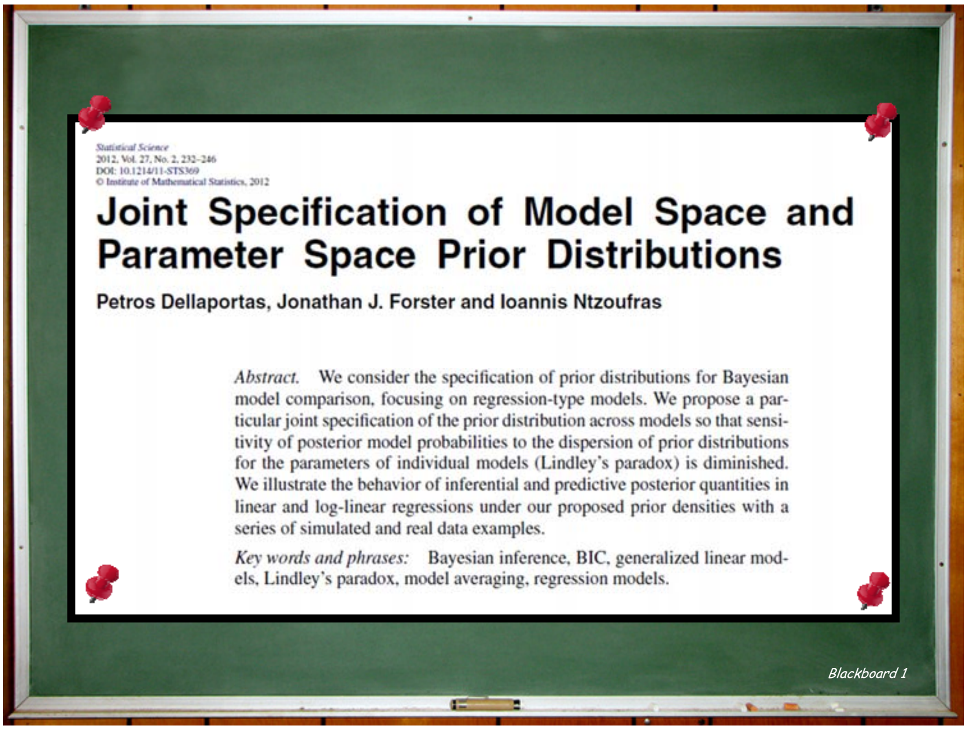


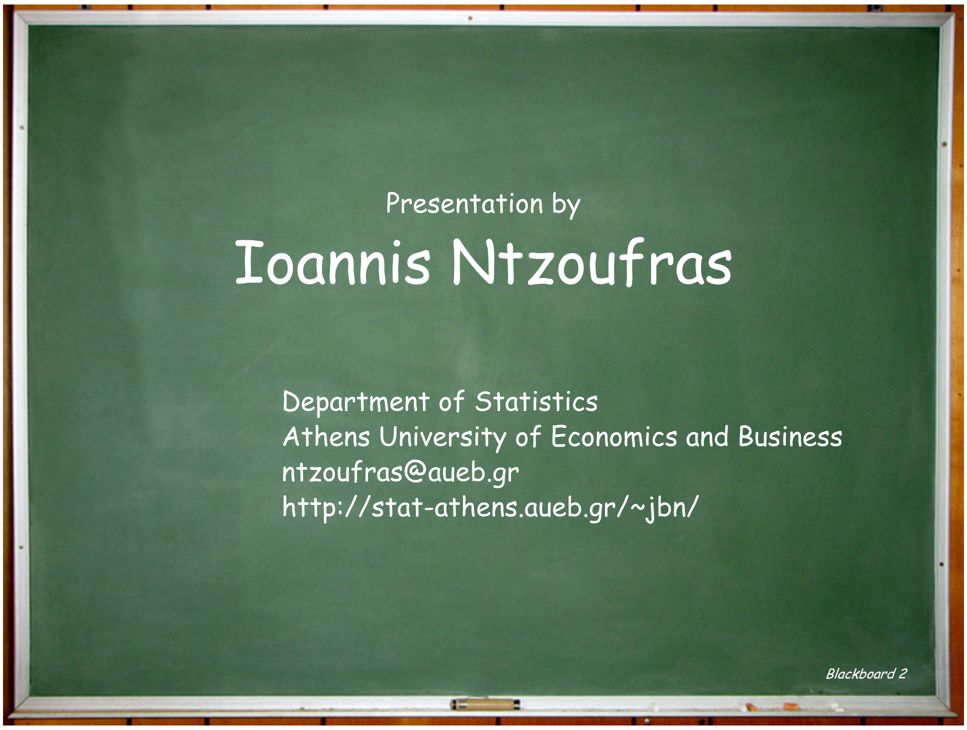
Statistical Science
2012, Vol. 27, No. 2, 232–246
DOI: 10.1214/11-STS369

Joint Specification of Model Space and Parameter Space Prior Distributions

Petros Dellaportas, Jonathan J. Forster and Ioannis Ntzoufras

Abstract. We consider the specification of prior distributions for Bayesian model comparison, focusing on regression-type models. We propose a particular joint specification of the prior distribution across models so that sensitivity of posterior model probabilities to the dispersion of prior distributions for the parameters of individual models (Lindley's paradox) is diminished. We illustrate the behavior of inferential and predictive posterior quantities in linear and log-linear regressions under our proposed prior densities with a series of simulated and real data examples.

Key words and phrases: Bayesian inference, BIC, generalized linear models, Lindley's paradox, model averaging, regression models.



Agenda

1. An Obsession: Model selection and the Paradox
2. An Idea: Avoiding (?) the paradox
3. Illustrations and comparisons
4. Conclusion
5. Break: At last time for coffee

Blackboard 3

1. An Obsession
Model selection and the Paradox

A Bayesian approach to inference under model uncertainty proceeds as follows.

Suppose

- data y generated by a model $m \in M$
- Each model specifies the distribution of y, $f(y|m, \beta_m)$
- *β_m is the parameter vector for model m.*
- *$f(m)$* is the prior probability of model *m*

Then posterior inference is based on *posterior model probabilities*

$$f(m|y) = \frac{f(m)f(y|m)}{\sum_{m \in M} f(m)f(y|m)}, \qquad m \in M$$

Blackboard 4

1. An Obsession
Model selection and the Paradox

or on *posterior model odds*

$$\frac{f(m_1|y)}{f(m_2|y)} = \frac{f(m_1)}{f(m_2)} \frac{f(y|m_1)}{f(y|m_2)}$$

Posterior model odds (PO_{12}) — Prior model odds — Bayes factor (BF_{12})

Usually inference is based on Bayes factors (BF) since a natural (?) choice is to assume that the two models under consideration are a-priori equal.

Blackboard 5

1. An Obsession
Model selection and the Paradox

The Lindley-Bartlett-Jeffreys Paradox (1)

For a single model inference,

a highly diffuse prior on the model parameters is often used (to represent ignorance).

