Bayesian Model Averaging

Hoff Chapter 9, Liang et al 2007, Hoeting et al (1999), Clyde & George (2004) Statistical Science

October 26, 2020

Prior Distributions

- Bayesian Model choice requires proper prior distributions on parameters that are not common across models

Prior Distributions

- Bayesian Model choice requires proper prior distributions on parameters that are not common across models
- Vague but proper priors may lead to paradoxes!

Prior Distributions

- Bayesian Model choice requires proper prior distributions on parameters that are not common across models
- Vague but proper priors may lead to paradoxes!
- Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner's g-prior is the most popular.

Prior Distributions

- Bayesian Model choice requires proper prior distributions on parameters that are not common across models
- Vague but proper priors may lead to paradoxes!
- Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner's g-prior is the most popular.

Prior Distributions

- "Spike and Slab" - Lempers (1971) Mitchell & Beauchamp (1988)

Prior Distributions

- "Spike and Slab" - Lempers (1971) Mitchell & Beauchamp (1988)
- "Spike and Bell" Leamer (1978) in BMA

Prior Distributions

- "Spike and Slab" - Lempers (1971) Mitchell & Beauchamp (1988)
- "Spike and Bell" Leamer (1978) in BMA
- mixture of 2 normals - concentrated and dispersed - SSVS Gibbs Sampler - George & McCulloch (1993)

Prior Distributions

- “Spike and Slab” - Lempers (1971) Mitchell & Beauchamp (1988)
- “Spike and Bell” Leamer (1978) in BMA
- mixture of 2 normals - concentrated and dispersed - SSVS Gibbs Sampler - George & McCulloch (1993)
- Back to “Spike and Bell” Hoeting, Raftery & Madigan MC^3 (1997) and George & McCulloch (1997) collapsed MCMC after integrating out β_γ
- Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner’s g-prior is the most popular.

Prior Distributions

- "Spike and Slab" - Lempers (1971) Mitchell & Beauchamp (1988)
- "Spike and Bell" Leamer (1978) in BMA
- mixture of 2 normals - concentrated and dispersed - SSVS Gibbs Sampler - George & McCulloch (1993)
- Back to "Spike and Bell" Hoeting, Raftery & Madigan MC^3 (1997) and George & McCulloch (1997) collapsed MCMC after integrating out $\boldsymbol{\beta}_\gamma$
- Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner's g-prior is the most popular.

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha) \propto 1$

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha) \propto 1$
- $p(\sigma^2) \propto 1/\sigma^2$

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha) \propto 1$
- $p(\sigma^2) \propto 1/\sigma^2$
- $\boldsymbol{\beta}_\gamma \mid \alpha, \sigma^2, \boldsymbol{\gamma} \sim \mathsf{N}(0, g\sigma^2(\mathbf{X}_\gamma^{c\,\prime}\mathbf{X}_\gamma^c)^{-1})$

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha) \propto 1$
- $p(\sigma^2) \propto 1/\sigma^2$
- $\boldsymbol{\beta}_\gamma \mid \alpha, \sigma^2, \boldsymbol{\gamma} \sim \mathsf{N}(0, g\sigma^2(\mathbf{X}_\gamma^{c\,\prime}\mathbf{X}_\gamma^c)^{-1})$

which leads to marginal likelihood of $\mathcal{M}_\gamma$ that is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = C(1+g)^{\frac{n-p-1}{2}}(1 + g(1 - R^2_\gamma))^{-\frac{(n-1)}{2}}$$

where R^2 is the usual coefficient of determination for model $\mathcal{M}_\gamma$.

Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}^c \boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}^c$ is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha) \propto 1$
- $p(\sigma^2) \propto 1/\sigma^2$
- $\boldsymbol{\beta}_\gamma \mid \alpha, \sigma^2, \boldsymbol{\gamma} \sim \mathrm{N}(0, g\sigma^2(\mathbf{X}_\gamma^{c\prime}\mathbf{X}_\gamma^c)^{-1})$

which leads to marginal likelihood of $\mathcal{M}_\gamma$ that is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = C(1+g)^{\frac{n-p-1}{2}}(1+g(1-R_\gamma^2))^{-\frac{(n-1)}{2}}$$

where R^2 is the usual coefficient of determination for model $\mathcal{M}_\gamma$.
Trade-off of model complexity versus goodness of fit

