ"= α and "x"= $\mathbf{y}$

- Prior: $\pi(\alpha) = e^{-\gamma L_1(\alpha)}$ $[\gamma = 0.37]$

- Then posterior mode at Lasso $\hat{\alpha}_\gamma$

- Subject 125: $\theta_{125} = \mathbf{x}_{125}^T \alpha$
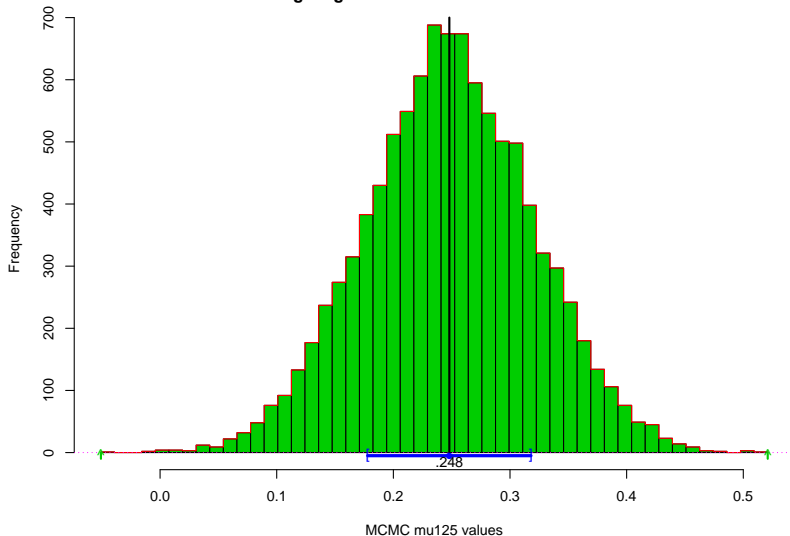
- How accurate are Bayes posterior inferences for θ_{125}?

Bayesian Analysis

- MCMC: posterior sample $\left\{\alpha_i^* \text{ for } i = 1, 2, \ldots, 10,000\right\}$

- Gives $\left\{\theta_{125,i}^* = \boldsymbol{x}_{125}^T \alpha_i^*, \ i = 1, 2, \ldots, 10,000\right\}$

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

- 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%
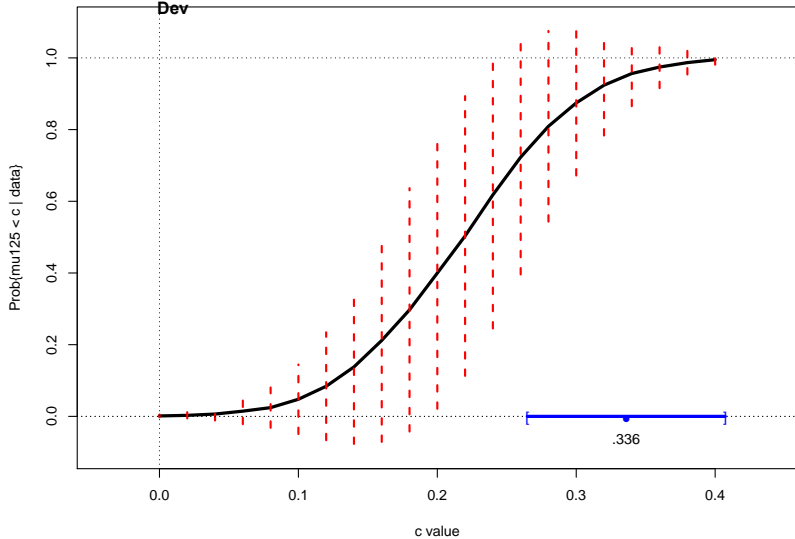
Posterior CDF of θ_{125}

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

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

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

$$\nearrow \qquad \nearrow$$

Bayes   GAC sd
estimate

Figure 3. MCMC posterior cdf of mu125, Diabetes data, Prior exp{−.37*L1(alpha)}; verts are +− One Frequentist Standard Dev

c value
Upper 95% credible limit is .336 +− .071

Exponential Families

- $f_\alpha\left(\hat{\beta}\right) = e^{\alpha^T \hat{\beta} - \psi(\alpha)} f_0\left(\hat{\beta}\right)$ "x"$= \hat{\beta}$ and "μ"$= \alpha$

- **G**eneral **A**ccuracy **F**ormula $\alpha = \alpha_x(\mu)$

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

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

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

Bayesian Estimation Using the Parametric Bootstrap

Efron (2012)

- Parametric bootstrap: $f_{\hat{\alpha}}(\cdot) \rightarrow \left\{ \hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_B^* \right\}$

- Want to calculate: $\hat{E} = E\left\{ t(\alpha) \big| \hat{\beta} \right\}$ for prior $\pi(\alpha)$

- Importance sampling:

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

$t_i = t\left(\hat{\alpha}_i^*\right)$, $\pi_i = \pi\left(\hat{\alpha}_i^*\right)$,   and $R_i = $ "conversion factor"

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

(Easy in exponential families.)