Then the posterior density takes the shape of the likelihood and is insensitive to the exact value of the prior density function, provided that the prior is relatively flat over the range of parameter values with non-negligible likelihood.

Blackboard 6

1. An Obsession
Model selection and the Paradox

The Lindley-Bartlett-Jeffreys Paradox (2)

For multiple models inference:

The use of such a prior creates an apparent difficulty.

For illustration, let us consider the simple case where model m_1 is completely specified (no unknown parameters) and model m_2 has parameter β_{m_2}

- *Then, for any observed data y, BF_{12} can be made arbitrarily large by choosing a sufficiently diffuse prior distribution for β_{m_2}*
- *Hence, under model uncertainty, two different diffuse prior distributions for model parameters might lead to essentially the same posterior distributions for those parameters, but very different BFs.*

Blackboard 7

1. An Obsession
Model selection and the Paradox

The Lindley-Bartlett-Jeffreys Paradox (3)

Therefore

$$BF_{12} \rightarrow \infty$$

when the Prior variance of $\beta_{m_2} \rightarrow \infty$

whatever data y we have…

Fully supporting the simpler model and making the procedure informative (?)

Blackboard 8

1. An Obsession
Model selection and the Paradox

The Lindley-Bartlett-Jeffreys Paradox (4)

Discussed by

- *Lindley (1957, Bka) ; referred to as `Lindley's paradox'*

  he actually noted the sensitivity of BF on the sample size and not on the prior

- *it is also variously attributed to Bartlett (1957, Bka) and*

  he also added the sensitivity on the prior variance in a note complementary to the publication of Lindley (1957); published in the next issue of Bka.

- *Also discussed by Jeffreys in his book*

As you can understand this became my obsession (and of many others). The aim was to overcome this paradoxical behavior…

Blackboard 9

1. An Obsession
Model selection and the Paradox

The Lindley-Bartlett-Jeffreys Paradox (5)

Dawid (2011)

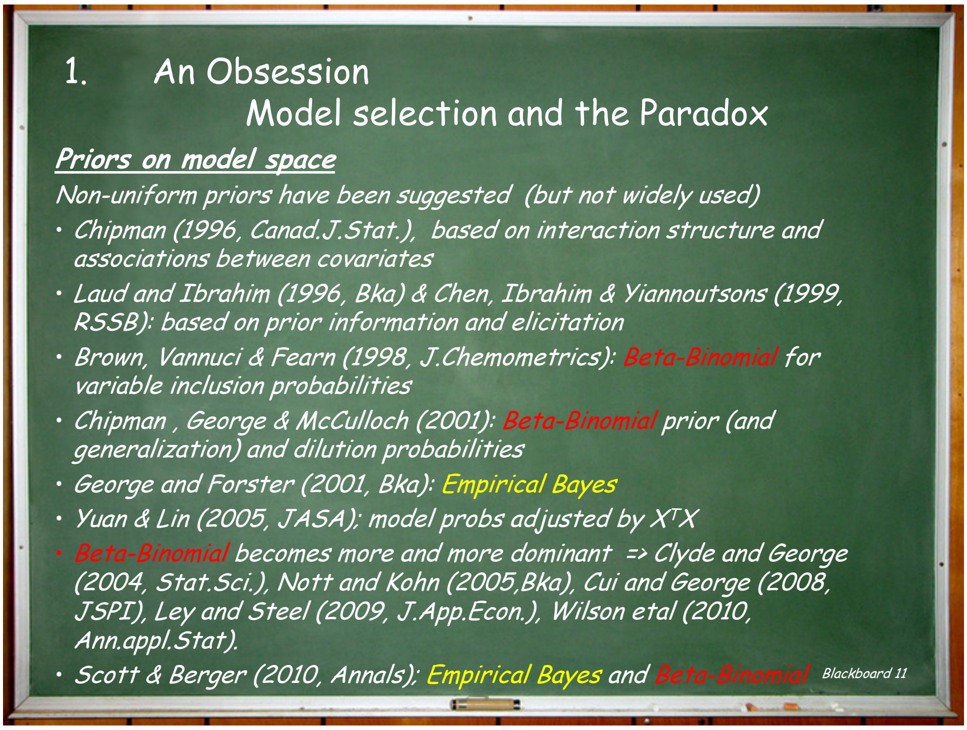
==> the Bayes factor is only one of the two elements on the posterior model odds.

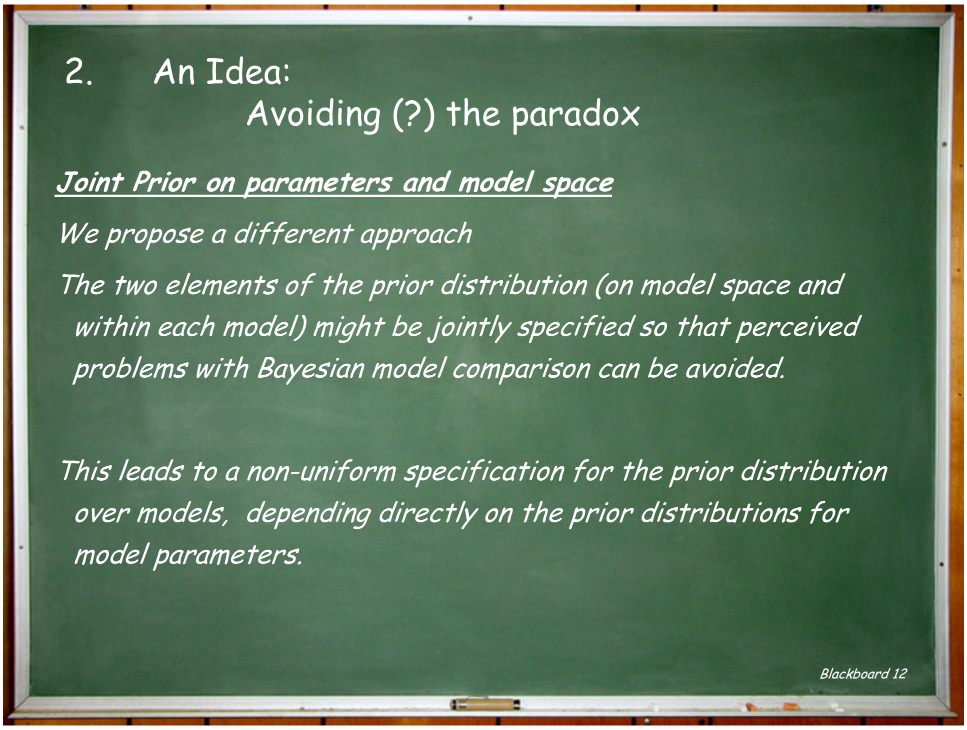
==> The prior model probabilities are of equal significance.

By focusing on the impact of the prior distributions for model parameters on the Bayes factor, there is an implicit understanding that the prior model probabilities are specified independently of these prior distributions.

This is often the case in practice, where a uniform prior distribution over models is commonly adopted, as a reference position.

