Bayesian POst model Selection Inference

Edward George
Daniel Yekutieli

December 13, 2012

Plan

1. Background
   - Bayesian Selective inference
   - The submodel interpretation of parameters in model selection
   - Frequentist post model selection inference

2. Bayesian post model selection inference
   - Phrasing POSI as a Bayesian selective inference problem
   - Bayesian POSI in a simplified example

Selective Inference?

Benjamini and Yekutieli '05, two separate type problems can arise when providing inference for multiple parameters:

1. Simultaneity is the need to provide inferences that apply to all the parameters.
2. Selective inference refers to the practice of providing marginal statistical inferences for parameters that are selected after viewing the data.

Selective inference a new problem?

- Post hoc inferences are selective inferences. Tukey '53, ''The Problem of Multiple Comparisons," suggests controlling $FWE \leq \alpha$ because it ensures that the probability of making any selected type-I error is less than α.
- Soric '89: Goal of many studies is making statistical discoveries (= finding non-zero effects); it is mainly the discoveries that are reported and included into science; unless the proportion of false discoveries is kept small there is danger that a large part of science is untrue ...
- Ioannidis '05, "Why most research findings are false?"
- Berk et al. '12, "Valid Post-Selection Inference"
- Weinstein et al. '12, "Selection Adjusted Confidence Intervals with More Power to Determine the Sign"

FDR control – a frequentist mechanism for ensuring that $\approx 95\%$ of the research findings are true

Benjamini and Hochberg '95 considered the problem of testing $H_1 \cdots H_m$

- A discovery is rejecting a null hypothesis
- A false discovery is erroneously rejecting a true null hypothesis
- Define

$$FDR = E\{V/\max(R,1)\}$$

R – number of discoveries
V – number of false discoveries

- BH procedure: a multiple testing procedure that controls $FDR \leq q$, i.e. ensure that $\approx 1 - q$ of the discoveries are true discoveries

FCR control – a frequentist mechanism for ensuring that $\approx 95\%$ of the inferences for the selected parameters are true

Benjamini and Yekutieli '05 generalize the BH testing framework:

1. m parameters $\theta_1 \cdots \theta_m$ with corresponding estimators $T_1 \cdots T_m$
2. Construct CI's for selected parameters $\tilde{S}(T_1 \cdots T_m) \subseteq \{1 \cdots m\}$

- $1 - q$ CIs for selected parameters don't offer marginal $1 - q$ coverage probability.
- Suggest the False Coverage-statement Rate as a measure for the validity of CI's constructed for the selected parameters $FCR = E\{V/\max(R, 1)\}$ where $R = |\tilde{S}(T_1 \cdots T_m)|$, V – number of non-covering CIs,
- For independent $T_1 \cdots T_m$ and any selection rule: constructing marginal $1 - R \cdot q/m$ CI's for each selected parameter ensures $FCR \leq q$

Bayesian selective inference

Bayesian selective inference framework:

- θ is the parameter, Y is the data and Ω is the data sample space.
- $\pi(\theta)$ is the prior distribution and $f(y|\theta)$ is the likelihood function.
- The multiple parameters, for which inference may or may not be provided, are actually multiple functions of $\theta : h_1(\theta), h_2(\theta), \ \ldots$
- For each $h_i(\theta)$ there is a given subset $S^i_\Omega \subseteq \Omega$, such that inference is provided for $h_i(\theta)$ only if $y \in S^i_\Omega$ is observed.

Bayesian selective inference – a truncated data problem

- As inference is provided for $h_i(\theta)$ only if $y \in S^i_\Omega$, $Y = y$ used for providing selective inference for $h_i(\theta)$ is a realization of $f_{S^i}(\theta, y)$, the joint distribution of (θ, Y) truncated by the event that $y \in S^i_\Omega$.