Lastly, assign distribution to space of models

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma})$

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma})$
- Integrate out intercept (normal)

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma})$
- Integrate out intercept (normal)
- Integrate out ϕ (gamma)

Sketch

- Integrate out β_γ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma})$
- Integrate out intercept (normal)
- Integrate out ϕ (gamma)
- algebra to simplify in from quadratic forms to R^2_γ

Priors on Model Space

$p(\mathcal{M}_\gamma) \Leftrightarrow p(\boldsymbol{\gamma})$

- $p(\gamma_j = 1) = .5 \Rightarrow P(\mathcal{M}_\gamma) = .5^p$ Uniform on space of models

Priors on Model Space

$p(\mathcal{M}_\gamma) \Leftrightarrow p(\boldsymbol{\gamma})$

- $p(\gamma_j = 1) = .5 \Rightarrow P(\mathcal{M}_\gamma) = .5^p$ Uniform on space of models
  $p_{\boldsymbol{\gamma}} \sim \text{Bin}(p, .5)$
- $\gamma_j \mid \pi \overset{\text{iid}}{\sim} \text{Ber}(\pi)$ and $\pi \sim \text{Beta}(a, b)$ then $p_{\boldsymbol{\gamma}} \sim \text{BB}_p(a, b)$

$$p(p_{\boldsymbol{\gamma}} \mid p, a, b) = \frac{\Gamma(p+1)\Gamma(p_{\boldsymbol{\gamma}} + a)\Gamma(p - p_{\boldsymbol{\gamma}} + b)\Gamma(a+b)}{\Gamma(p_{\boldsymbol{\gamma}} + 1)\Gamma(p - p_{\boldsymbol{\gamma}} + 1)\Gamma(p + a + b)\Gamma(a)\Gamma(b)}$$

- $p_{\boldsymbol{\gamma}} \sim \text{BB}_p(1, 1) \sim \text{Unif}(0, p)$

USair Data

```r
library(BAS)
data(usair, package="HH")
poll.bma = bas.lm(log(SO2) ~ temp + log(mfgfirms) +
                             log(popn) + wind +
                             precip + raindays,
                  data=usair,
                  prior="g-prior",
                  alpha=nrow(usair), # g = n
                  n.models=2^6,
                  modelprior = uniform(),
                  method="deterministic")
```

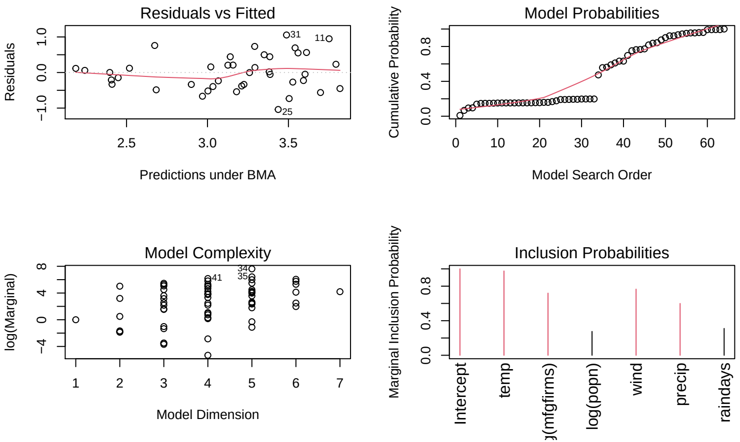
Summary

```
poll.bma

##
## Call:
## bas.lm(formula = log(SO2) ~ temp + log(mfgfirms) + log(popn)
##     wind + precip + raindays, data = usair, n.models = 2^6, p
##     alpha = nrow(usair), modelprior = uniform(), method = "de
##
##
##  Marginal Posterior Inclusion Probabilities:
##     Intercept          temp  log(mfgfirms)      log(popn)
##        1.0000        0.9755         0.7190         0.2757
##        precip      raindays
##        0.5994        0.3104
```