Blackboard 10


1. An Obsession
Model selection and the Paradox
Priors on model space
Non-uniform priors have been suggested (but not widely used)
• Chipman (1996, Canad.J.Stat.), based on interaction structure and associations between covariates
• Laud and Ibrahim (1996, Bka) & Chen, Ibrahim & Yiannoutsons (1999, RSSB): based on prior information and elicitation
• Brown, Vannuci & Fearn (1998, J.Chemometrics): Beta-Binomial for variable inclusion probabilities
• Chipman , George & McCulloch (2001): Beta-Binomial prior (and generalization) and dilution probabilities
• George and Forster (2001, Bka): Empirical Bayes
• Yuan & Lin (2005, JASA): model probs adjusted by X^TX
• Beta-Binomial becomes more and more dominant => Clyde and George (2004, Stat.Sci.), Nott and Kohn (2005,Bka), Cui and George (2008, JSPI), Ley and Steel (2009, J.App.Econ.), Wilson etal (2010, Ann.appl.Stat).
• Scott & Berger (2010, Annals): Empirical Bayes and Beta-Binomial
Blackboard 11


2. An Idea:
Avoiding (?) the paradox
Joint Prior on parameters and model space
We propose a different approach
The two elements of the prior distribution (on model space and within each model) might be jointly specified so that perceived problems with Bayesian model comparison can be avoided.
This leads to a non-uniform specification for the prior distribution over models, depending directly on the prior distributions for model parameters.
Blackboard 12

2. An Idea: Avoiding (?) the paradox

We focus on models in which

- *the parameters can be a-priori expressed by a multivariate normal prior density with mean* μ_{β_m} *and variance-covariance matrix* V_m
- *the likelihood is sufficiently regular for standard asymptotic results to apply.*

Linear regression models and GLMs are such models.

Blackboard 13

2. An Idea: Avoiding (?) the paradox

We rewrite the prior variance matrix as $V_m = c_m^2 \Sigma_m$

where

- c_m *is the scale of the prior dispersion*
- Σ_m *is a semi-positive matrix with a fixed volume* $|\Sigma_m|$

Then, the posterior is given by

$$f(m|y) \propto f(m)(2\pi)^{-d_m/2}|\Sigma_m|^{-1/2}c_m^{-d_m}$$
$$\times \int \exp\left(-\frac{1}{2c_m^2}(\beta_m - \mu_{\beta_m})^T \Sigma_m^{-1}(\beta_m - \mu_{\beta_m})\right) f(y|m,\beta_m)d\beta_m$$

[d_m stands for the dimension of β_m]

Blackboard 14

2. An Idea:
Avoiding (?) the paradox

and for suitably large c_m

$$f(m|y) \approx f(m)(2\pi)^{-d_m/2}|\Sigma_m|^{-1/2}c_m^{-d_m}\int f(y|m,\beta_m)d\beta_m.$$

[d_m stands for the dimension of β_m]

Hence, as c_m gets larger, $f(m|y)$ gets smaller, assuming everything else remains fixed.

Therefore, for two models of different dimension and equal $c_m=c$, the posterior odds in favor of the more complex model tends to zero as c_m gets larger.

This is essentially the Lindley-Bartlett-Jeffreys paradox.

Blackboard 15

2. An Idea:
Avoiding (?) the paradox

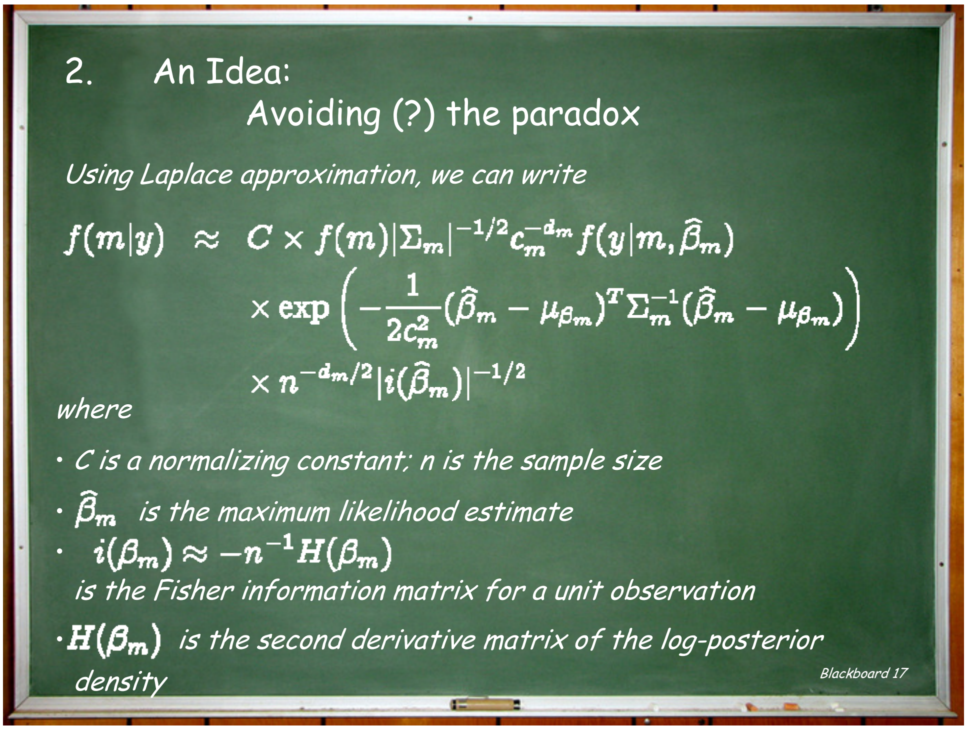
Using Laplace approximation, we can write