- We define $f_{S^i}(\theta, y)$ through a average risk:
  if selective inference for $h_i(\theta)$ involves an action $\delta_i(Y)$ associated with a loss function $L(h_i(\theta), \delta_i)$, then $f_{S^i}(\theta, y)$ is the distribution over which the expected loss

$$\int_\theta \int_{y \in S^i_\Omega} f_{S^i}(\theta, y) \cdot L(h_i(\theta), \delta_i(y)) \, dy d\theta \tag{1}$$

  is the average risk incurred in selective inference for $h_i(\theta)$.

Example 1: Predicting students' academic abilities

We wish to predict a student's "true" academic ability from his/her observed academic ability – but only for students that are admitted to college

- True academic ability $\theta_i \sim N(0, 1)$
- Observed academic ability in high school $Y_i \sim N(\theta_i, 1)$
- Only Students with $0 < Y_i$ are admitted to college.

Find $f_S(\theta_i, y_i)$ for which prediction error is

$$\int_{\theta_i=-\infty}^{\infty} \int_{y_i=0}^{\infty} f_S(\theta_i, y_i) \cdot \{\theta_i - \delta(y_i)\}^2 \, dy_i d\theta_i$$

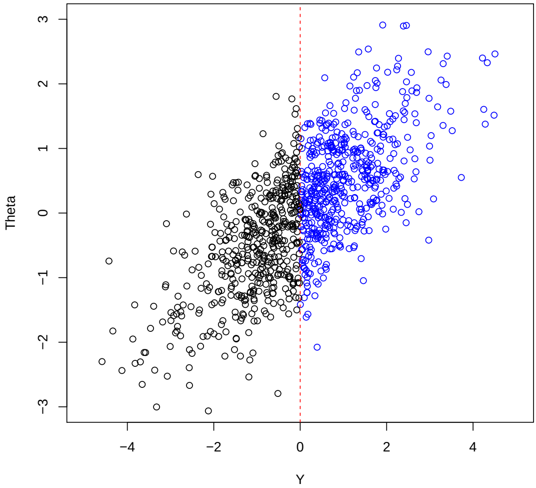
Predicting true ability for a random college student

Consider the case of a college professor predicting θ_i for a student in his class

- Joint distribution of (θ_i, Y_i) for college student is generated by drawing (θ_i, Y_i) for a high school student and keeping it only if $0 < Y_i$.
- Thus, joint density of (θ_i, y_i) used for *predicting* θ_i is

$$f_S(\theta_i, y_i) \;\propto\; e^{-\frac{\theta_i^2}{2}} \cdot e^{-\frac{(\theta_i - y_i)^2}{2}} \,/\, \Pr(Y_i > 0) \;\propto\; e^{-\frac{(\theta - y_i/2)^2}{2\cdot(1/2)}}$$

Distribution of (θ_i, Y_i) for college students



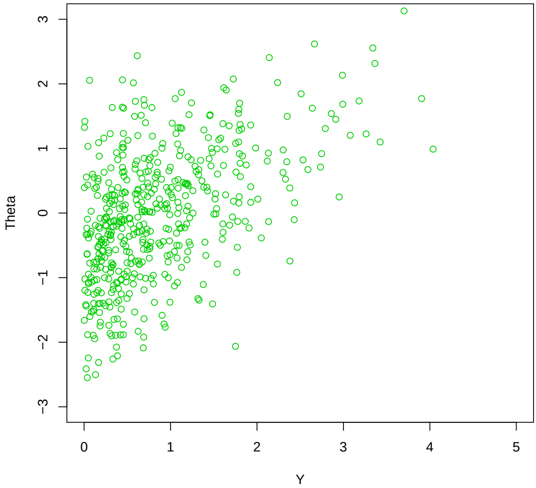
Predicting true ability for a high school student

Now consider the case of a counselor at high school predicting θ_i for a student arriving for mandatory counseling when there is a high school regulation instructing counselors to predict academic ability only for students that can be admitted to college