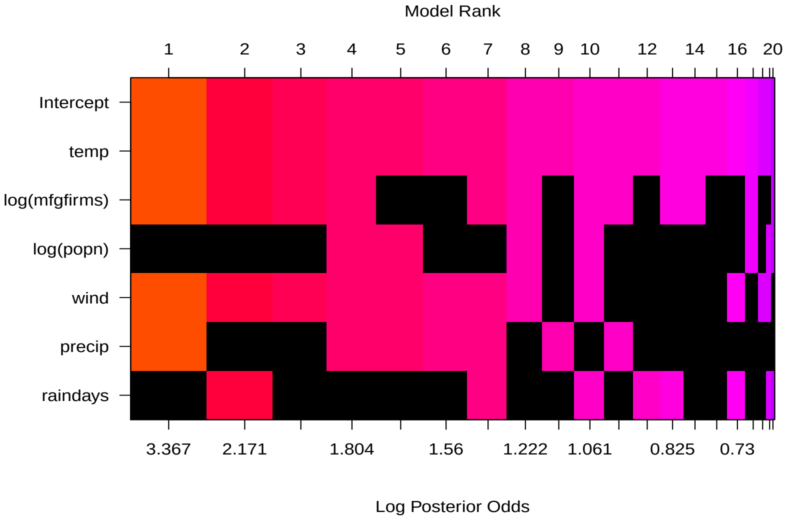
Plots

```r
par(mfrow=c(2,2))
plot(poll.bma, ask=F)
```



Posterior Distribution with Uniform Prior on Model Space

```r
image(poll.bma, rotate=FALSE)
```

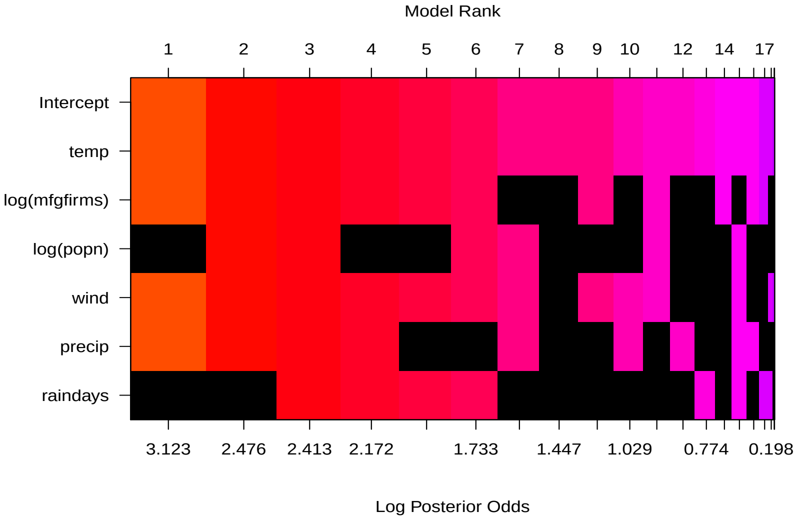


Posterior Distribution with BB(1,1) Prior on Model Space

```
poll.bb.bma = bas.lm(log(SO2) ~ temp + log(mfgfirms) +
                                log(popn) + wind +
                                precip + raindays,
                      data=usair,
                      prior="g-prior",
                      alpha=nrow(usair),
                      n.models=2^6,  #enumerate
                      modelprior=beta.binomial(1,1))
```

BB(1,1) Prior on Model Space

```
image(poll.bb.bma, rotate=FALSE)
```



Jeffreys Scale of Evidence

- Bayes Factor = ratio of marginal likelihoods

Jeffreys Scale of Evidence

- Bayes Factor $=$ ratio of marginal likelihoods
- Posterior odds $=$ Bayes Factor $\times$ Prior odds

Jeffreys Scale of Evidence

- Bayes Factor $=$ ratio of marginal likelihoods
- Posterior odds $=$ Bayes Factor $\times$ Prior odds

$B = BF[\mathcal{M}_0 : \mathcal{M}_\gamma]$ and $1/B = BF[\mathcal{M}_\gamma : \mathcal{M}_0]$

Jeffreys Scale of Evidence

- Bayes Factor = ratio of marginal likelihoods
- Posterior odds = Bayes Factor × Prior odds

$B = BF[\mathcal{M}_0 : \mathcal{M}_\gamma]$ and $1/B = BF[\mathcal{M}_\gamma : \mathcal{M}_0]$