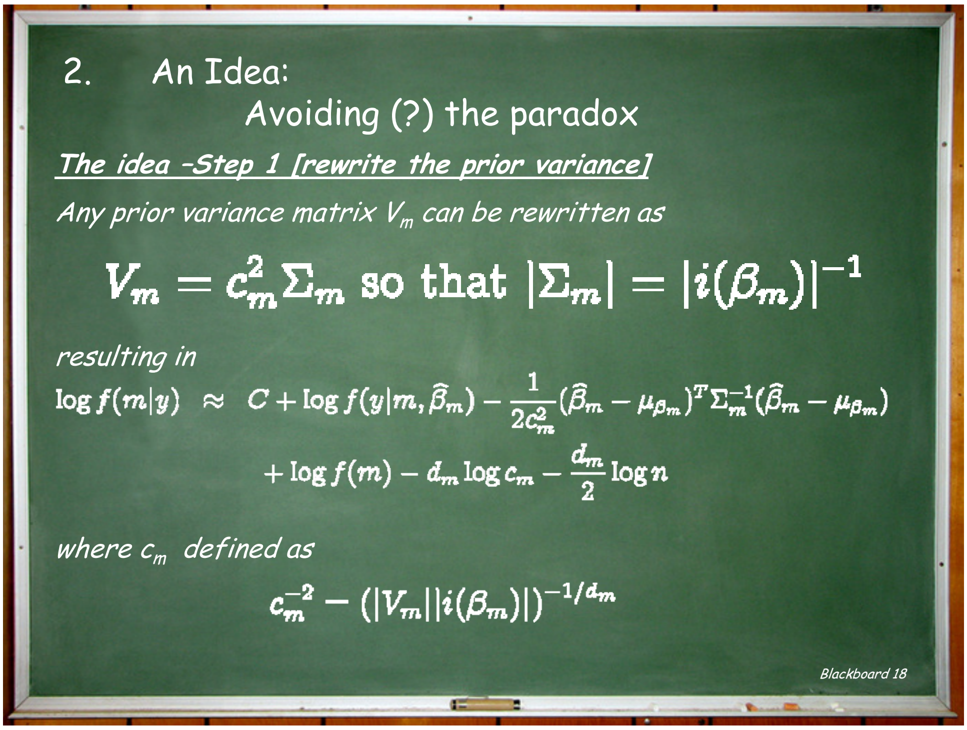
$$\begin{aligned} f(m|y) \;\approx\; & C \times f(m)|\Sigma_m|^{-1/2}c_m^{-d_m}f(y|m,\widehat{\beta}_m) \\ & \times \exp\left(-\frac{1}{2c_m^2}(\widehat{\beta}_m-\mu_{\beta_m})^T\Sigma_m^{-1}(\widehat{\beta}_m-\mu_{\beta_m})\right) \\ & \times |c_m^{-2}\Sigma_m^{-1}-H(\widehat{\beta}_m)|^{-1/2} \end{aligned}$$

where

- *C is a normalizing constant;*
- *$\widehat{\beta}_m$ is the maximum likelihood estimate and*
- *$H(\beta_m)$ is the second derivative matrix of the log-posterior density*

Blackboard 16


2. An Idea:
Avoiding (?) the paradox
Using Laplace approximation, we can write
$$f(m|y) \approx C \times f(m)|\Sigma_m|^{-1/2} c_m^{-d_m} f(y|m, \widehat{\beta}_m) \times \exp\left(-\frac{1}{2c_m^2}(\widehat{\beta}_m - \mu_{\beta_m})^T \Sigma_m^{-1} (\widehat{\beta}_m - \mu_{\beta_m})\right) \times n^{-d_m/2} |i(\widehat{\beta}_m)|^{-1/2}$$
where
• C is a normalizing constant; n is the sample size
• $\widehat{\beta}_m$ is the maximum likelihood estimate
• $i(\beta_m) \approx -n^{-1} H(\beta_m)$ is the Fisher information matrix for a unit observation
• $H(\beta_m)$ is the second derivative matrix of the log-posterior density
Blackboard 17


2. An Idea:
Avoiding (?) the paradox
The idea -Step 1 [rewrite the prior variance]
Any prior variance matrix V_m can be rewritten as
$$V_m = c_m^2 \Sigma_m \text{ so that } |\Sigma_m| = |i(\beta_m)|^{-1}$$
resulting in
$$\log f(m|y) \approx C + \log f(y|m, \widehat{\beta}_m) - \frac{1}{2c_m^2}(\widehat{\beta}_m - \mu_{\beta_m})^T \Sigma_m^{-1} (\widehat{\beta}_m - \mu_{\beta_m}) + \log f(m) - d_m \log c_m - \frac{d_m}{2} \log n$$
where c_m defined as
$$c_m^{-2} - (|V_m||i(\beta_m)|)^{-1/d_m}$$
Blackboard 18

2. An Idea:
Avoiding (?) the paradox

The idea - Step 2 [express posterior probs as BIC and additional penalties]

Additional dimension penalty (1)

$$\log f(m|y) \approx C + \overbrace{\log f(y|m,\widehat{\beta}_m) - \frac{d_m}{2}\log n}^{\text{BIC}} - \frac{d_m}{2}\log c_m^2$$
$$-\frac{1}{2c_m^2}(\widehat{\beta}_m - \mu_{\beta_m})^T\Sigma_m^{-1}(\widehat{\beta}_m - \mu_{\beta_m}) + \log f(m)$$

Additional penalty 2
(Shrinkage/Ridge type penalty)

Additional penalty 3
(from prior model probs)

Blackboard 19

2. An Idea:
Avoiding (?) the paradox

The idea -Step 2 [express posterior probs as BIC and additional penalties]

BIC can be obtained if

- *c_m=1*
- *Prior mean of β_m is set equal to its MLEs*
- *Prior model probabilities are assummed equal for all models.*

Similar to Kass and Wasserman (1995)

Blackboard 20

2. An Idea: Avoiding (?) the paradox

The idea - Step 2 [express posterior probs as BIC and additional penalties]

This term causes the Lindleys paradox since it explodes for large prior variances

$$\log f(m|y) \approx C + \log f(y|m,\widehat{\beta}_m) - \frac{d_m}{2}\log n - \frac{d_m}{2}\log c_m^2$$
$$-\frac{1}{2c_m^2}(\widehat{\beta}_m - \mu_{\beta_m})^T\Sigma_m^{-1}(\widehat{\beta}_m - \mu_{\beta_m}) + \log f(m)$$

This is eliminated for large prior variances

This still remains unspecified

And it can be used to effectively eliminate the prior penalty 1 which causes the paradoxical behavior

Blackboard 21

2. An Idea: Avoiding (?) the paradox

The idea -Step 3 [eliminating additional penalty 1]

We suggest choosing the cm freely to express the desired amount of shrinkage (to the prior mean), and choose prior model probabilities to adjust for the resulting effect this will have on the posterior model probabilities.

$$f(m) \propto p(m) c_m^{d_m} = p(m)\left(|V_m||i(_m)|\right)^{1/2}$$

where p(m) are some baseline model probabilities.

Blackboard 22

2. An Idea:
Avoiding (?) the paradox

The idea – Step 3 [eliminating additional penalty 1]

Under this prior set-up

$$\log f(m|y) \approx C + \log f(y|m, \widehat{\beta}_m) - \frac{d_m}{2}\log n - \cancel{\frac{d_m}{2}\log c_m^2}$$
$$-\frac{1}{2c_m^2}(\widehat{\beta}_m - \mu_{\beta_m})^T \Sigma_m^{-1}(\widehat{\beta}_m - \mu_{\beta_m}) + \cancel{\log f(m)}$$
$$\log p(m)$$

Log p(m) can be also interpreted as an additional dimension penalty

Blackboard 23

2. An Idea:
Avoiding (?) the paradox

For normal models

Using Normal-inverse-gamma prior set-up results become exact.