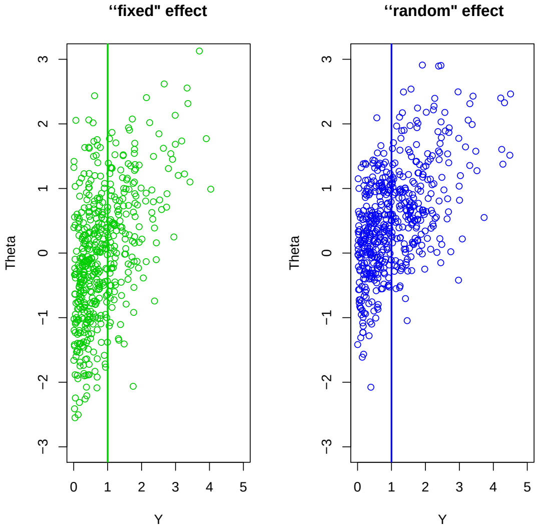
- Thus in this case $\theta_i \sim N(0, 1)$
- and then Y_i used to predict θ_i is drawn from the $N(\theta_i, 1)$ density truncated by the event $0 < Y_i$
- Now the joint density of (θ_i, Y_i) used for predicting θ_i is

$$f_S(\theta_i, y_i) \;\propto\; e^{-\frac{\theta_i^2}{2}} \cdot e^{-\frac{(\theta_i - y_i)^2}{2}} / \Pr(Y_i > 0|\theta_i)$$

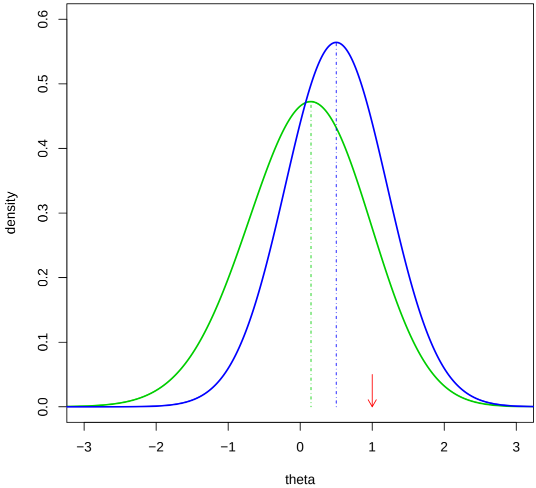
Distribution of (θ_i, Y_i) for high school students



Joint distribution of selected (θ_i, Y_i)



Conditional density of $\theta_i | Y_i = 1$



Return to distribution of (θ_i, Y_i) for college students

- The joint truncated distribution of (θ_i, Y_i)

$$f_S(\theta_i, y_i) = \frac{I_{S_\Omega}(y_i) \cdot \pi(\theta_i) \cdot f(y_i|\,\theta_i)}{\Pr(Y_i \in S_\Omega)} \tag{2}$$

- The marginal truncated distribution of θ_i

$$\pi_S(\theta_i) \;=\; \int_{S_\Omega} \frac{\pi(\theta_i) \cdot f(y_i|\,\theta_i)}{\Pr(Y_i \in S_\Omega)} dy_i \;=\; \frac{\pi(\theta_i) \cdot \Pr(Y_i \in S_\Omega|\,\theta_i)}{\Pr(Y_i \in S_\Omega)} \tag{3}$$

- Dividing (2) by (3) yields the truncated conditional distribution of $Y_i|\theta_i$

$$f_S(y_i|\theta_i) \;=\; I_{S_\Omega}(y_i) \cdot f(y_i|\,\theta_i) / \Pr(Y_i \in S_\Omega|\,\theta_i)$$

Return to distribution of (θ_i, Y_i) for high-school students

- The joint truncated distribution of (θ_i, Y_i)

$$f_S(\theta_i, y_i) = \frac{I_{S_\Omega}(y_i) \cdot \pi(\theta_i) \cdot f(y_i|\ \theta_i)}{\Pr(Y_i \in S_\Omega|\theta_i)} \tag{4}$$

- The marginal truncated distribution of θ_i

$$\pi_S(\theta_i) \ = \ \int_{S_\Omega} \frac{\pi(\theta_i) \cdot f(y_i|\ \theta_i)}{\Pr(Y_i \in S_\Omega|\theta_i)} dy_i \ = \ \pi(\theta_i) \tag{5}$$