Bootstrap for GAC

- $p_i = \pi_i R_i \big/ \sum_1^B \pi_j R_j$ is weight on ith Bootstrap Replication

- $\hat{E} = \sum_{i=1}^B p_i t_i^*$

- $\widehat{\text{cov}} = \sum_{i=1}^B p_i \hat{\alpha}_i^* \left(t_i^* - \hat{E} \right)$ estimates $\text{cov}\left(\alpha, t \big| \hat{\beta} \right)$

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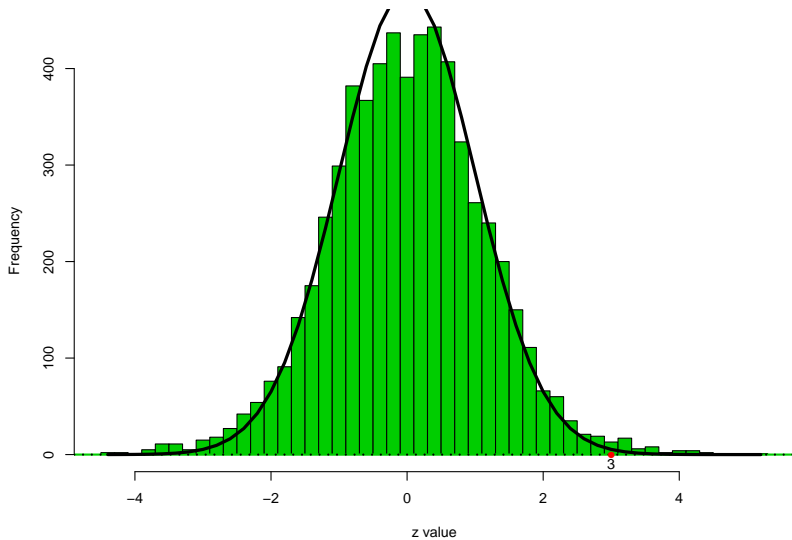
Prostate Cancer Study

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: $y_j \overset{\text{ind}}{\sim} \text{Poi}(\mu_j)$

$$\boxed{\log(\boldsymbol{\mu}) = \text{poly}(\mathbf{c}, \text{degree=8})}$$
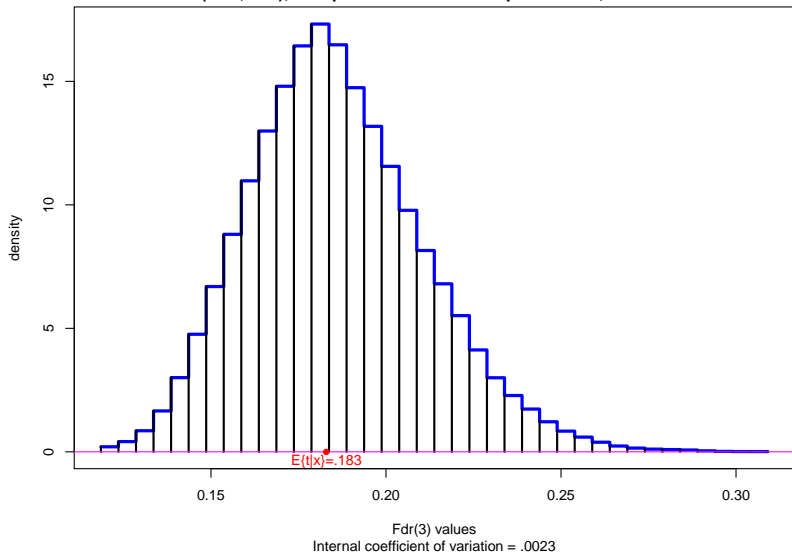
[MLE: `glm(`$\mathbf{y}$` ~ poly(`$\mathbf{c}$`,8), Poisson)`]

Bayesian Estimation for the Poisson Model

- Model: $\mathbf{y} \sim \text{Poi}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = e^{\mathbf{X}\alpha} \qquad [\mathbf{X} = \text{poly}(\mathbf{c}, 8)]$

- Prior: Jeffreys prior for α

- cdf: $\alpha \rightarrow \boldsymbol{\mu} \rightarrow$ cdf: $F_\alpha(z) = \sum_{c_i \leq z} \mu_i \Big/ \sum \mu_i$

- Fdr parameter: $t(\alpha) = \dfrac{1 - \Phi(3)}{1 - F(3)} = \text{Fdr}(3)$

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

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

Fdr(3) values
Internal coefficient of variation = .0023

Model Selection Calculations

- Full model: $\boldsymbol{y} \sim \text{Poi}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = e^{\boldsymbol{X}\alpha} \qquad \left[\alpha \in \mathcal{R}^9, \ \boldsymbol{\mu} \in \mathcal{R}^{49} \right]$

- Submodel $\mathcal{M}_m$: $\{\boldsymbol{\mu} :$ only 1st $m+1$ coordinates of $\beta \neq 0\}$