Here we present results using a multivariate normal prior for β_m with mean μ_{β_m} and variance $V_m \sigma^2$ and $f(\sigma^2) \propto 1/\sigma^2$

Using the prior model probabilities of type

$$f(m) \propto p(m)|V_m|^{\frac{1}{2}}|i(\widehat{\beta}_m) + n^{-1}V_m^{-1}|^{\frac{1}{2}}$$
$$= p(m)n^{-d_m/2}|V_m|^{\frac{1}{2}}|X_m^T X_m + V_m^{-1}|^{\frac{1}{2}}$$

Results in posterior model probabilities

$$\log f(m|y) = C - \frac{n}{2}\log\Big((y - X_m\widehat{\beta}_m)^T(y - X_m\widehat{\beta}_m)$$ Residual sum of squares

$$+(\widehat{\beta}_m - \mu_{\beta_m})^T V_m^{-1}(\widehat{\beta}_m - \mu_{\beta_m})\Big)$$ Shrinkage penalty

$$-\frac{d_m}{2}\log n + \log p(m)$$ Dimension penalty

Blackboard 24

2. An Idea: Avoiding (?) the paradox

What do we achieve

- *Separate the prior effect within each model from the posterior inference on model space*
- *The prior of the parameters contributes on the model evaluation through a shrinkage term measuring the difference between data and the prior*
- *The posterior (dimension) penalty on model space is solely controlled by p(m)*
- Setting all p(m) equal leads to a model determination based on a modified BIC involving penalized maximum likelihood.

Blackboard 25

2. An Idea: Avoiding (?) the paradox

What is p(m)?

p(m) can be based on a model complexity penalty which is a-priori seems to be appropriate.

Default option => Setting all p(m) equal leading to a modified BIC procedure

Hence, the impact of the prior distribution of the model parameters is through the shrinkage factor (additional penalty 2) and it is straightforward to assess, and any undesirable side effects of large prior variances are eliminated.

Blackboard 26

2. An Idea: Avoiding (?) the paradox

What is p(m)?

To chose p(m) such that it corresponds to a particular complexity penalty, we need to evaluate c_m^{-2} (i.e. the number of units of information introduced by the prior of β_m).

Except in certain cases, e.g. normal linear models, this quantity depends on the unknown model parameters β_m.

This is not appropriate as a specification for the marginal prior distribution over model space.

One possibility is to use a sample-based estimate in the Fisher information matrix to determine the `prior' model probability (not fully Bayesian).

Alternatively we may substitute β_m by its prior mean into the Fisher information matrix. This has a unit information interpretation but the model comparison is not asymptotically based the procedure described above (a correction term is required)

Blackboard 27

2. An Idea: Avoiding (?) the paradox

Some arguments in favor of this approach

ARGUMENT 1: *Constant probability in a neighborhood of the prior mean*

Let us consider the prior probability of the event

$$E = \{\text{model } m \text{ is 'true' }\} \cap \{(\beta_m - \mu_{\beta_m})^T V_m^{-1} (\beta_m - \mu_{\beta_m}) < \epsilon^2\}$$

Then, for any ε>0,

$$P(E) \;=\; f(m) P\left(\chi^2_{d_m} < \frac{\epsilon^2}{c_m^2}\right) \approx \frac{f(m)\epsilon^{d_m}}{2^{d_m/2-1}\Gamma(d_m/2) c_m^{d_m}}$$

Blackboard 28

2. An Idea:
Avoiding (?) the paradox

<u>*Some arguments in favor of this approach*</u>

ARGUMENT 1: Constant probability in a neighborhood of the prior mean

$$P(E) \;=\; f(m)P\left(\chi^2_{d_m} < \frac{\epsilon^2}{c_m^2}\right) \approx \frac{f(m)\epsilon^{d_m}}{2^{d_m/2-1}\Gamma(d_m/2)c_m^{d_m}}$$

Therefore, if the joint prior probability of model m in conjunction with β_m being in some specified neighborhood of its prior mean is to be uniform across models then we require

$$f(m) \propto p(m)c_m^{d_m} \text{ with } p(m) = 2^{d_m/2-1}\Gamma(d_m/2)/\epsilon^{d_m}$$

Blackboard 29

2. An Idea:
Avoiding (?) the paradox

<u>*Some arguments in favor of this approach*</u>

ARGUMENT 2: Flattening prior densities

Assume a baseline prior: $\beta_m|m \sim N(\mu_{\beta m}, \Sigma_m)$

We can raise this prior to the power of $1/c^2$ to make it flatter (and renormalize to make it again density); for $c^2>1$

The larger the c^2 the flatter the resulted prior.

For the above normal baseline prior, the new, heated, prior is

$$f_{heated}(\beta_m|m) \propto f(\beta_m|m)^{1/c^2}$$
$$= f_{Normal}\left(\beta_m; \mu_{\beta_m}, c^2\Sigma_m\right)$$

Blackboard 30

2. An Idea:
Avoiding (?) the paradox

Some arguments in favor of this approach

ARGUMENT 2: Flattening prior densities

Doing the same procedure for the joint prior on parameter and model space we end up to

$$f_{heated}(\beta_m, m) \propto f(\beta_m, m)^{1/c^2} = f(\beta_m|m)^{1/c^2} f(m)^{1/c^2}$$

$$= f_{Normal}\left(\beta_m; \mu_{\beta_m}, c^2\Sigma_m\right) \times [C_0 f(m)]^{1/c^2} \times c_m^{d_m} C_0^{-1}$$

Where $C_0 = (2\pi)^{-1/2} |\Sigma_m|^{-1/2}$ is the normalizing constant of the baseline prior

$$f_{heated}(m) = [C_0 f(m)]^{1/c^2} \times c_m^{d_m} C_0^{-1} \rightarrow (2\pi)^{d_m/2} c_m^{d_m} |\Sigma_m|^{1/2}$$

for large c_m

Blackboard 31

2. An Idea:
Avoiding (?) the paradox

Some arguments in favor of this approach

ARGUMENT 2: Flattening prior densities

Hence the heated prior model probabilities are equivalent to our proposal with

$$p(m) = (2\pi)^{d_m} \left(i(\beta_m)|\right)^{-1/2}$$

Implementing a procedure similar to Fisher Information Criterion (FIC, Wei, 1992, Annals of Statistics)

The above choice of p(m) does not requires to evaluate the Fisher information matrix

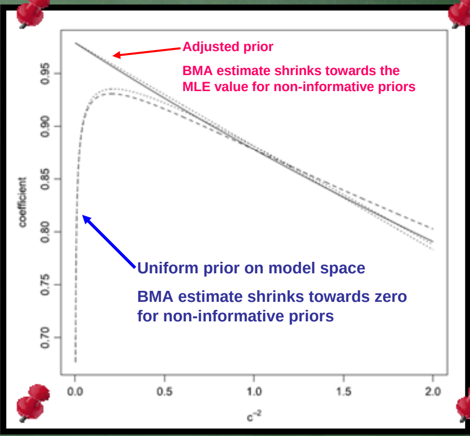
Blackboard 32

2. An Idea: Avoiding (?) the paradox

Some arguments in favor of this approach

ARGUMENT 3: *Bayesian model averaging and shrinkage*

Even the BMA estimates are affected by the Lindley's paradox and the proposed adjustment avoids the incoherent behavior