- And again, the truncated conditional distribution of $Y_i|\theta_i$

$$f_S(y_i|\theta_i) \ = \ I_{S_\Omega}(y_i) \cdot f(y_i|\ \theta_i) / \Pr(Y_i \in S_\Omega|\ \theta_i)$$

Selection-adjusted Bayesian inference

1. The selection-adjusted prior distribution is $\pi_S(\theta)$.
2. The selection adjusted likelihood is the truncated distribution of $Y|\theta$

$$f_S(y|\theta) \;=\; I_{S_\Omega}(y) \cdot f(y|\,\theta) / \Pr(Y \in S_\Omega|\,\theta)$$

3. Bayes rules are based on the selection-adjusted posterior distribution

$$\pi_S(\theta|\,y) \propto \pi_S(\theta) \cdot f_S(y|\theta)$$

Fixed and Random parameters

- θ is called a *fixed parameter* in cases where only $Y|\theta$ is truncated. Fixed θ are generally fixed effects whose value can be thought to be generated once from $\pi(\theta)$ and remain unchanged. For fixed θ

$$\pi_S(\theta) = \pi(\theta)$$

- θ is called a *random parameter* in cases where the joint distribution of (θ, Y) is truncated. Random θ's are usually the random effects whose values are generated, and thus truncated, concurrently with the data. For random θ

$$\pi_S(\theta_i) \propto \pi(\theta_i) \cdot \Pr(Y_i \in S_\Omega | \, \theta_i)$$

- If θ is assigned a *non informative prior* than the same non informative prior is also used for the selection-adjusted prior

$$\pi_S(\theta) = \pi(\theta)$$

Berk et el. '12: the full model interpretation of parameters

- The full model is

$$\boldsymbol{Y} = \boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}\ , \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I_{n\times n})$$

- Submodels are denoted

$$M = \{j_1 \cdots j_m) \subset \{1 \cdots p\}, \;\; \boldsymbol{X}_M = \{\boldsymbol{X}_{j_1} \cdots \boldsymbol{X}_{j_m}\}.$$

- The target of inference is $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ or some functionals thereof.
- $\boldsymbol{\beta}_M = \{\beta_i : i \in M\}$ selected after viewing the data assumed to consist of the non-zero components of $\boldsymbol{\beta}$.
- Thus, $\boldsymbol{\mu}_M = \boldsymbol{X}_M\boldsymbol{\beta}_M$ are regarded as a computational compression and a parsimonious statistical summary $\boldsymbol{\mu}$, and neither as models in their own right nor as objects of future scientific research.

The submodel interpretation of parameters

Berk et al. '12, suggest an alternative interpretation for selected sub models

- *The submodel interpretation of parameters.* A submodel M corresponds to a subset of $\boldsymbol{\beta}$ and $\{\beta_i : i \notin M\}$, the deselected components of $\boldsymbol{\beta}$, are non-existent.
- Hence the relevant components are only those in $\boldsymbol{\beta}_M$ and these will generally differ from their siblings in $\boldsymbol{\beta}$.
- Thus selecting model M implies that the target of estimation

$$\boldsymbol{\beta}_M = (\mathbf{X}_M^T \mathbf{X}_M)^{-1} \mathbf{X}_M^T X \boldsymbol{\beta} \iff \boldsymbol{\mu}_M = X_M \boldsymbol{\beta}_M.$$

- With $\beta_{j \cdot M}$ denoting is the coefficient of the j'th predictor in $\boldsymbol{X}$ "adjusted" for the other predictors in M.
- $\boldsymbol{\mu}_M$ is the projection of $\boldsymbol{\mu}$ on the vector space spanned by $\mathbf{X}_M$.