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

- Poor man's model estimates: partition $\mathcal{R}^{49}$ into
  "preference regions" $\mathcal{R}_m = \{\boldsymbol{\mu}$ closest to $\mathcal{M}_m\}$

- Calculate posterior probabilities $\text{Pr}\{\mathcal{R}_m | \boldsymbol{y}\}$

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 $\quad g(\cdot) \rightarrow \theta_1, \theta_2, \ldots, \theta_N \qquad [N = 6033]$

- From observations $z_1, z_2, \ldots, z_N$ we wish to estimate

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

[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., $q(\theta) = \left(\delta_0(\theta), \theta, \theta^2, \theta^3, \theta^4, \theta^5\right)^T$ ["spike and slab"]

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

- $\hat{E}_z = \int t(\theta) f_\theta(z) g_{\hat{\alpha}}(\theta) \, d\theta \Big/ \int f_\theta(z) g_{\hat{\alpha}}(\theta) \, d\theta \pm ??$

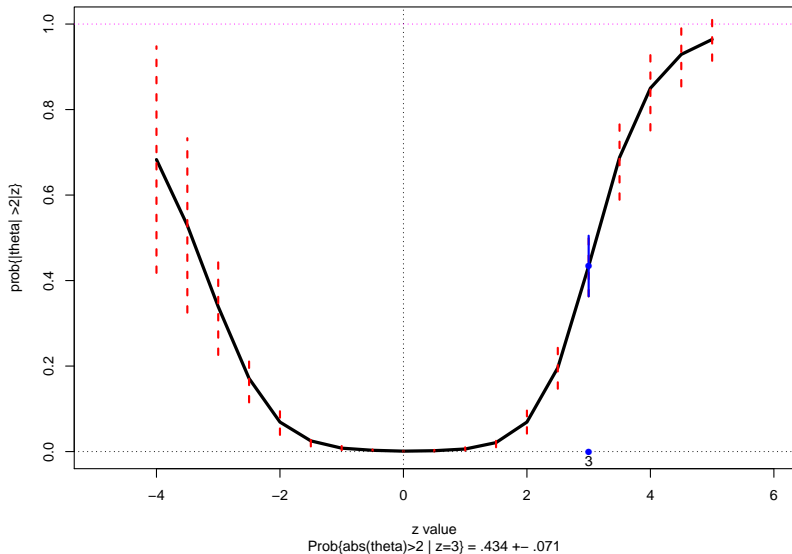*Delta-Method Standard Deviation of $\hat{E}_z$*

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

- Fisher information matrix:

  - $\bar{q}_\alpha = \int q(\theta) g_\alpha(\theta) \, d\theta$
  - $h_\alpha(z) = \int f_\theta(z) g_\alpha(\theta) \left[q(\theta) - \bar{q}\right] \, d\theta$

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

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

$$w(\theta) = t(\theta) f_\theta(z) g_\alpha(\theta) \Big/ \int t f g_\alpha - f_\theta(z) g_\alpha(\theta) \Big/ \int f g_\alpha$$
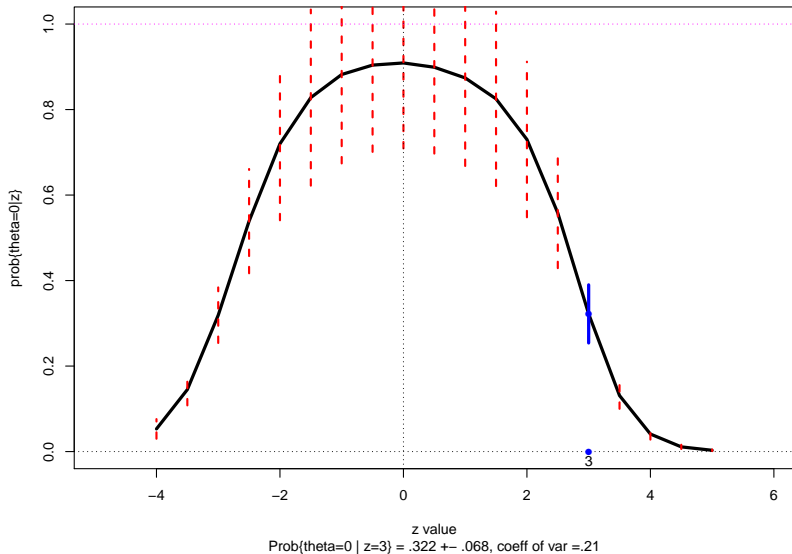
Figure 6. Estimated prob{abs(theta)>2 | z value} Prostate study;
Bars show +− one freq stdev. Using g−model {0,ns(5)}

z value
Prob{abs(theta)>2 | z=3} = .434 +− .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 prob{theta=0 | z value} Prostate study; Bars show +− one freq stdev. Using g−model {0,ns(5)}

z value

Prob{theta=0 | z=3} = .322 +− .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