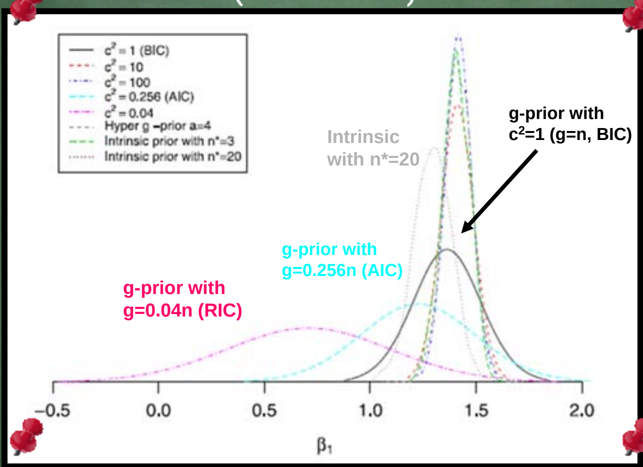
Comparison of two simple models differing by one coefficient β with MLE value equal to one

Blackboard 33

3. Illustrations and comparisons:

Example 1: A simple linear regression example

$$y_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right),\ i = 1, \ldots, n$$

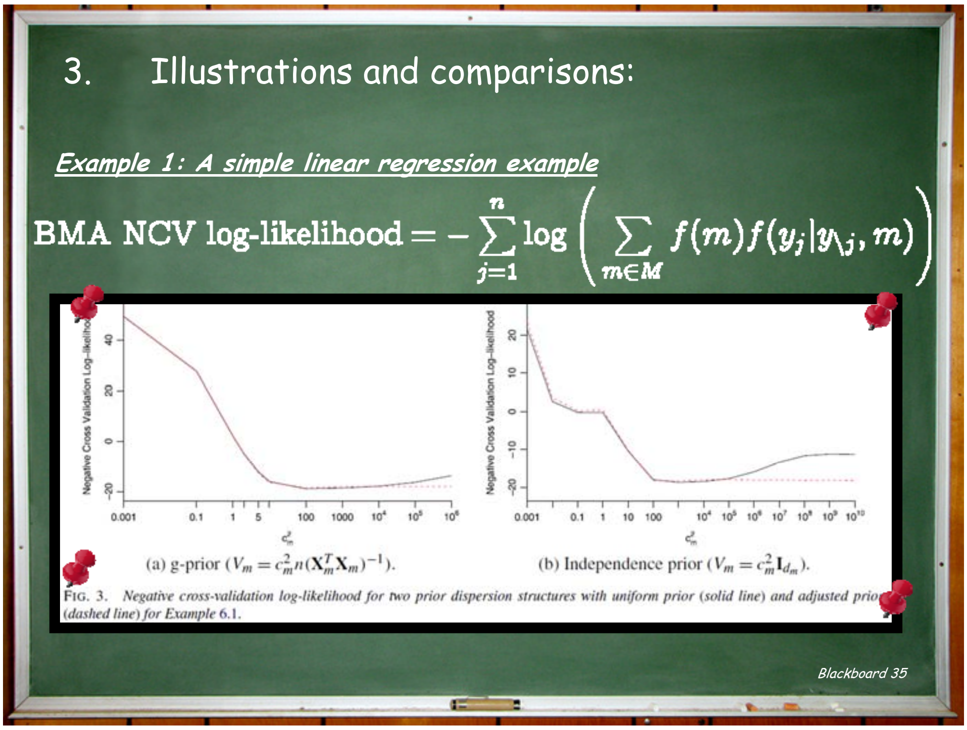


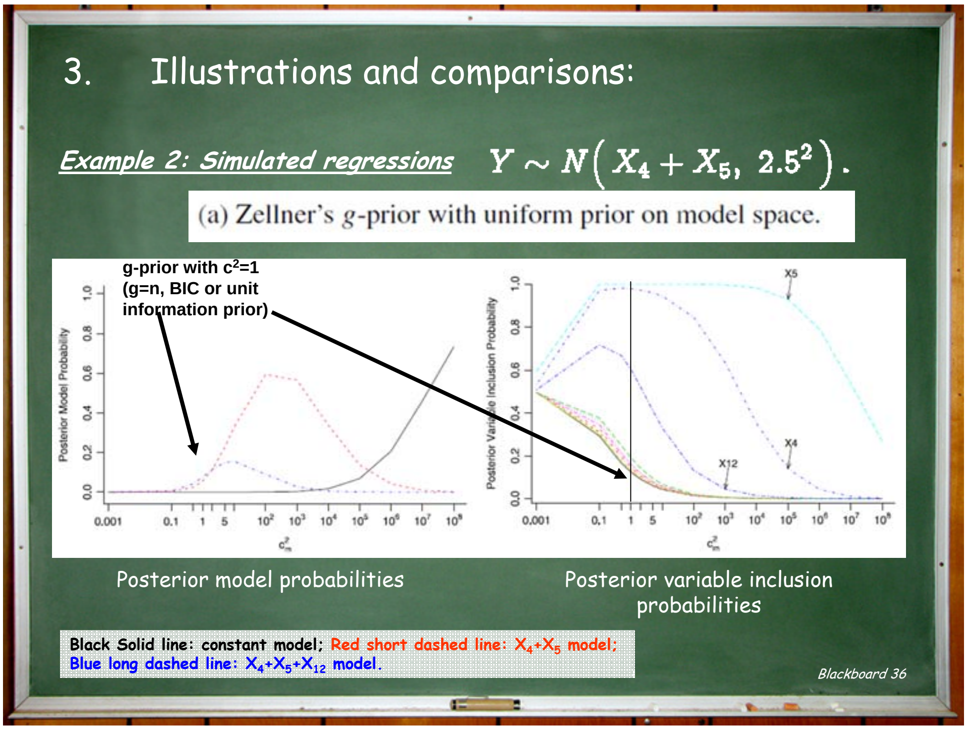
Data from Montgomery, Peck and Vining (2001) with n=25