Berk et al. '12: valid frequentist POSI

- Estimator: $\hat{\boldsymbol{\beta}}_M = (\mathbf{X}_M^T\mathbf{X}_M)^{-1}\mathbf{X}_M^T\boldsymbol{Y}$ with $\hat{\beta}_{j\cdot m} \sim N(\beta_{j\cdot m}, (\mathbf{X}_M^T\mathbf{X}_M)_{jj}^{-1}\sigma^2)$
- A marginal $1-\alpha$ CI for $\beta_{j\cdot M}$:

$$\hat{\beta}_{j\cdot m} \pm K \cdot \sqrt{(\mathbf{X}_M^T\mathbf{X}_M)_{jj}^{-1} \cdot s^2}\ , \quad K = t_{n-p,1-\alpha/2}$$

- To ensure valid POSI (for any coeff. in any selected model) Berk et al. '12 propose using a larger $K = K(\boldsymbol{X})$ ensuring simultaneous coverage

$$\Pr\{\ \forall M,\ \forall j \in M,\ \ \beta_{j\cdot M} \in \hat{\beta}_{j\cdot m} \pm K(\boldsymbol{X}) \cdot S.E.(\hat{\beta}_{j\cdot M})\ \}\ \geq\ 1-\alpha$$

- $K(\boldsymbol{X})$ can be $O(\sqrt{p})$, for orthogonal designs $K(\boldsymbol{X}) \sim O(\sqrt{2\log p})$, and there are also cases where $K(\boldsymbol{X}) \sim O(\sqrt{\log p})$

Phrasing POSI as a Bayesian selective inference problem

Adopting the Berk et al. '12 submodel Interpretation of parameters:

- Given data generating model

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon},\ \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta},\ \boldsymbol{\epsilon} \sim N(0, \sigma^2 I_{n\times n}).$$

- The likelihood is $f(\boldsymbol{y}|\ \boldsymbol{\mu}, \sigma^2) = \Pi_{i=1}^n \phi((y_i - \mu_i)/\sigma)$.
- The prior distribution is $\pi(\boldsymbol{\mu}, \sigma^2)$.
- The posterior given by $\pi(\boldsymbol{\mu}, \sigma^2 |\, \boldsymbol{y}) \propto f(\boldsymbol{y}|\ \boldsymbol{\mu}, \sigma^2) \cdot \pi(\boldsymbol{\mu}, \sigma^2)$.
- *There* is a model selection scheme: $M \to S_\Omega^M$,

$$\boldsymbol{y} \in S_\Omega^M \qquad \Rightarrow \quad \text{target for inference is } \ h_M(\boldsymbol{\mu}) = \boldsymbol{\mu}_M.$$

Bayes rules in Bayesian POSI

- Thus the distribution of $(\boldsymbol{\mu}, \sigma^2, \boldsymbol{Y})$ used for estimating $\boldsymbol{\mu}_M$ is truncated by the event $\boldsymbol{Y} \in S_\Omega^M$.
- The average risk incurred by estimating $\boldsymbol{\mu}_M$ is

$$\int_{\boldsymbol{\mu}} \int_{\sigma^2} \int_{\boldsymbol{y} \in S_\Omega^M} L(\boldsymbol{\mu}_M, \delta_M(\boldsymbol{Y}))\, f_S(\boldsymbol{\mu}, \sigma^2, \boldsymbol{Y}) d\boldsymbol{\mu}\, d\sigma^2\, d\boldsymbol{Y}.$$

- This implies that the Bayes rules in POSI are based on the selection adjusted posterior distribution

$$\pi_S(\boldsymbol{\mu}, \sigma^2 | \, \boldsymbol{Y} = \boldsymbol{y}) \propto f_S(\boldsymbol{\mu}, \sigma^2, \boldsymbol{y}).$$

Example 2: Linear regression with orthonrmal X-matrix, known σ^2 and marginal selection rule

- We study model selection on a linear regression

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon},\ \ \boldsymbol{\epsilon}_{n\times 1} \sim N(0, 1\cdot \boldsymbol{I}_{n\times n})$$