Bayes Factor	Interpretation
$B \geq 1$	H_0 supported
$1 > B \geq 10^{-\frac{1}{2}}$	minimal evidence against H_0
$10^{-\frac{1}{2}} > B \geq 10^{-1}$	substantial evidence against H_0
$10^{-1} > B \geq 10^{-2}$	strong evidence against H_0
$10^{-2} > B$	decisive evidence against H_0

Jeffreys Scale of Evidence

- Bayes Factor $=$ ratio of marginal likelihoods
- Posterior odds $=$ Bayes Factor $\times$ Prior odds

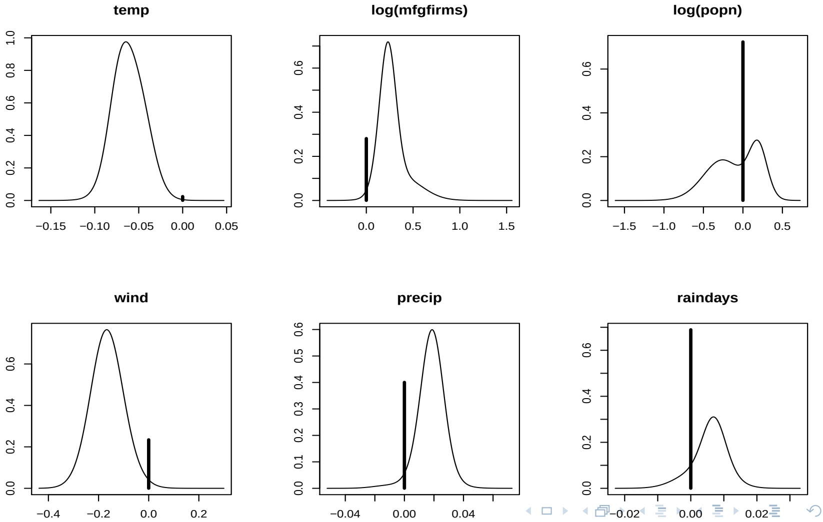
$B = BF[\mathcal{M}_0 : \mathcal{M}_\gamma]$ and $1/B = BF[\mathcal{M}_\gamma : \mathcal{M}_0]$

Bayes Factor	Interpretation
$B \geq 1$	H_0 supported
$1 > B \geq 10^{-\frac{1}{2}}$	minimal evidence against H_0
$10^{-\frac{1}{2}} > B \geq 10^{-1}$	substantial evidence against H_0
$10^{-1} > B \geq 10^{-2}$	strong evidence against H_0
$10^{-2} > B$	decisive evidence against H_0

in context of testing one hypothesis with equal prior odds Kass & Raftery (JASA 1996)

Coefficients

```
beta = coef(poll.bma)
par(mfrow=c(2,3));  plot(beta, subset=2:7,ask=F)
```



Bartlett's Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

Bartlett's Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- For fixed sample size n and R^2, consider taking values of g that go to infinity

Bartlett's Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- For fixed sample size n and R^2, consider taking values of g that go to infinity
- Increasing vagueness in prior

Bartlett's Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- For fixed sample size n and R^2, consider taking values of g that go to infinity
- Increasing vagueness in prior
- What happens to BF as $g \to \infty$?

Information Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

Information Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathbf{\mathcal{M}}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- Let g be a fixed constant and take n fixed.

Information Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathcal{M}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- Let g be a fixed constant and take n fixed.
- Let $F = \frac{R^2_\gamma/p_\gamma}{(1-R^2_\gamma)/(n-1-p_\gamma)}$

Information Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathbf{\mathcal{M}}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- Let g be a fixed constant and take n fixed.
- Let $F = \frac{R^2_\gamma / p_\gamma}{(1-R^2_\gamma)/(n-1-p_\gamma)}$
- As $R^2_\gamma \to 1$, $F \to \infty$ LR test would reject $\mathbf{\mathcal{M}}_0$ where F is the usual F statistic for comparing model $\mathcal{M}_\gamma$ to $\mathbf{\mathcal{M}}_0$

Information Paradox

The Bayes factor for comparing $\mathcal{M}_\gamma$ to the null model:

$$BF(\mathcal{M}_\gamma : \mathbf{\mathcal{M}}_0) = (1+g)^{(n-1-p_\gamma)/2}(1+g(1-R^2))^{-(n-1)/2}$$