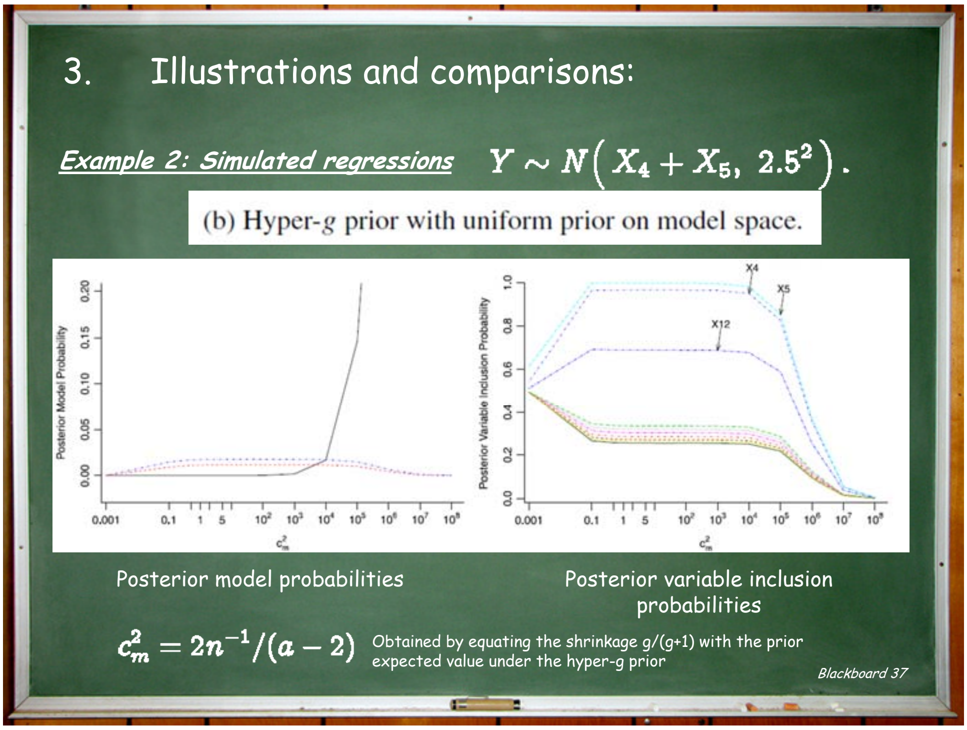
MLE for β_1=1.417

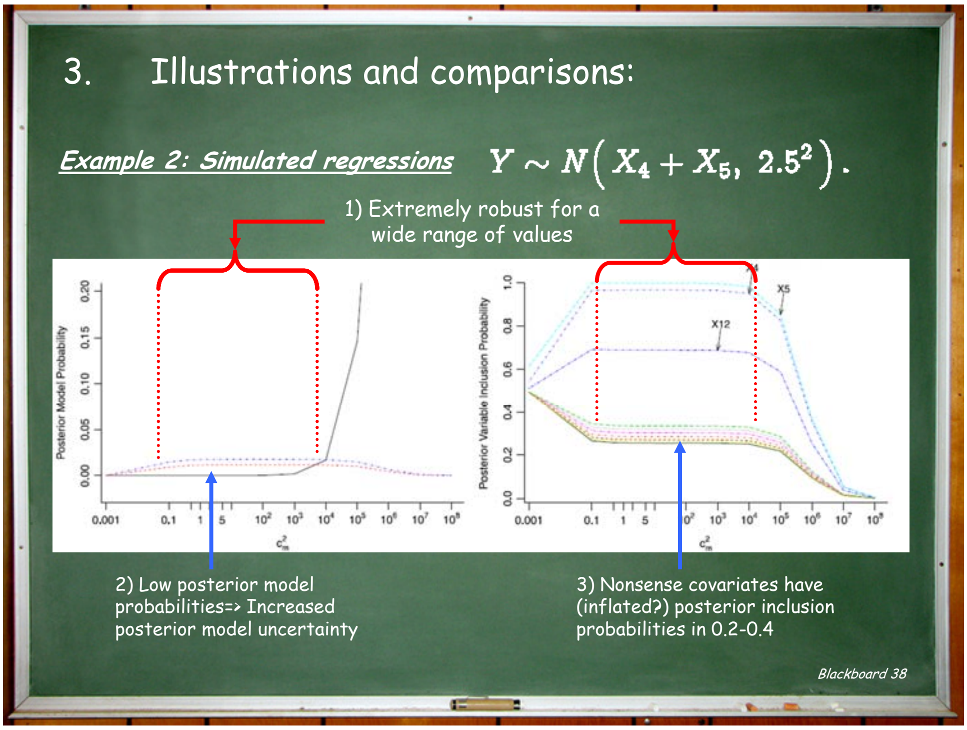
ρ= 0.98

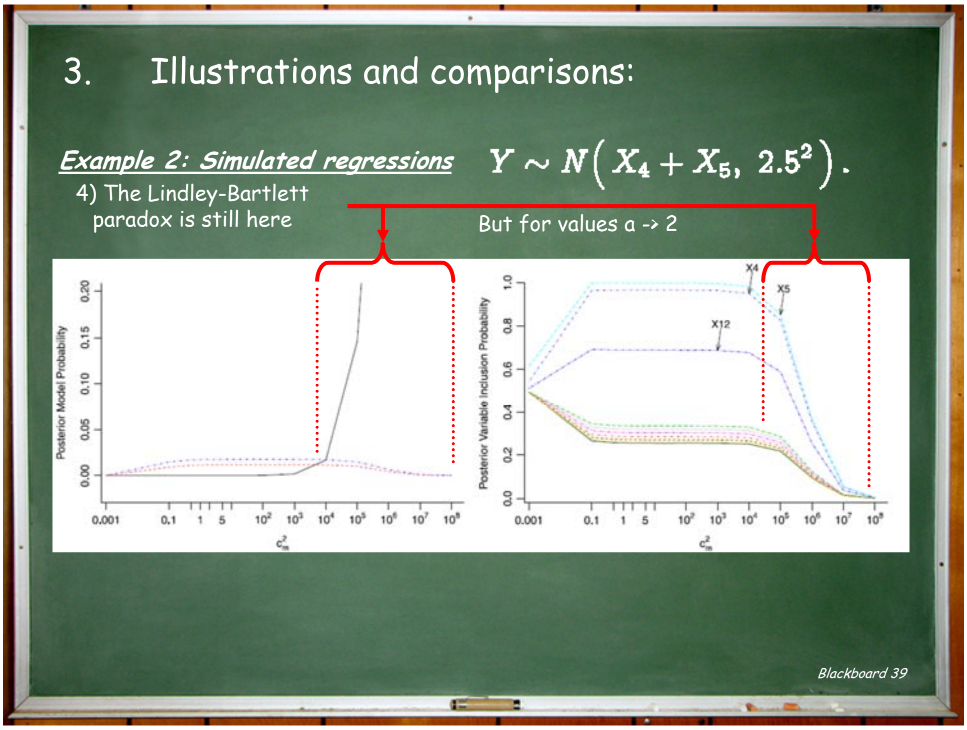
Blackboard 34

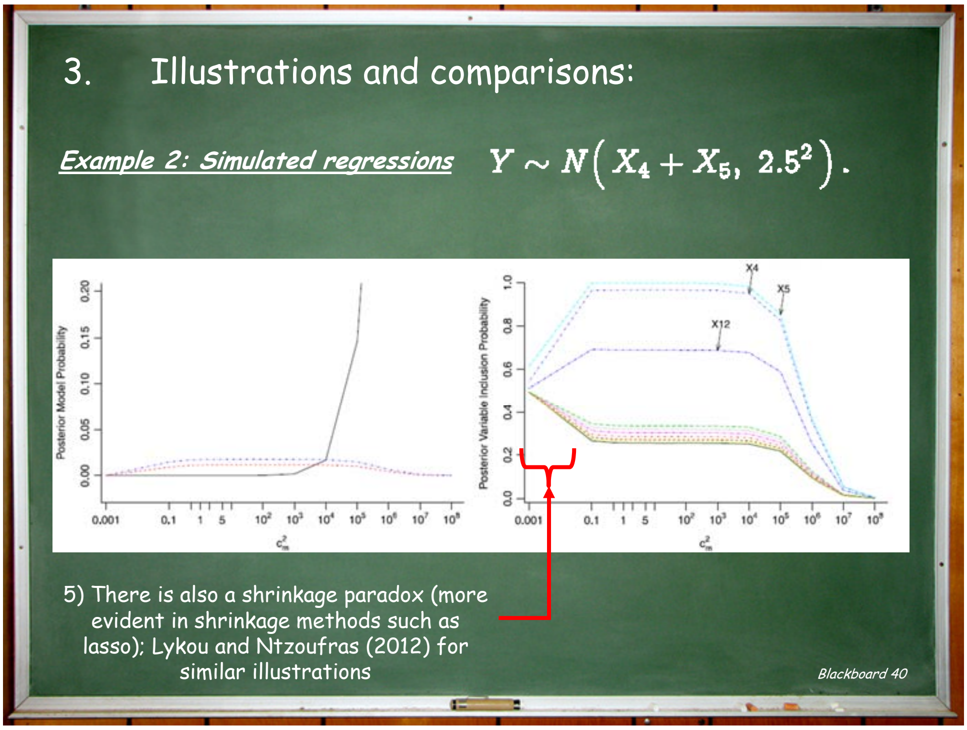

3. Illustrations and comparisons:
Example 1: A simple linear regression example
BMA NCV log-likelihood $= -\sum_{j=1}^{n} \log \left(\sum_{m \in M} f(m) f(y_j | y_{\setminus j}, m) \right)$
Negative Cross Validation Log-likelihood
c_m^2
(a) g-prior ($V_m = c_m^2 n (\mathbf{X}_m^T \mathbf{X}_m)^{-1}$).
(b) Independence prior ($V_m = c_m^2 \mathbf{I}_{d_m}$).
FIG. 3. Negative cross-validation log-likelihood for two prior dispersion structures with uniform prior (solid line) and adjusted prior (dashed line) for Example 6.1.
Blackboard 35


3. Illustrations and comparisons:
Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right)$.
(a) Zellner's g-prior with uniform prior on model space.
g-prior with c^2=1
(g=n, BIC or unit
information prior)
Posterior Model Probability
Posterior Variable Inclusion Probability
X5
X4
X12
c_m^2
Posterior model probabilities
Posterior variable inclusion probabilities
Black Solid line: constant model; Red short dashed line: X_4+X_5 model;
Blue long dashed line: X_4+X_5+X_{12} model.
Blackboard 36


3. Illustrations and comparisons:
Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right)$.
(b) Hyper-g prior with uniform prior on model space.
Posterior Model Probability
Posterior Variable Inclusion Probability
X4
X5
X12
c_m^2
Posterior model probabilities
Posterior variable inclusion probabilities
$c_m^2 = 2n^{-1}/(a-2)$
Obtained by equating the shrinkage g/(g+1) with the prior expected value under the hyper-g prior
Blackboard 37


3. Illustrations and comparisons:
Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right)$.
1) Extremely robust for a wide range of values
Posterior Model Probability
Posterior Variable Inclusion Probability
X5
X12
c_m^2
2) Low posterior model probabilities=> Increased posterior model uncertainty
3) Nonsense covariates have (inflated?) posterior inclusion probabilities in 0.2-0.4
Blackboard 38

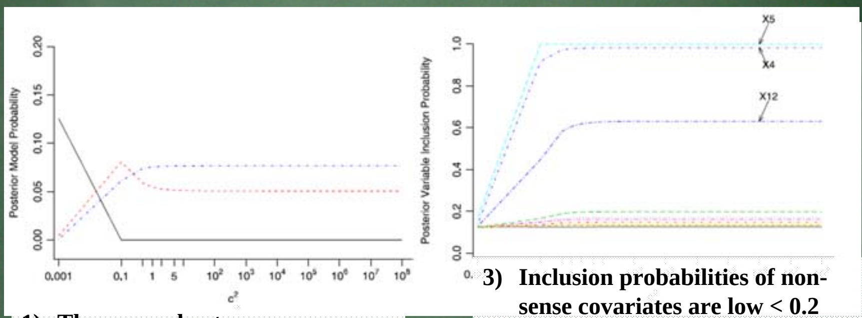

3. Illustrations and comparisons:
Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right)$.
4) The Lindley-Bartlett paradox is still here