- $\boldsymbol{X}_{n\times p} = \{X_1 \cdots X_p\}$ with $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_{n\times n}$
- Coefficient vector $\boldsymbol{\beta}_{p\times 1} = (\beta_1 \cdots \beta_p)^T$ with iid

$$\beta_i \sim 0.9 \cdot N(0, 0.2^2) + 0.1 \cdot N(0, 2^2) \tag{6}$$

- Coefficient vector estimator

$$\boldsymbol{b}_{p\times 1} = \boldsymbol{X}^T\boldsymbol{Y},\ \boldsymbol{b}|\boldsymbol{\beta} \sim N(\boldsymbol{\beta}, 1\cdot \boldsymbol{I}_{p\times p})$$

- Marginal selection rule

$$M = \{i : B^{Crit} \leq |b_i|\}$$

Full model estimation error

Average l_2 risk for estimators of the form $\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$, $\hat{\beta}_i = \hat{\beta}_i(b_i)$:

$$
\begin{aligned}
E_{\boldsymbol{\beta},\boldsymbol{b}} \sum_{i=1}^{n} (\mu_i - \hat{\mu}_i)^2 &= E_{\boldsymbol{\beta},\boldsymbol{b}} \{ (\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T (\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \} \\
&= E_{\boldsymbol{\beta},\boldsymbol{b}} \{ (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \boldsymbol{X}^T \boldsymbol{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \} \\
&= E_{\boldsymbol{\beta},\boldsymbol{b}} \{ (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \} \\
&= E_{\boldsymbol{\beta},\boldsymbol{b}} \sum_{i=1}^{p} (\beta_i - \hat{\beta}_i)^2 \\
&= p \cdot E_{\beta_i,b_i} \, (\beta_i - \hat{\beta}_i)^2 \qquad (7) \\
&= p \cdot E_{b_i} \, E_{\beta_i|b_i} \, (\beta_i - \hat{\beta}_i)^2
\end{aligned}
$$

Three full model estimators

- $\hat{\beta}_i^{Bayes} = E_{\beta_i|b_i}\beta_i$ the Bayes estimator with marginal average risk:

$$E_{b_i}\ \{E_{\beta_i|b_i}(\beta_i - E_{\beta_i|b_i}\beta_i)^2\} = E_{b_i} Var(\beta_i|b_i)$$

- $\hat{\beta}_i^{MLE} = b_i$ the MLE with marginal average risk:

$$E_{\beta_i}\ E_{b_i|\beta_i}(\beta_i - b_i)^2 = 1$$

- $\hat{\beta}_i^{Thrsh} = I(B^{Crit} \leq |b_i|) \cdot b_i$ the hard thresholding estimator with ... :

$$E_{b_i}\ \{I(|b_i| < B^{Crit}) \cdot E_{\beta_i|b_i}\beta_i^2 + I(B^{Crit} \leq |b_i|) \cdot E_{\beta_i|b_i}(\beta_i - b_i)^2\}$$

Model selection

- Model M is selected when the following occurs

$$S_{\Omega}^{M} = \{\, \boldsymbol{y} : B^{Crit} \leq b_i \ \ \forall i \in M,\ b_i < B^{Crit} \ \ \forall i \notin M \,\}$$

- The corresponding X-matrix is $\boldsymbol{X}^M = \{X_i : \ i \in M\}$
- As $\boldsymbol{X}$ is orthomormal the target of estimation is

$$\boldsymbol{\beta}_M = ((\mathbf{X}^M)^T \mathbf{X}^M)^{-1} (\mathbf{X}^M)^T \boldsymbol{X}\boldsymbol{\beta} = \{\beta_i : i \in M\}$$

that can also be expressed $\boldsymbol{\mu}^M = \boldsymbol{X}^M \boldsymbol{\beta}_M$

Errors in model selection

1. *Submodel selection error:* a bias term quantifying the loss incurred by selecting a submodel

$$E\|\boldsymbol{\mu} - \boldsymbol{\mu}^M\|^2$$

$$(= \; p \cdot E_{b_i} \; \{I(|b_i| < B^{Crit}) \cdot E_{\beta_i|b_i}(\beta_i)^2 \; \} \;)$$

2. *Submodel estimation error:* a variance term quantifying how difficult it is to estimate the selected model

$$E_{f_S(\boldsymbol{\beta},\boldsymbol{b})}\|\boldsymbol{\mu}^M - \hat{\boldsymbol{\mu}}^M\|^2$$

(this is of course the selective inference problem!!)