- Let g be a fixed constant and take n fixed.
- Let $F = \frac{R^2_\gamma / p_\gamma}{(1-R^2_\gamma)/(n-1-p_\gamma)}$
- As $R^2_\gamma \to 1$, $F \to \infty$ LR test would reject $\mathbf{\mathcal{M}}_0$ where F is the usual F statistic for comparing model $\mathcal{M}_\gamma$ to $\mathbf{\mathcal{M}}_0$
- BF converges to a fixed constant $(1+g)^{-p_\gamma/2}$ (does not go to infinity

"Information Inconsistency" see Liang et al JASA 2008

Mixtures of g priors & Information consistency

Need $BF \to \infty$ if $\mathrm{R}^2 \to 1 \Leftrightarrow \mathrm{E}_g[(1+g)^{-p_\gamma/2}]$ diverges for $p_\gamma < n-1$ (proof in Liang et al)

Mixtures of g priors & Information consistency

Need $BF \to \infty$ if $\mathrm{R}^2 \to 1 \Leftrightarrow \mathrm{E}_g[(1+g)^{-p_\gamma/2}]$ diverges for $p_\gamma < n-1$ (proof in Liang et al)

- Zellner-Siow Cauchy prior

Mixtures of g priors & Information consistency

Need $BF \to \infty$ if $\mathrm{R}^2 \to 1 \Leftrightarrow \mathrm{E}_g[(1+g)^{-p_\gamma/2}]$ diverges for $p_\gamma < n-1$ (proof in Liang et al)

- Zellner-Siow Cauchy prior
- hyper-g prior or hyper-g/n (Liang et al JASA 2008)

Mixtures of g priors & Information consistency

Need $BF \to \infty$ if $\mathrm{R}^2 \to 1 \Leftrightarrow \mathrm{E}_g[(1+g)^{-p_\gamma/2}]$ diverges for $p_\gamma < n-1$ (proof in Liang et al)

- Zellner-Siow Cauchy prior
- hyper-g prior or hyper-g/n (Liang et al JASA 2008)
- robust prior (Bayarrri et al Annals of Statistics 2012

Mixtures of g priors & Information consistency

Need $BF \to \infty$ if $\mathrm{R}^2 \to 1 \Leftrightarrow \mathrm{E}_g[(1+g)^{-p_\gamma/2}]$ diverges for $p_\gamma < n-1$ (proof in Liang et al)

- Zellner-Siow Cauchy prior
- hyper-g prior or hyper-g/n (Liang et al JASA 2008)
- robust prior (Bayarrri et al Annals of Statistics 2012

All have tails that behave like a Cauchy distribution

Mortality & Pollution

- Data 60 cities from Statistical Sleuth 12.17

Mortality & Pollution

- Data 60 cities from Statistical Sleuth 12.17
- response Mortality

Mortality & Pollution

- Data 60 cities from Statistical Sleuth 12.17
- response Mortality
- 15 predictors; measures of HC, NOX, SO2

Mortality & Pollution

- Data 60 cities from Statistical Sleuth 12.17
- response Mortality
- 15 predictors; measures of HC, NOX, SO2
- Is pollution associated with mortality after adjusting for other socio-economic and meteorological factors?

Mortality & Pollution

- Data 60 cities from Statistical Sleuth 12.17
- response Mortality
- 15 predictors; measures of HC, NOX, SO2
- Is pollution associated with mortality after adjusting for other socio-economic and meteorological factors?
- 15 predictor variables implies $2^{15} = 32,768$ possible models

```r
data(ex1217, package="Sleuth3")
suppressWarnings(library(dplyr))
mortality = mutate(ex1217,
                   logHC = log(HC),
                   logNOX = log(NOX),
                   logSO2 = log(SO2)) %>%
            select(-CITY, -HC, -NOX, -SO2)
```