But for values a -> 2
Posterior Model Probability
Posterior Variable Inclusion Probability
X4
X5
X12
Blackboard 39


3. Illustrations and comparisons:
Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right)$.
Posterior Model Probability
Posterior Variable Inclusion Probability
X4
X5
X12
5) There is also a shrinkage paradox (more evident in shrinkage methods such as lasso); Lykou and Ntzoufras (2012) for similar illustrations
Blackboard 40

3. Illustrations and comparisons:

Example 2: Simulated regressions $Y \sim N\left(X_4 + X_5,\ 2.5^2\right).$

(c) Zellner's g-prior with adjusted prior on model space.



1) They are robust
2) They do give reasonably high posterior model probabilities to best models
3) Inclusion probabilities of non-sense covariates are low < 0.2
4) They do not suffer from Lindleys paradox
5) No shrinkage paradox

3. Illustrations and comparisons:

Example 3: A 3x2x4 contingency table example with available prior information

	Parameter prior	Model space prior	Posterior model probabilities O + H + A	OH + A	O + HA	OH + HA
1.	DF	uniform	0.657	0.336	0.004	0.002
2.	KS	uniform	0.075	0.000	0.923	0.002
3.	KS/DF	uniform	0.059	0.023	0.638	0.280
4.	DF	adjusted	0.677	0.317	0.004	0.002
5.	KS	adjusted	0.665	0.335	0.000	0.000
6.	KS/DF	adjusted	0.690	0.310	0.000	0.000
7.	IND	adjusted	0.690	0.303	0.004	0.003

DF=Dellaportas & Forster prior (non-informative for model comparison);
KS=Knuiman and Speeed prior (informative within each model)
IND=Independence prior

Blackboard 42

4. Conclusion

What we do argue is:

1) there is nothing sacred about a uniform prior distribution over models.
2) It is reasonable to consider specifying jointly $f(\beta_m, m)$ and hence $f(m)$ in a way which takes account of the prior distributions for the model parameters for individual models.

We propose priors of type $f(m) \propto p(m) c_m^{d_m}$ as a possible choice which

a) Separates (in a reasonable way) inference within and across models

β) Avoids Lindley-Bartlett paradox

γ) Implements a desired complexity/dimensionality penalty and a shrinkage penalty (on the same time)

δ) Can be used <u>even when the prior is informative</u> for some parameters

Blackboard 43

5. Coffee time

At last …

Blackboard 44