Random parameter truncated distribution of $(\boldsymbol{\beta}, \boldsymbol{b})$

$$
\begin{aligned}
f_S(\boldsymbol{\beta}, \boldsymbol{b}) \quad = \quad & \Pi_{i \in M} \frac{I(B^{Crit} \leq |b_i|) \cdot \pi(\beta_i) \cdot f(b_i|\beta_i)}{\Pr(B^{Crit} \leq |b_i|)} \\
& \times \Pi_{i \notin M} \frac{I(|b_i| < B^{Crit}) \cdot \pi(\beta_i) \cdot f(b_i|\beta_i)}{\Pr(|b_i| < B^{Crit})}
\end{aligned}
$$

- Thus (β_i, b_i) independent
- For $i \in M$

$$
f_S(\beta_i, b_i) \propto \pi(\beta_i) \cdot f(b_i|\beta_i)
$$

Fixed parameter truncated distribution of $(\boldsymbol{\beta}, \boldsymbol{b})$

$$
\begin{aligned}
f_S(\boldsymbol{\beta}, \boldsymbol{b}) \quad = \quad & \Pi_{i \in M} \frac{I(B^{Crit} \leq |b_i|) \cdot \pi(\beta_i) \cdot f(b_i|\beta_i)}{\Pr(B^{Crit} \leq |b_i| \mid \beta_i)} \\
& \times \Pi_{i \notin M} \frac{I(|b_i| < B^{Crit}) \cdot \pi(\beta_i) \cdot f(b_i|\beta_i)}{\Pr(|b_i| < B^{Crit} \mid \beta_i)}
\end{aligned}
$$

- Thus (β_i, b_i) independent
- For $i \in M$

$$
f_S(\beta_i, b_i) \propto \pi(\beta_i) \cdot f(b_i|\beta_i) / \Pr(B^{Crit} \leq |b_i| \mid \beta_i)
$$

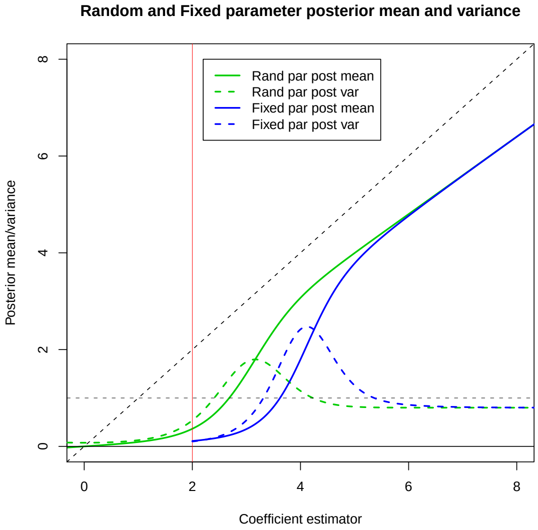
Assessing the average risk in submodel

$$
\begin{aligned}
& E_{f_S(\boldsymbol{\beta},\boldsymbol{b})} \sum_{i=1}^{n} (\mu_i^M - \hat{\mu}_i^M)^2 \\
&= E_{f_S(\boldsymbol{\beta},\boldsymbol{b})} \{ (\boldsymbol{X}^M \boldsymbol{\beta}_M - \boldsymbol{X}^M \hat{\boldsymbol{\beta}}_M)^T (\boldsymbol{X}^M \boldsymbol{\beta}_M - \boldsymbol{X}^M \hat{\boldsymbol{\beta}}_M) \} \\
&= E_{f_S(\boldsymbol{\beta},\boldsymbol{b})} \{ (\boldsymbol{\beta}_M - \hat{\boldsymbol{\beta}}_M)^T (\boldsymbol{X}^M)^T \boldsymbol{X}^M (\boldsymbol{\beta}_M - \hat{\boldsymbol{\beta}}_M) \} \\
&= E_{f_S(\boldsymbol{\beta},\boldsymbol{b})} \{ (\boldsymbol{\beta}_M - \hat{\boldsymbol{\beta}}_M)^T (\boldsymbol{\beta}_M - \hat{\boldsymbol{\beta}}_M) \} \\
&= E_{f_S(\boldsymbol{\beta},\boldsymbol{b})} \sum_{i \in M} (\beta_i - \hat{\beta}_i)^2 = \sum_{i \in M} E_{f_S(\beta_i, b_i)} (\beta_i - \hat{\beta}_i)^2
\end{aligned}
$$