Jeffreys Zellner-Siow Cauchy Prior

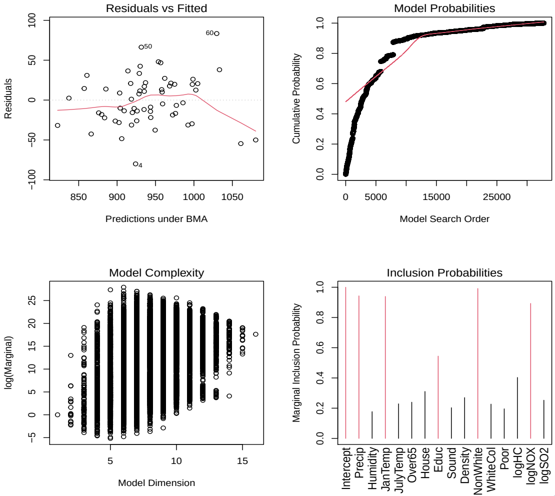
- Jeffreys "reference" prior on α and σ^2
- Zellner-Siow Cauchy prior

$$
\begin{aligned}
1/g &\sim G(1/2, n/2) \\
\boldsymbol{\beta}_\gamma \mid g, \sigma^2 &\sim N(0, g\sigma^2(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)^{-1}) \\
\Rightarrow \boldsymbol{\beta}_\gamma \mid \sigma^2 &\sim C(0, \sigma^2(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)^{-1})
\end{aligned}
$$

```
mort.bma = bas.lm(Mortality ~ ., data=mortality,
                  prior="JZS",
                  alpha=1,
                  n.models=2^15,
                  initprobs="eplogp",
                  method='BAS')
```

Posterior Plots

```
par(mfrow=c(2,2))
plot(mort.bma, ask=FALSE)
```



Posterior Probabilities

- What is the probability that there is no pollution effect?

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

```
models = list2matrix.which(mort.bma)
poll.inclusion = (models[, 14:16] %*% rep(1, 3)) > 0
prob.poll = sum(poll.inclusion * mort.bma$postprobs)
prob.poll

## [1] 0.9829953
```

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

```
models = list2matrix.which(mort.bma)
poll.inclusion = (models[, 14:16] %*% rep(1, 3)) > 0
prob.poll = sum(poll.inclusion * mort.bma$postprobs)
prob.poll
## [1] 0.9829953
```

- Posterior probability no effect is 0.017

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

```
models = list2matrix.which(mort.bma)
poll.inclusion = (models[, 14:16] %*% rep(1, 3)) > 0
prob.poll = sum(poll.inclusion * mort.bma$postprobs)
prob.poll
## [1] 0.9829953
```

- Posterior probability no effect is 0.017
- Odds that there is an effect $0.983/0.017 = 57.8073$

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

```
models = list2matrix.which(mort.bma)
poll.inclusion = (models[, 14:16] %*% rep(1, 3)) > 0
prob.poll = sum(poll.inclusion * mort.bma$postprobs)
prob.poll

## [1] 0.9829953
```

- Posterior probability no effect is 0.017
- Odds that there is an effect $0.983/0.017 = 57.8073$
- Prior Odds $7 = (1 - .5^3)/.5^3$

Posterior Probabilities

- What is the probability that there is no pollution effect?
- Sum posterior model probabilities over all models that include at least one pollution variable

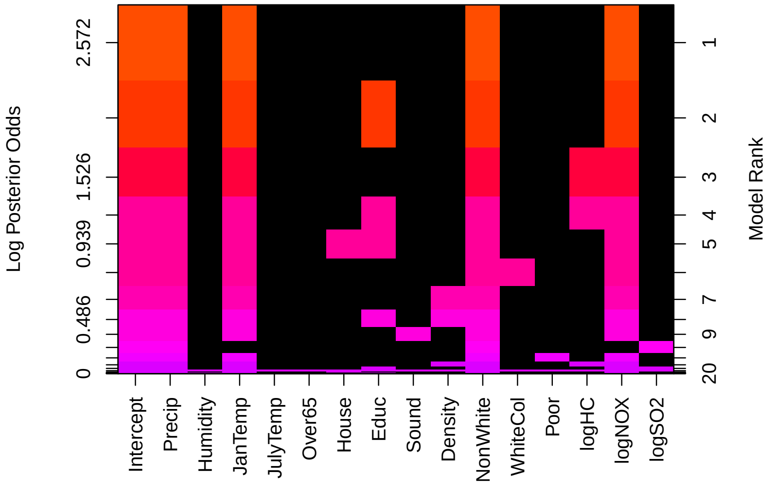
```
models = list2matrix.which(mort.bma)
poll.inclusion = (models[, 14:16] %*% rep(1, 3)) > 0
prob.poll = sum(poll.inclusion * mort.bma$postprobs)
prob.poll

## [1] 0.9829953
```