Bayes estimator $\tilde{\boldsymbol{\beta}}_M = \{ \tilde{\beta}_i = E_{\pi_S(\beta_i |\, b_i)} \beta_i : \ i \in M \}$ with average risk

$$
|M| \cdot E_{f_S(b_i)} \{ E_{\pi_S(\beta_i |\, b_i)} (\beta_i)^2 - (E_{\pi_S(\beta_i |\, b_i)} (\beta_i - \hat{\beta}_i))^2 \}
$$

Marginal saBayes posterior mean and variance for $B^{Crit} = 2$



Full model estimation error for $p = 20$

$$\beta_i \sim 0.9 \cdot N(0, 0.2^2) + 0.1 \cdot N(0, 2^2)$$

- Mean marginal effect size: $E\beta_i^2 = 0.436$
- Mean effect size is $E\|\boldsymbol{\beta}\|^2 = 20 \cdot 0.436 = 8.7$
- Mean marginal estimation error: $E(\beta_i - \hat{\beta}_i^{Bayes})^2 = 0.210$
- Relative estimation error is $0.482 = 0.210/0.436$

Submodel selection and estimation error for $p = 20$

- Marginal model selection error $E_{b_i}\,\{I(|b_i| < B^{Crit}) \cdot E_{\beta_i|b_i}(\beta_i)^2\} = 0.132$ thus mean submodel effect-size $E\|\mu^M\|^2 = 20 \cdot (0.436 - 0.132) = 6.08$
- Marginal selection probability $\Pr(B^{Crit} \leq |b_i|) = 0.082$ thus mean number of selected parameters $E|M| = 20 \cdot 0.082 = 1.64$
- Random β marginal submodel estimation error is $E_{f_S(\beta_i,b)}(\beta_i - \tilde{\beta}_i^{Bayes})^2 = 1.065$, thus the submodel estimation error $1.747 = 1.65 \cdot 1.065$
- Fixed β marginal submodel estimation error is $E_{f_S(\beta_i,b)}(\beta_i - \tilde{\beta}_i^{Bayes})^2 = 0.261$, thus the submodel estimation error $0.42 = 1.65 \cdot 0.261$
- Thus relative estimation errors are $1.747/6.08$ and $0.42/6.08$

Some final points

- Our Bayesian POSI provides comprehensive selection-adjusted inference
- Our selection adjustment generally much smaller than that of Berk et al. ‘12.
- In particular, if the selected model is strongly supported then the selection adjustment is negligible.
- Work in progress ...

References

- Benjamini, Y., Yekutieli, D., (2005) "False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Parameters" *JASA*, **100**, 71-81.
- Berk, R., Brown, L. D., Buja, A., Zhang, K., Zhao, L. (2012) "Valid Post-Selection Inference", can be downloaded from authors' website.
- Tukey J. W. (1953), *The Problem of Multiple Comparisons*, Mimeographed monograph.
- Weinstein, A., Fithian, W., Benjamini, Y., (2012) "Selection Adjusted Confidence Intervals with More Power to Determine the Sign" *JASA* DOI: 10.1080/01621459.2012.737740
- Yekutieli D. (2012) "Adjusted Bayesian inference for selected parameters" *JRSS B*, **74**, 3, 515 - 541.