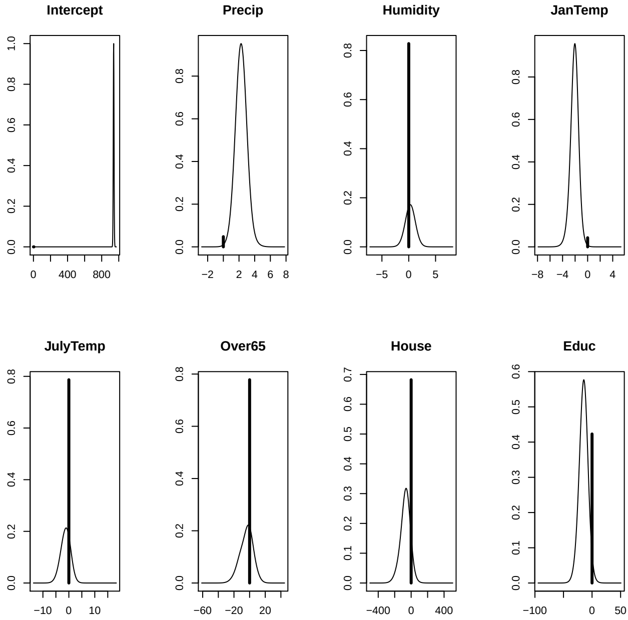
- Posterior probability no effect is 0.017
- Odds that there is an effect $0.983/0.017 = 57.8073$
- Prior Odds $7 = (1 - .5^3)/.5^3$

Model Space

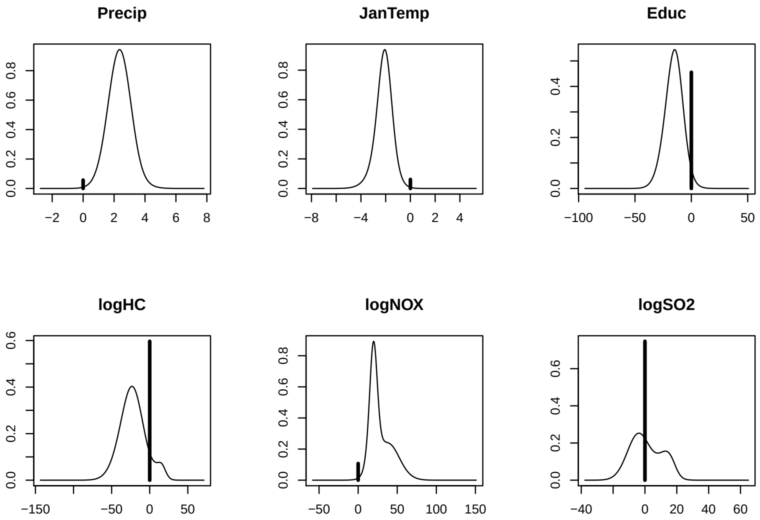
```r
image(mort.bma)
```



Coefficients



Coefficients



Effect Estimation

- Coefficients in each model are adjusted for other variables in the model
- OLS: leave out a predictor with a non-zero coefficient then estimates are biased!
- Model Selection in the presence of high correlation, may leave out "redundant" variables;
- improved MSE for prediction (Bias-variance tradeoff)
- in BMA all variables are included, but coefficients are shrunk to 0
- Care for "causal" questions and confounder adjustment!

Other Problems

- Computational

Other Problems

- Computational if $p > 35$ enumeration is difficult

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ
  - poor convergence/mixing with high correlations

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ
  - poor convergence/mixing with high correlations
  - Metropolis Hastings algorithms more flexibility

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ
  - poor convergence/mixing with high correlations
  - Metropolis Hastings algorithms more flexibility
  - "Stochastic Search" (no guarantee samples represent posterior)

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ
  - poor convergence/mixing with high correlations
  - Metropolis Hastings algorithms more flexibility
  - "Stochastic Search" (no guarantee samples represent posterior)
- Prior Choice: Choice of prior distributions on β and on γ

Other Problems

- Computational if $p > 35$ enumeration is difficult
  - Gibbs sampler or Random-Walk algorithm on γ
  - poor convergence/mixing with high correlations
  - Metropolis Hastings algorithms more flexibility
  - "Stochastic Search" (no guarantee samples represent posterior)
- Prior Choice: Choice of prior distributions on β and on γ

Model averaging versus Model Selection – what are objectives?